Wanda Andreoni
Sidney Yip
Editors

Handbook of Materials Modeling

Applications: Current and Emerging Materials

Second Edition

Springer

Handbook of Materials Modeling

Wanda Andreoni • Sidney Yip

Editors

Handbook of Materials Modeling

Applications: Current and Emerging Materials

Second Edition

With 930 Figures and 39 Tables

Springer

Editors
Wanda Andreoni
Institute of Physics
Swiss Federal Institute of Technology – Lausanne
Lausanne, Switzerland

Sidney Yip
Department of Nuclear Science and Engineering
Department of Materials Science and Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

We wish to dedicate this book to the memory of Walter Kohn and Aneesur Rahman, founders of the modern approach to the physics and chemistry of real materials, both theoretical and computational. While their contributions will forever stand as milestones in the progress of science, we wish to commemorate their courage and vision in proposing new methods that would be fully implemented only decades later. We – and everyone who has known them personally – continue to remember them also for their tireless enthusiasm for science, their generous sharing of ideas, their unique kindness, and their charisma. As their articles start to disappear from citation lists in the same way as those of the founders of quantum mechanics, it is both humbling and comforting to know that any present and future development rests on the work of two wonderful people who have truly embodied the values of science.

Wanda Andreoni and Sidney Yip

Preface to *HMM2*

The *Handbook of Materials Modeling*, published in 2005, was the first reference work of its kind, providing a comprehensive description of an emerging field of research and helping to grow a community at the intersection of *Computational Science* and *Materials Science and Technology* through the development of the theory, modeling, and simulation of materials across the disciplines of physics, chemistry, and engineering. In producing the first edition, the editors were guided by a simple vision – target students, young researchers, and nonspecialists as the primary audience, introduce the concept and practice of linking materials research at different resolution levels (from the atomic to the meso- to the macroscale), and stimulate exploration and applications. The e-book version became available two years later.

The *Handbook of Materials Modeling Second Edition* (*HMM2*) follows its predecessor in serving the broad community of *Computational Science* and *Materials Science and Technology*. Its goal is to update the current state of the art of materials modeling and simulation in two complementary ways:

- *Strengthen* the foundations of materials theory, modeling, and simulation
- *Broaden* the scope to include challenges and opportunities of interdisciplinary interest

In both objectives, *HMM2* strives to promote open discussion of the capabilities and limitations of current methodologies and applications. The Second Edition is composed of two companion sets: *Handbook of Materials Modeling – Methods: Theory and Modeling*, and *Handbook of Materials Modeling – Applications: Current and Emerging Materials*. These will be abbreviated as *MTM* and *ACE*, respectively. Each set is a stand-alone publication with its own Table of Contents and individual chapters that are grouped under a number of parts. Both *MTM* and *ACE* are three-volume sets. There are 11 parts in *MTM* containing, beyond the Introduction, 10 sections with 88 chapters and 16 parts in *ACE* containing, beyond the Introduction, 15 sections with 120 chapters. The chapters which constitute the primary scientific content of the handbook are organized by the section editors in each set, working in coordination with the editors in chief. The section editors also provide an introductory overview of their own section, thus facilitating access to the following chapters.

Logistical issues aside, we think it is important to emphasize the holistic aspects of materials modeling and simulation, as an evolving discipline across science and technology and the complementarity of the two sets. As the community grows in scope and relevance, learning "in depth" with the aim of mastering the relevant theoretical and computational algorithms becomes more and more important. Equally, it becomes important to be able to discern unresolved materials issues for which information from simulations could "make a difference" and to properly apply and critically assess the chosen models and methods. Thus *all readers* should recognize the beneficial overlap of the two sets. In the respective introductions we will elaborate further on this point.

An undertaking of this magnitude can only succeed as a collective effort. The *Handbook of Materials Modeling Second Edition* would not be a reality without the dedicated efforts of former Springer Editor *Maria Bellantone*, to whom we are especially indebted. She brought us together and provided us with essential logistical support. We thank *Lydia Mueller* for her early guidance. Our deepest gratitude goes to *Juby George* and *Sunaina Dadhwal* for their invaluable editorial assistance, efficient communication with authors and editors, and efficient management of the Springer platform *Meteor*. Their assistance made it possible to complete the project on an admittedly ambitious schedule. That 200 plus chapters could be commissioned, manuscripts collected, revisions, etc., be undertaken and published in less than four years is an accomplishment for which the credit properly should go to the section editors and their authors. We would like to think all participants will take pride in what we have achieved together. This is a journey that will continue. We trust the community will be diligent in upholding standards and maintaining respect and collegiality among all stakeholders.

February 2020   Wanda Andreoni
                                                                                                   Sidney Yip
                                                                                                    Editors

Contents

Contents   xi

Volume 2

Volume 3

About the Editors

Wanda Andreoni
Institute of Physics
Swiss Federal Institute
of Technology – Lausanne
Lausanne, Switzerland

Wanda Andreoni is Emeritus Professor of Physics at the Swiss Federal Institute of Technology of Lausanne (EPFL). She has been involved in Computational Materials Science since the early days of her research activity, in both academic and industrial institutions, covering also teaching, organizational, and research managing positions. She is a Fellow of the American Physical Society. She was Zernike Professor at the Groningen University (NL) (2011) and CECAM Director (2009–2012). Regarding editorial work, she was Coeditor of *Europhysics Letters* (1990–1993) and Editor of *The Chemical Physics of Fullerenes 10 (and 5) Years Later*, NATO ASI Series E: Applied Sciences, Vol. 316 (Kluwer, 1996), and of *The Physics of Fullerene-Based and Fullerene-Related Materials*, Series on the Physics and Chemistry of Materials with Low-Dimensional Structures, Vol. 23 (Kluwer, 2000).

Sidney Yip
Department of Nuclear Science
and Engineering
Department of Materials Science
and Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

Sidney Yip immigrated to the USA from China in 1950 at age 14. After receiving degrees in Mechanical Engineering and Nuclear Engineering from the University of Michigan and spending two postdoctoral years at Cornell University, he joined the Nuclear Engineering Faculty at MIT in 1965. From early research in theoretical studies of particle and fluid transport, he became broadly involved in atomistic modeling and simulation of materials. He edited the first edition of the *Handbook of Materials Modeling* as well as *Spectroscopy in Biology and Chemistry: Neutron, X-Ray, Laser* (1974) and *Materials Interfaces: Atomic-Level Structure and Properties* (1993). Other books include the monographs, *Foundations of Neutron Transport Theory* (1967), *Neutron Molecular Spectroscopy* (1968), and *Molecular Hydrodynamics* (1980), and a text, *Nuclear Radiation Interactions* (2014). He received Guggenheim Fellowship, US Senior Scientist Prize of the Alexander von Humboldt Foundation, Distinguished Alumnus Award of the University of Michigan, and the Robert Cahn Award. A Fellow of the American Physical Society, he became Professor Emeritus in 2009.

Section Editors

Wanda Andreoni Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland

Part: Plenary Topics

Marco Bernardi Department of Applied Physics and Materials Science, California Institute of Technology, Pasadena, CA, USA

Part: Photovoltaics: First Principles Modeling

Markus J. Buehler Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Part: Hierarchical Materials Modeling: Mechanical Performance

James R. Chelikowsky Center for Computational Materials, Oden Institute for Computational Engineering and Sciences, Departments of Physics and Chemical Engineering, The University of Texas at Austin, Austin, TX, USA

Part: Low-Dimensional Materials at the Nanoscale

Ming Dao Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Part: Multiscale Modeling of Diseases

Emanuela Del Gado Department of Physics, Institute for Soft Matter Synthesis and Metrology, Georgetown University, Washington, DC, USA

Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA, USA

Part: Modeling the Structural Development and Mechanics of Complex Soft Materials

Ram Devanathan Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA

Part: Nuclear Materials

Davide Donadio Department of Chemistry, University of California Davis, Davis, USA

IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Part: Thermal Transport

Cesare Franchini Faculty of Physics and Center for Computational Materials Science, University of Vienna, Vienna, Austria

Part: Oxides in Energy and Information Technologies

Jeffrey C. Grossman Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Part: Photovoltaics: First Principles Modeling

George E. Karniadakis Division of Applied Mathematics, Brown University, Providence, RI, USA

Part: Multiscale Modeling of Diseases

Jaime Marian Department of Materials Science and Engineering and Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA

Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA

Part: Nuclear Materials

Francisco J. Martin-Martinez Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Part: Hierarchical Materials Modeling: Mechanical Performance

John C. Mauro Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA, USA

Part: Glass Science and Technology: Predictive Modeling

Horia Metiu   Department of Chemistry and Biochemistry, University of California at Santa Barbara, Santa Barbara, CA, USA

Part: Surface Catalysis

Kai Nordlund   Department of Physics and Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland

Part: Radiation Damage

Mehmet C. Onbaşlı Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Electrical and Electronics Engineering, Koç University, Istanbul, Turkey

Part: Glass Science and Technology: Predictive Modeling

Roland J.-M. Pellenq Concrete Sustainability Hub, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

MSE2, the MIT/CNRS/Aix-Marseille University Joint Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Centre Interdisciplinaire de Nanosciences de Marseille, CINaM, CNRS/Aix-Marseille Université, Marseille, France

UMI MSE2, CNRS/MIT/AMU Joint Laboratory "MultiScale Materials Science for Energy and Environment", Marseille, France

Part: Modeling the Structural Development and Mechanics of Complex Soft Materials

Karsten Reuter Chair for Theoretical Chemistry and Catalysis Research Center, Technical University of Munich, Garching, Germany

Part: Surface Catalysis

Stefano Sanvito Center for Materials Genomics, Duke University, Durham, NC, USA

School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Part: Modeling Applications to Magnetism, Magnetic Materials, and Spintronics

Michael P. Short Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Part: Radiation Damage

Cai-Zhuang Wang Ames Laboratory – USDOE and Department of Physics and Astronomy, Iowa State University, Ames, IA, USA

Part: Computational Crystal Structure: Prediction and Materials Discovery

Christopher Wolverton Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA

Part: Computational Crystal Structure: Prediction and Materials Discovery

Bilge Yildiz Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Part: Oxides in Energy and Information Technologies

Sidney Yip Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Part: Atomistic Simulations; Plenary Topics

Ting Zhu Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

School of Material Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Part: Nanomechanics of Materials: Structure and Deformation

Contributors

Vishal Agarwal Department of Chemical Engineering, Indian Institute of Technology, Kanpur, Uttar Pradesh, India

Karim Ahmed Department of Nuclear Engineering, Texas A&M University, College Station, TX, USA

Zlatan Aksamija Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA

Konstantinos Alexopoulos Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA

Zahed Allahyari Materials science and Engineering, Skolkovo Institute of Science and Technology, Moscow, Russia
Moscow Institute of Physics and Technology, Moscow, Russia

Julio A. Alonso Departamento de Física Teórica, Atómica y Optica, University of Valladolid, Valladolid, Spain

Maximilian Amsler Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY, USA

David Andersson Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA

Wanda Andreoni Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland

Aleandro Antidormi Department of Physics, University of Cagliari, Cittadella Universitaria, Monserrato(Ca), Italy

Kazuto Arakawa Department of Materials Science, Faculty of Science and Engineering, Shimane University, Matsue, Japan

T. Archer School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Sina Askarinejad Department of Civil and Environmental Engineering, Worcester Polytechnic Institute, Worcester, MA, USA

Alexander Bagger Department of Chemistry, University of Copenhagen, Copenhagen, Denmark

Chunguang G. Bai Institute of Metal Research, Chinese Academy of Sciences, Shenyang, China

J. P. Balbuena Departamento de Física Aplicada, Facultad de Ciencias, Fase II, Universidad de Alicante, Alicante, Spain

Stefano Baroni SISSA – Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy

CNR Istituto Officina dei Materiali, Trieste, Italy

Diego López Barreiro Laboratory for Atomistic and Molecular Mechanics Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Mathieu Bauchy Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA

Charlotte S. Becquart Univ.Lille, CNRS, INRA, ENSCL, UMR 8207, UMET, Unité Matériaux et Transformations, Lille, France

Marco Bernardi Department of Applied Physics and Materials Science, California Institute of Technology, Pasadena, CA, USA

Riccardo Bertossa SISSA, Trieste, Italy

Theodore M. Besmann College of Engineering and Computing, University of South Carolina, Columbia, SC, USA

Xin Bian Computational Science and Engineering Laboratory, Clausiusstrasse, ETH Zurich, Switzerland

Danny Bluestein Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA

Lilia Boeri Dipartimento di Fisica, Sapienza Università di Roma, Roma, Italy

Venkatesh Botu Science and Technology Division, Corning Incorporated, Corning, NY, USA

Mehdi Bouzid LPTMS, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Orsay, France

Markus J. Buehler Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Oral Büyüköztürk Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Penghui Cao Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA

Ivano E. Castelli Department of Energy Conversion and Storage, Technical University of Denmark, Lyngby, Denmark

Nicolas Castin The Belgian Nuclear Research Centre (SCK· CEN), Nuclear Materials Science (NMS) Institute, Structural Materials (SMA) Expert Group, Mol, CP, Belgium

María José Caturla Departamento de Física Aplicada, Facultad de Ciencias, Fase II, Universidad de Alicante, Alicante, Spain

Andrea Cepellotti Department of Physics, University of California at Berkeley, Berkeley, CA, USA

Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Theory and Simulations of Materials (THEOS), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Yann Chalopin EM2C Lab, Ecole CentraleSupelec/CNRS, Gif-sur-Yvette, France

James R. Chelikowsky Center for Computational Materials, Oden Institute for Computational Engineering and Sciences, Departments of Physics and Chemical Engineering, The University of Texas at Austin, Austin, TX, USA

Dengke Chen Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Li Chen Schlumberger Oilfield Services, Houston, TX, USA

Long-Qing Chen Department of Materials Science and Engineering, The Pennsylvania State University, Pennsylvania, PA, USA

Eunseog Cho Samsung Advanced Institute of Technology, Suwon, Gyeonggi-do, Republic of Korea

Deung-Jang Choi Centro de Física de Materiales CFM/MPC (CSIC-UPV/EHU), Donostia-San Sebastian, Spain

Donostia International Physics Center (DIPC), Donostia-San Sebastian, Spain

Jiwoong Choi Department of Mechanical Engineering, The University of Iowa, Iowa City, IA, USA

IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA, USA

Sanghun Choi School of Mechanical Engineering, Kyungpook National University, Daegu, Republic of Korea

Iek-Heng Chu Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA

Matteo Ciccotti Laboratoire Sciences et Ingénierie de la Matière Molle (SIMM), ESPCI Paris, PSL University, Sorbonne Université, CNRS, Paris, France

Luciano Colombo Department of Physics, University of Cagliari, Cittadella Universitaria, Monserrato(Ca), Italy

S. Curtarolo Center for Materials Genomics, Duke University, Durham, NC, USA
Departments of Mechanical Engineering and Materials Science, Physics, and Chemistry, Duke University, Durham, NC, USA

Gustavo Martini Dalpian Center for Natural and Human Sciences, Federal University of ABC, Santo André, SP, Brazil

Ming Dao Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Anil K. Dasanna BioQuant-Center for Quantitative Biology and Institute of Theoretical Physics, Heidelberg University, Heidelberg, Germany

Emanuela Del Gado Department of Physics, Institute for Soft Matter Synthesis and Metrology, Georgetown University, Washington, DC, USA
Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA, USA

Alexander A. Demkov Department of Physics, The University of Texas, Austin, TX, USA

Yuefan Deng Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA

Olaf Deutschmann Institute for Chemical Technology and Polymer Chemistry and Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Ram Devanathan Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA

Ulrike Diebold Institute of Applied Physics, Technische Universität Wien, Vienna, Austria

Flyura Djurabekova Department of Physics and Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland

Jose A. Dobado Facultad de Ciencias, Departamento de Química Orgánica, Universidad de Granada, Granada, Spain

Kapildeb Dolui Department of Physics and Astronomy, University of Delaware, Newark, DE, USA

Christophe Domain Département MMC, Les Renardières, Moret sur Loing, France

Davide Donadio Department of Chemistry, University of California Davis, Davis, USA

IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Hadrien Dumont Schlumberger Oilfield Services, Houston, TX, USA

Davoud Ebrahimi Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Anter El-Azab School of Materials Engineering, Purdue University, West Lafayette, IN, USA

Loris Ercole SISSA, Trieste, Italy

Richard F. L. Evans The Department of Physics, The University of York, Heslington, York, UK

Yue Fan Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA

Dmitry A. Fedosov Institute of Complex Systems and Institute for Advanced Simulation, Forschungszentrum Jülich, Jülich, Germany

Marina R. Filip Department of Materials, University of Oxford, Oxford, UK

Steve Fitzgerald School of Mathematics, University of Leeds, Leeds, UK

Cesare Franchini Faculty of Physics and Center for Computational Materials Science, University of Vienna, Vienna, Austria

Kurt D. Fredrickson Patterson+Sheridan LLC, San Jose, CA, USA

Huajian Gao School of Engineering, Brown University, Providence, RI, USA

Matthieu George Laboratoire Charles Coulomb (L2C), Université de Montpellier, CNRS, Montpellier, France

Nasr M. Ghoniem Department of Mechanical and Aerospace Engineering, University of California, Los Angeles (UCLA), Los Angeles, CA, USA

Feliciano Giustino Department of Materials Science and Engineering, Cornell University, Ithaca, NY, USA

Department of Materials, University of Oxford, Oxford, UK

Gerhard Gompper Institute of Complex Systems and Institute for Advanced Simulation, Forschungszentrum Jülich, Jülich, Germany

Xin-Gao Gong Key Laboratory for Computational Physical Science (Ministry of Education), Fudan University, Shanghai, China

Sushmit Goyal Science and Technology Division, Corning Incorporated, Corning, NY, USA

Federico Grasselli SISSA, Trieste, Italy

Simon J. Greaves Research Institute of Electrical Communication, Tohoku University, Sendai, Japan

Axel Groß Helmholtz Institute Ulm (HIU) Electrochemical Energy Storage, Ulm, Germany

Institute of Theoretical Chemistry, Ulm University, Ulm, Germany

Jeffrey C. Grossman Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Hui-Jun Gu Key Laboratory for Computational Physical Science (Ministry of Education), Fudan University, Shanghai, China

Babak Haghighi Department of Mechanical Engineering, The University of Iowa, Iowa City, IA, USA

IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA, USA

Jason Hales Fuel Modeling and Simulation Department, Idaho National Laboratory, Idaho Falls, ID, USA

Martin Hangaard Hansen SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, California, CA, USA

Riccardo Hertel Université de Strasbourg, CNRS, Institut de Physique et Chimie des Matériaux de Strasbourg, Strasbourg, France

Franziska Hess Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Matthias Hettel Institute for Chemical Technology and Polymer Chemistry, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Kai-Ming Ho Ames Laboratory and Department of Physics, Iowa State University, Ames, IA, USA

Oded Hod Department of Physical Chemistry, School of Chemistry, The Raymond and Beverly Sackler Faculty of Exact Sciences and The Sackler Center for Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv, Israel

Eric A. Hoffman Department of Radiology, The University of Iowa, Iowa City, IA, USA

Department of Biomedical Engineering, The University of Iowa, Iowa City, IA, USA

Hossein Honarvar Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, USA

Bing Huang Materials and Energy Division, Beijing Computational Science Research Center, Beijing, China

Mahmoud I. Hussein Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, USA

Katerina Ioannidou Department of Civil and Environmental Engineering, MIT Energy Initiative, Multiscale Materials Science for Energy and Environment, Joint MIT-CNRS laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Thomas Jourdan DEN-Service de Recherches de Métallurgie Physique, CEA, Université Paris-Saclay, Gif-sur-Yvette, France

Ken Kamrin Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

George E. Karniadakis Division of Applied Mathematics, Brown University, Providence, RI, USA

Pawel Keblinski Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

John Kilner International Institute for Carbon Neutral Energy Research (WPI-I2CNER), Kyushu University, Fukuoka, Japan

Department of Materials, Imperial College London, London, UK

Hyunbin Kim Corning Technology Center Korea,Corning Precision Materials Co., Ltd., Asan, Republic of Korea

Boris Kozinsky John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Georg Kresse University of Vienna, Vienna, Austria

Henrik H. Kristoffersen DTU Energy, Technical University of Denmark, Kgs. Lyngby, Denmark

Frank Lechermann Institut für Theoretische Physik, Universität Hamburg, Hamburg, Germany

Sung Hoon Lee Corning Technology Center Korea, Corning Precision Materials Co., Ltd., Asan, Republic of Korea

Huan Lei Advanced Computing, Mathematics and Data, Pacific Northwest National Laboratory, Richland, WA, USA

David M. Leitner Department of Chemistry, University of Nevada, Reno, NV, USA

Ju Li Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Mingda Li Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Quan Li State Key Laboratory of Superhard Materials and Innovation Center of Computational Physics Methods and Software, College of Physics, Jilin University, Changchun, China

International Center of Future Science, Jilin University, Changchun, China

Sichi Li Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA

Weidong Li Department of Materials Science and Engineering, The University of Tennessee, Knoxville, TN, USA

Xiaoyan Li Center for Advanced Mechanics and Materials, Applied Mechanics Laboratory, Department of Engineering Mechanics, Tsinghua University, Beijing, China

Xuejin Li Department of Engineering Mechanics, Zhejiang University, Zhejiang, People's Republic of China

Division of Applied Mathematics, Brown University, Providence, RI, USA

Zhen Li Department of Mechanical Engineering, Clemson University, Clemson, SC, USA

Ching-Long Lin Department of Mechanical Engineering, The University of Iowa, Iowa City, IA, USA

Department of Radiology, The University of Iowa, Iowa City, IA, USA

IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA, USA

Lucas Lindsay Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Zishun Liu International Center for Applied Mechanics, State Key Laboratory for Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an, China

Zhi-Pan Liu Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science (Ministry of Education), Department of Chemistry, Fudan University, Shanghai, China

María J. López Departamento de Física Teórica, Atómica y Optica, University of Valladolid, Valladolid, Spain

Nicolás Lorente Centro de Física de Materiales CFM/MPC (CSIC-UPV/EHU), Donostia-San Sebastian, Spain

Donostia International Physics Center (DIPC), Donostia-San Sebastian, Spain

Huijie Lu Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN, USA

Xue-Zeng Lu Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA

Jian Lv State Key Laboratory of Superhard Materials and Innovation Center of Computational Physics Methods and Software, College of Physics, Jilin University, Changchun, China

Kirill Lykov Faculty of Informatics, Institute of Computational Science, Universita della Svizzera Italiana, Lugano, Switzerland

Yanming Ma State Key Laboratory of Superhard Materials and Innovation Center of Computational Physics Methods and Software, College of Physics, Jilin University, Changchun, China

International Center of Future Science, Jilin University, Changchun, China

Lorenzo Malerba The Belgian Nuclear Research Centre (SCK· CEN), Nuclear Materials Science (NMS) Institute, Structural Materials (SMA) Expert Group, Mol, CP, Belgium

Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avda., Spain

Aris Marcolongo Cognitive Computing and Computational Sciences Department, IBM Research, Zürich, Switzerland

THEOS-MARVEL, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Jaime Marian Department of Materials Science and Engineering and Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA

Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA

Troels Markussen Synopsys QuantumWise, Copenhagen, Denmark

Kirsten Martens Université Grenoble Alpes, CNRS, LIPhy, Grenoble, France

Enrique Martínez Material Science and Technology Division, MST-8, Los Alamos National Laboratory, Los Alamos, NM, USA

Francisco J. Martin-Martinez Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Nicola Marzari National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Theory and Simulations of Materials (THEOS), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Enrico Masoero School of Engineering, Newcastle University, Newcastle upon Tyne, UK

John C. Mauro Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA, USA

Mathew McKenzie Science and Technology Division, Corning Incorporated, Corning, NY, USA

Jacob W. McMurray Materials Science and Technology, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Ying Shirley Meng Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA

Bryce Meredig Citrine Informatics Inc., Redwood City, CA, USA

Horia Metiu Department of Chemistry and Biochemistry, University of California at Santa Barbara, Santa Barbara, CA, USA

Kyle Michel Citrine Informatics Inc., Redwood City, CA, USA

Kyoungmin Mim Samsung Advanced Institute of Technology, Suwon, Gyeonggido, Republic of Korea

Vinay K. Mishra Schlumberger Oilfield Services, Houston, TX, USA

Normand Mousseau Département de physique and Regroupement québécois sur les matériaux de pointe, Université de Montréal, Montréal, QC, Canada

Kosar Mozaffari Department of Mechanical Engineering, Cullen College of Engineering, University of Houston, Houston, TX, USA

Oliver C. Mullins Schlumberger Oilfield Services, Houston, TX, USA

J. Nelson School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Teng Yong Ng School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore, Singapore

Manh Cuong Nguyen Ames Laboratory and Department of Physics, Iowa State University, Ames, IA, USA

Branislav K. Nikolić Department of Physics and Astronomy, University of Delaware, Newark, DE, USA

Kai Nordlund Department of Physics and Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland

Harald Oberhofer Chair for Theoretical Chemistry and Catalysis Research Center, Technische Universität München, Garching, Germany

Artem R. Oganov Materials science and Engineering, Skolkovo Institute of Science and Technology, Moscow, Russia

Moscow Institute of Physics and Technology, Moscow, Russia

International Center for Materials Design, Northwestern Polytechnical University, Xi'an, China

Shigenobu Ogata Department of Mechanical Science and Bioengineering, Graduate School of Engineering Science, Osaka University, Osaka, Japan

Mehmet C. Onbaşlı Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Electrical and Electronics Engineering, Koç University, Istanbul, Turkey

Shyue Ping Ong Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA

C. Oses Center for Materials Genomics, Duke University, Durham, NC, USA

Departments of Mechanical Engineering and Materials Science, Physics, and Chemistry, Duke University, Durham, NC, USA

Andrew O'Hara Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA

Ricardo Pablo-Pedro Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Steven D. Palkovic Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Wenxiao Pan Department of Mechanical Engineering, University of Wisconsin-Madison, Madison, WI, USA

Harold S. Park Department of Mechanical Engineering, Boston University, Boston, MA, USA

Hyunhang Park Corning Technology Center Korea, Corning Precision Materials Co., Ltd., Asan, Republic of Korea

Mike Payne Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK

Q. X. Pei Institute of High Performance Computing, A*STAR, Singapore, Singapore

Roland J.-M. Pellenq Concrete Sustainability Hub, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

MSE2, the MIT/CNRS/Aix-Marseille University Joint Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Centre Interdisciplinaire de Nanosciences de Marseille, CINaM, CNRS/Aix-Marseille Université, Marseille, France

UMI MSE, CNRS/MIT/AMU Joint Laboratory "MultiScale Materials Science for Energy and Environment", Marseille, France

Zhangli Peng Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN, USA

Marko D. Petrović Department of Mathematical Sciences, University of Delaware, Newark, DE, USA

Silvia Picozzi Consiglio Nazionale delle Ricerche, CNR-SPIN, c/o Univ. Chieti-Pescara "G. D'Annunzio", Chieti, Italy

Carlo A. Pignedoli Swiss Federal Laboratories for Materials Science and Technology, Empa, Dübendorf, Switzerland

Stefano Pittalis Istituto Nanoscienze – CNR, Modena, Italy

Igor V. Pivkin Faculty of Informatics, Institute of Computational Science, Universita della Svizzera Italiana, Lugano, Switzerland

Petr Plecháč Department of Mathematical Sciences, University of Delaware, Newark, DE, USA

Giacomo Po Department of Mechanical and Aerospace Engineering, University of California Los Angeles, Los Angeles, CA, USA

Carlos A. Polanco Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Matej Praprotnik Laboratory for Molecular Modeling, National Institute of Chemistry, Ljubljana, Slovenia

Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

Zhao Qin Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Mohammad Javad Abdolhosseini Qomi Advanced Infrastructure Materials for Sustainability Laboratory (AIMS Lab), Department of Civil and Environmental Engineering, Henry Samueli School of Engineering, E4130 Engineering Gateway, University of California, Irvine, Irvine, CA, USA

Nima Rahbar Department of Civil and Environmental Engineering, Worcester Polytechnic Institute, Worcester, MA, USA

Abhinav S. Raman Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA, USA

Aravind Rammohan Science and Technology Division, Corning Incorporated, Corning, NY, USA

Michele Reticcioli University of Vienna, Vienna, Austria

Karsten Reuter Chair for Theoretical Chemistry and Catalysis Research Center, Technical University of Munich, Garching, Germany

James M. Rondinelli Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA

Jan Rossmeisl Department of Chemistry, University of Copenhagen, Copenhagen, Denmark

Carlo Andrea Rozzi Istituto Nanoscienze – CNR, Modena, Italy

Angel Sanchez-Gonzalez Facultad de Ciencias, Departamento de Química Orgánica, Universidad de Granada, Granada, Spain

Andrea E. Sand Department of Physics, University of Helsinki, Helsinki, Finland

Stefano Sanvito Center for Materials Genomics, Duke University, Durham, NC, USA

School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

William F. Schneider Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA

Ulrich S. Schwarz BioQuant-Center for Quantitative Biology and Institute of Theoretical Physics, Heidelberg University, Heidelberg, Germany

Daniel Schwen Fuel Modeling and Simulation Department, Idaho National Laboratory, Idaho Falls, ID, USA

Annabella Selloni Department of Chemistry, Princeton University, Princeton, NJ, USA

Hosung Seo Department of Physics, Ajou University, Suwon, Republic of Korea

Cheng Shang Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science (Ministry of Education), Department of Chemistry, Fudan University, Shanghai, China

Pradeep Sharma Department of Mechanical Engineering, Cullen College of Engineering, University of Houston, Houston, TX, USA

Jawaad Sheriff Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA

Alexander Shluger Department of Physics and Astronomy, University College London, London, UK

Michael P. Short Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Marvin J. Slepian Department of Medicine, Biomedical Engineering Department, University of Arizona, Tucson, AZ, USA

V. Sorkin Institute of High Performance Computing, A*STAR, Singapore, Singapore

Benjamin Spencer Fuel Modeling and Simulation Department, Idaho National Laboratory, Idaho Falls, ID, USA

Aleksandar T. Staykov International Institute for Carbon Neutral Energy Research (WPI-I2CNER), Kyushu University, Fukuoka, Japan

Ross Stewart Science and Technology Division, Corning Incorporated, Corning, NY, USA

Kurt Stokbro Synopsys QuantumWise, Copenhagen, Denmark

Haregewine Tadesse Science and Technology Division, Corning Incorporated, Corning, NY, USA

Leopold Talirz Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, Sion, Switzerland

Adama Tandia Science and Technology Division, Corning Incorporated, Corning, NY, USA

Anna Tarakanova Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Alexandre M. Tartakovsky Computational Mathematics Group, Pacific Northwest National Laboratory, Richland, WA, USA

Meenakshi Upadhyaya Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA

Dionisios G. Vlachos Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA

Aleksandra Vojvodic Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA, USA

George Volonakis Department of Materials, University of Oxford, Oxford, UK

Cai-Zhuang Wang Ames Laboratory – USDOE and Department of Physics and Astronomy, Iowa State University, Ames, IA, USA

Hao Wang Institute of Metal Research, Chinese Academy of Sciences, Shenyang, China

Hui Wang State Key Laboratory of Superhard Materials and Innovation Center of Computational Physics Methods and Software, College of Physics, Jilin University, Changchun, China

Jian-Jun Wang Department of Materials Science and Engineering, The Pennsylvania State University, Pennsylvania, PA, USA

Yanchao Wang State Key Laboratory of Superhard Materials and Innovation Center of Computational Physics Methods and Software, College of Physics, Jilin University, Changchun, China

Logan Ward Computation Institute, University of Chicago, Chicago, IL, USA

Su-Huai Wei Materials and Energy Division, Beijing Computational Science Research Center, Beijing, China

Gerhard R. Wittreich Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA

Christopher Wolverton Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA

Chung H. Woo The Hong Kong Polytechnic University, Hong Kong SAR, China

Martin Wörner Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Shunqing Wu Department of Physics, Xiamen University, Xiamen, China

Dongsheng S. Xu Institute of Metal Research, Chinese Academy of Sciences, Shenyang, China

Jing Yang Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Li Yang Department of Physics, Institute of Materials Science and Engineering, Washington University, St Louis, MO, USA

Rui Yang Institute of Metal Research, Chinese Academy of Sciences, Shenyang, China

Shengyou Yang Department of Mechanical Engineering, Cullen College of Engineering, University of Houston, Houston, TX, USA

Yang Yang Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Alireza Yazdani Division of Applied Mathematics, Brown University, Providence, RI, USA

Oleg V. Yazyev Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Jingjie Yeo Department of Biomedical Engineering, Tufts University, Medford, MA, USA

Laboratory for Atomistic and Molecular Mechanics (LAMM)/Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

Bilge Yildiz Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Wan-Jian Yin School of Energy and Soochow Institute for Energy and Materials InnovationS (SIEMIS), Soochow University, Suzhou, Jiangsu, China

Sidney Yip Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Mostafa Youssef Department of Mechanical Engineering, The American University in Cairo, New Cairo, Egypt

Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Sina Youssefian Department of Civil and Environmental Engineering, Worcester Polytechnic Institute, Worcester, MA, USA

Jinhu H. Zhang Institute of Metal Research, Chinese Academy of Sciences, Shenyang, China

Minghao Zhang Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA

Peng Zhang Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA

Xuan Zhang Center for Advanced Mechanics and Materials, Applied Mechanics Laboratory, Department of Engineering Mechanics, Tsinghua University, Beijing, China

Y. W. Zhang Institute of High Performance Computing, A*STAR, Singapore, Singapore

Xin Zhao Ames Laboratory and Department of Physics, Iowa State University, Ames, IA, USA

Ting Zhu Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

School of Material Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA

M. Žic School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Julian Y. Zuo Schlumberger Oilfield Services, Houston, TX, USA

Part I
Introduction

Applications of Materials Modeling and Simulation: An Introduction

1

Wanda Andreoni and Sidney Yip

Contents

Abstract

The ultimate goals of research on materials are to improve on their properties and control processes such as chemical degradation or mechanical failure as well as to design new materials and processing routes. The *Handbook of Materials Modeling*, in its former and present editions, aims to clarify how modeling and virtual experiments contribute to this effort and to monitor their development over the years. In particular, the set *Handbook of Materials Modeling: Applications – Current and Emerging Materials* (*ACE*) offers an ample description of recent achievements in several areas of materials science and technology. It covers research on systems – such as catalysts or magnetic materials or complex oxides – on which modeling and simulations have traditionally focused, and especially on a few "discovered" during the last 15 years or so, such as 2D materials or novel photovoltaics or bio-inspired hierarchical systems. The former studies can be considered to be an update of those reported in the 2005 edition

W. Andreoni (✉)
Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland
e-mail: wanda.andreoni@epfl.ch

S. Yip (✉)
Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: syip@mit.edu

of the *Handbook of Materials Modeling* (*HMM*). Therefore, by comparison, *ACE* provides the reader with a view of how computational materials science has developed since then. On the other hand, the articles on emergent materials show how critical the role of computations has been as an aid to experiment in tackling the (often unprecedented) related scientific and technological issues. Here we briefly introduce the 15 parts of *ACE* and emphasize the importance of topics treated therein and of the major findings. Several chapters provide a brief assessment of the applied methodologies by reporting on both achievements and failures. We shall underline what one can learn from these results and also point out connections between *ACE* and the companion handbook (*MTM*), which is devoted to the methods.

1   Introduction and Connection with *MTM*

Recent progress in technology has largely been related to the use of new materials of unprecedented complexity whose understanding appears to require the use of advanced modeling and simulation. Indeed, during the past decade or so, the realm of applications of computational materials science has broadened remarkably, their protocols have changed from the use of single to multiple methods for a more comprehensive understanding, and their objectives have gradually moved toward the requirements of rational design. Many are the challenges, however, that are still open so as to render the paradigms of virtual experiment essential companions to real experiment and thus help further progress in technology and fundamental understanding.

The *Handbook of Materials: Applications – Current and Emerging Materials* (*ACE*) gives an account of the various steps, successes, and failures in this endeavor to date. While we do not pretend that the scenario presented is complete, we are confident it gives a broad view of the diverse branches of materials science in which computations play a significant role.

ACE is the companion publication (set) to the *Handbook of Materials Modeling: Methods – Theory and Modeling* (*MTM*). Together these two stand-alone, multivolume publications constitute the *Handbook of Materials Modeling – Second Edition*. There is necessarily considerable overlap both amongst the parts in each set and amongst the individual chapters in each section. However this overlap is deliberate and intended to aid the reader. To access all the information available from the handbook, it would be advisable to examine the Table of Contents (*TOC*) for each set, along with consulting the present chapter (and its counterpart for *MTM*) for an overall introduction.

To aid readers relatively new to the handbook, we start by providing a brief description of the contents of the set *ACE*. As shown in the *Table of Contents*, the set consists of 15 parts and a total of 120 chapters. In analogy with the Introduction chapter to *MTM* (Table 1), we consider a list of keyword descriptors as a way of concisely characterizing the contents of *ACE*. This is shown in Table 1. Comparison

Table 1 List of keywords describing the topical themes of each of the 15 parts in *ACE*. For full section titles and the individual chapters in each section, see the *Table of Contents*

Plenary topics
Photovoltaics
Magnetism and spintronics
Low-dimensional materials
Nanoscale thermal transport
Electrochemical oxides
Surface catalysis
Nanomechanics
Hierarchical materials
Complex soft matter
Predictive modeling of glasses
Nuclear materials
Radiation damage
Human disease modeling
Crystal structure predictions

with the counterpart for *MTM* clarifies the commonalities and distinctions between the two sets of the handbook, if one simply notes the different contexts, namely frontier applications for *ACE* and fundamental methods for *MTM*.

Obviously method development is ahead of application: even when appreciation of new algorithms comes soon after their proposal, efficient implementation in computational codes is rarely fast, and a relevant application takes even longer. Therefore most of the results discussed in *ACE* can be considered a test of methods that are by now commonly used. *ACE* shows that computational materials scientists are setting increasingly more ambitious goals for their research that are often beyond the reach of available tools. On the other hand, *MTM* reports not only on methods that have been establishing themselves during the last decade but also on very recent highly promising developments. The latter clearly appear to be in line with the demand arising from the research work described in *ACE*.

2   A Brief Guide to the Parts of *ACE*

Here we will briefly introduce the topics addressed in *ACE* and highlight the relevance of each of them as well as methodological problems which seem to be common to diverse applications. For a more extensive and specific discussion, we refer the reader to the overviews by the section editors. In particular, they provide a global outlook of the current status of modeling and simulations and clarify both short- and long-term challenges.

Before focusing on each section separately, we note that, in each of them, particular attention is paid to the need to go beyond the understanding of a phenomenon obtained through the application of certain models and calculations and to translate it into suggestions for improvement or innovation. In fact, this drive toward the

practical use of the results of computational approaches has been developing over the last 10–15 years, certainly motivated by the increased and widespread need for "better" materials in current technologies. Simultaneously, computational scientists have become increasingly aware of the complexity of real materials. This implies, on the one hand, the need to acquire a global understanding of their behavior, e.g., concerning the interplay of chemical and mechanical properties in determining thermodynamic stability or concerning the performance of a given compound in real devices at operating conditions, and, on the other hand, the need to learn and make use of diverse theories and computational tools. Common concerns about the limitations of the latter can also be identified across the sections, e.g., about the lack of accuracy of standard approximations in DFT-based calculations, and the need for algorithms capable of bridging the description of a system at different length scales so as to unravel, for example, the influence that the interatomic processes have on the behavior at the mesoscale. Moreover, as we will see in the following, machine learning techniques have already started to be employed in the search for new materials in several areas.

In analogy with *MTM*, *ACE* opens with a collection of articles on subjects which we expect will be of special interest to the reader ("Plenary Topics," Editors: Andreoni and Yip). As shown in our section overview, some of these chapters can be easily connected to more than one of the following topical sections, thus providing further information and new perspectives. Others deal with additional very interesting topics that did not fit into the specific classification of Table 1.

The first topical part, "Photovoltaics: First Principles Modeling" (Editors: Bernardi and Grossman), is focused on ab initio approaches to the study of "modern" photovoltaics, namely, organic-inorganic halide perovskites (prototype $CH_3NH_3PbI_3$) and ultrathin group 6 metal chalcogenides such as WS_2. Both classes of materials have emerged in the last decade as highly promising alternatives to amorphous silicon (see, e.g., Editorial 2019). In an attempt to understand their characteristics and help design better performing devices, one is faced with a number of basic questions that cannot all be treated with current first-principle methods. These issues are pointed out in the overview by the section editors. However, as thoroughly discussed in the chapter by Filip, Volonakis, and Giustino for the case of perovskites (▶ Chap. 14, "Hybrid Halide Perovskites: Fundamental Theory and Materials Design") current ab initio methods have been successful in predicting – not just reproducing – several optoelectronic properties and have been of invaluable support to experimental research from the early days to date, also in the quest for more stable and nontoxic alternatives to lead compounds. As an example, we direct the reader to a related patent for the synthesis and utilization of nontoxic double perovskites, which were identified by the mentioned calculations (Snaith et al 2018).

The next part entitled "Modeling Applications to Magnetism, Magnetic Materials and Spintronics" (Editor: Sanvito) is well connected to the companion section in *MTM* (Editor: Sanvito) describing state-of-the-art theoretical and computational methods for the investigation of different aspects of magnetism and spintronics. The applications presented here involve the description of macroscopic magnetization

dynamics and of the interaction between electrical currents and spins at the atomic level. Moreover, a chapter is included on the search for novel magnetic materials using combined machine learning strategies and high-throughput data from electronic structure calculations. (For a general perspective, see the part on "Materials Informatics" in *MTM*.)

The part on "Low-dimensional Materials at the Nanoscale" (Editor: Chelikowsky) records successful applications of computations – mainly DFT-based – of the structural and electronic properties of 2D materials, nanoclusters, nanowires, and graphene nanoribbons. For example, these calculations clarify the interplay between limited dimensionality and quantum confinement and their influence on doping which is crucial for material functionalization. They also allow one to understand the evolution of physical properties with increasing system size and thus to predict how certain characteristics of the bulk such as the energy gap establish themselves.

The part "Thermal Transport" (Editor: Donadio) provides a broad view of the ongoing efforts to unravel multiple issues related to thermal transport in realistic systems and obtain reliable modeling tools and simulations. On the one hand, a few chapters describe in detail recent advances made in theoretical algorithms and computational schemes and point out the remaining challenges for method development. On the other hand, application of current approaches to a variety of systems, from polymers to metal alloys to nanostructured materials, clarifies the complexity of the problems that simulations should tackle in the near future.

The part "Challenges and Opportunities in Modeling Oxides for Energy and Information Devices" (Editors: Franchini and Yildiz) introduces the reader to a vast research arena, namely that of electrochemical materials, mainly complex metal oxides. The great interest in (complex) metal oxides stems from their versatile applications, as part of energy conversion and storage devices as well as of logic and memory systems. The articles presented in this section tackle fundamental issues that are critical for the performance of the mentioned devices. In particular, there are concerns mainly about the nature and role of point defects, surfaces, and interfaces, their effects on the ion or electron transport, and also about the limitations of currently available calculations to provide quantitative information.

We note that one of the chapters focuses on perovskite oxides and their use as electro-catalysts in fuel cells. In particular, it points out the lack of knowledge of the catalyst surface, for which modeling and simulations could be relevant. Electro-catalysis is actually one of the subjects more extensively treated in the next section, which is specifically dedicated to surface catalysis. In conclusion, we remark that, besides the use of traditional computational methods that are critically discussed in this section, also machine learning strategies have started to be applied for the optimization of the characteristics of electromagnetic oxides (see, e.g., Min et al. 2018).

The part on "Surface Catalysis" (Editors: Reuter and Metiu) also relates to a vast area of research. It suffices to realize that catalytic reactions are ubiquitous in our life and in industrial processes and that tremendous effort has been ongoing for decades in the search for "optimum" catalysts. The set of articles collected here

introduce relevant sub-topics and especially illustrate how current computational methods – from ab initio calculations to microkinetic models to computational fluid dynamics to multiscale approaches – have been applied during the last 10–15 years. In particular, the change of paradigm of modeling and simulations that we have mentioned above towards in silico design is clearly shown. Indeed, besides the mechanisms of specific chemical reactions and the nature of possible active sites and point defects, issues of major interest that have recently been tackled regard numerous factors that affect the performance and lifetime of the catalyst. For example, current investigation focuses on the interaction of the catalyst with its environment, e.g., heterointerfaces, other chemical species introduced in the reactor as either promoters or inhibitors of catalytic activity, and the solvent (in the case of electro-catalysis) or concentrates on the dynamical restructuring of the catalyst under operating conditions.

The huge amount of available data on catalytic processes calls for the use of artificial intelligence procedures in catalyst design. In addition to some references provided in the overview by the section editors to emerging applications of machine learning techniques, we direct the reader to an interesting discussion of recent progress in machine learning in the field of catalysis and on several severe challenges (Li et al 2018). We also suggest for further reading a recent general review on inverse design approaches to catalysis (Freeze et al 2019).

The next two sections illustrate how state-of-the-art modeling at various length scales can be applied to "everyday" materials with complex structures to study their fundamental properties. Understanding the connection of the mechanical properties to the structural characteristics is certainly a primary goal of this research. We note that dynamics simulations at the molecular or nanoscale level – namely, the first step in the chain of models leading to a more comprehensive representation of these systems – rely either on well-established coarse-grained schemes or on simple interatomic potentials. For methodological issues and an extensive perspective, we refer the reader to *MTM* and especially to the part "Soft Matter/Polymers Simulations" (Editor: Kremer).

The part "Multiscale Modeling of Structural Materials: Chemistry and Mechanical Performance" (Editors: Buehler and Martin-Martinez) mainly focuses on hierarchical natural materials such as silk, wood, and bamboo and on infrastructure composites such as asphalt and concrete. It also includes a chapter about silica aerogels, namely, nanostructures with an unusual structural organization. The part "Modeling the Structural Development and Mechanics of Complex Soft Materials" (Editors: Del Gado and Pellenq) is devoted to another fascinating class of "soft materials" and in particular reports on modeling and simulations for soft gels, cement, and clay. We remark that both the above sections also serve as an introduction to the rich phenomenology of the broad class of materials treated therein and to the achievements of experimental investigation.

Mechanical behavior is also the focus of the next part entitled "Nanomechanics of Materials: Structure and Deformation" (Editor: Zhu). Nanomechanics covers a large variety of nanoscale materials, such as nanowires, nanotubes, nanomembranes,

nanomotors, and low-dimensional materials in general. Size reduction has strong and peculiar effects on the mechanical properties of a material relative to those of the bulk and often leads to a remarkable enhancement in strength and toughness. Also unusual mechanisms have emerged as responsible for deformation and fracture. Some examples are given in the selected chapters. Most computational studies discussed therein are based on classical molecular dynamics. These simulations provide insights into, e.g., the nature and role of predominant defects or the mechanism leading to deformation in nano-twinned materials or the auxetic behavior of 2D materials. The limitations of this approach are also underlined, especially in terms of length and timescale of the simulations. For a more comprehensive understanding of "nanomechanics," we suggest further reading, e.g., the chapters included in the section "Recent Advances in Crystal Plasticity Modeling" (Editors: Cai and Ghosh) in *MTM* and also a recent review (Taloni et al. 2018) describing the failure behavior of materials at different size scales. We also point out a connection with the previous *ACE* section on "Low-dimensional Materials at the Nanoscale" (Editor: Chelikowsky) dealing with different aspects, mainly electronic properties, of nanostructures and 2D materials.

The section "Glass Science and Technology: Predictive Modeling" (Editors: Mauro and Onbasli) shows how constitutive models are commonly utilized for practical engineering purposes, with specific reference to the glass industry. This choice is certainly due to the ubiquitous presence of functional glasses in areas as diverse as construction, transport, automotive, or photovoltaic applications. Indeed modeling has proven to be crucial for the optimization and design of a large variety of functional properties of glasses. For example, the manufacturing of glasses is required to ensure unique stable attributes, such as high optical transparency, high mechanical strength, and high sound transmission loss. Moreover, as the extensive overview provided by the section editors points out, the current development of predictive schemes in the glass industry is moving toward a multiscale and physically based approach.

The next two parts, "Nuclear Materials" (Editors: Marian and Devanathan) and "Radiation Damage" (Editors: Short and Nordlund), are connected to a large extent, because, although distinct, they treat similar problems requiring the same computational approaches. The former reports on studies of the fundamental properties of nuclear fissile fuels, e.g., uranium dioxide, and structural materials, e.g., zirconium alloys, as well as on the detrimental effects of neutron irradiation on the performance and lifetime of nuclear reactors. The latter considers both neutron and ion irradiation effects and also materials other than nuclear. Both parts provide a comprehensive description of the various modeling strategies and simulations in use, from the atomic level to the continuum level. However, it is fair to say that, despite the advances in theory, modeling, and simulation, and experiments made over 50 years, formidable challenges still remain. For further specific discussions, readers should consult major reference works dedicated to this topic (Cacuci 2010; Konings 2012; IAEA 2018).

We also remark that damage induced by ionizing radiation in biological matter, say DNA or proteins, is not considered in this handbook.

The part "Multiscale Modeling of Diseases" (Editors: Dao and Karnidakis) provides an introduction to the emerging field of experimental and computational modeling of human diseases. It is a field of biophysical and biomedical research driven principally during the past 10 years by the developments along three directions, advanced multiscale modeling capabilities, new mechanobiology experimental techniques, and interdisciplinary collaboration among modeling scientists, experimentalists, and clinicians in the targeted specialties. The chapters discuss various computational methods developed for modeling at the meso-scale (particle-based) and continuum levels and specific studies on the human red blood cell (relevant to malaria) and sickle cell anemia.

The *ACE* set of the handbook closes with a section entitled "Computational Crystal Structure: Prediction and Materials Discovery" (Editors: Wang and Wolverton) providing another interesting topic of materials development. Applications of a variety of methods for the prediction of crystalline structures are presented for materials of given chemical compositions. The development of the basic algorithms, based on the exploration of the potential energy surface and energy minimization procedures, dates back more than 20 years. However, several new and more accurate procedures have been put in place only in the last decade so as to increase the ability to predict low-energy structures for complex compounds and discover unforeseen polymorphs. In particular, we emphasize extensions and adaptations of the genetic algorithm and particle swarm optimization techniques. Moreover, we point out the advent of an original methodology that combines results obtained both from the analysis of X-ray diffraction (Rietveld refinement) and from ab initio (DFT) calculations.

3   Conclusion

ACE provides an overview of several areas of materials science in which computations at different levels of sophistication have been extensively applied during the last 10–15 years. The role of computational materials science is now established.

However, it is clear that further progress is needed for physically based modeling and simulation to become essential components in the design of either new complex materials with prescribed properties or new procedures for material process control and synthesis. To meet these ambitious objectives, an intensive exchange between academia and industry and society will be necessary.

While we may very possibly have overlooked some topics of interest to the reader, we trust that he/she will find other very interesting and timely articles and particularly instructive perspectives on foreseeable future applications of computational methods. Indeed, after comparing the needs for more robust and far-reaching algorithms as clarified in *ACE* and the ongoing as well as expected method developments discussed in *MTM*, we can safely conclude by repeating Leibniz's famous sentence: "The present is pregnant with the future."

References

Cacuci DG (ed) (2010) Handbook of nuclear engineering. Nuclear engineering fundamentals, vol. 1. Springer Science+Business Media, New York

Editorial (2019) A decade of perovskite photovoltaics. Nat Energy 4:1

Freeze JG, Kelly HR, Batista VS (2019) Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. Chem Rev 119:6595

IAEA (International Atomic Energy Agency) (2018) Accelerator simulation and theoretical modelling of radiation effects in structural materials. Nuclear Energy Series No. NF-T-2.2, Vienna

Konings RJM (ed) (2012) Comprehensive nuclear materials. Elsevier, Amsterdam

Li Z, Wang S, Xin H (2018) Toward artificial intelligence in catalysis. Nat Catal 1:641

Min K, Choi B, Park K, Cho E (2018) Machine learning assisted optimization of electrochemical properties for Ni-rich cathode materials. Sci Rep 8:15778

Snaith HJ, Hagighirad AA, Giustino F, Filip MR, Volonakis G (2018) Double perovskite. US Patent Appl. 20180290897

Taloni A, Vodret M, Costantini G, Zapperi S (2018) Size effects on the fracture of microscale and nanoscale materials. Nat Rev Mater 3:211

Part II
Plenary Topics

Plenary Topics: An Introduction

2

Wanda Andreoni and Sidney Yip

Contents

Abstract

This section is meant to complement the already extensive collection of articles classified in the following topical sections. In particular, additional applications of computational methods are discussed, which we believe will be of general interest to readers. They focus on different materials – ranging from graphene to titanium oxide to novel superconductors to cement to titanium alloys to glassy oxides and polymers – and/or critical material processes such as defect diffusion and fracture propagation, leading to degradation and failure. We will point out that some chapters can be related to several others across the sections of this part of the handbook (*ACE*) and add new information and perspectives. Connections with those of the companion set (*MTM*) will also be pointed out.

W. Andreoni (✉)
Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland
e-mail: wanda.andreoni@epfl.ch

S. Yip (✉)
Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: syip@mit.edu

1   Introduction

This is a brief overview of contributions collected in a special *ACE* section that – as its counterpart in *MTM* – is intended to complement the others, each focused on one specific theme, and also offer additional perspectives.

Most of the plenary topics, it is hoped, will be of general interest to the reader. For example, some chapters deal with materials, such as titanium alloys, cement, titania, and graphene, which are of outstanding relevance for a variety of applications. In particular, the importance of graphene has emerged during the last 15 years (see, e.g., Geim and Novoselov 2007) and thus was not part of the first issue of this handbook. Others focus on topics that did not fit into one of the following sections, for example, superconducting materials. Also in this case, reliable computational approaches have become available only in the last decade or so.

2   Brief Chapter Overviews

This section opens with Payne's view of materials modeling in/for industry, in particular in the UK (Payne, ▶ Chap. 3, "The Industrial Impact of Materials Modelling"). A report is given of recent initiatives in Europe to connect the academic community of computational physicists in the field of materials science to industrial partners. The problems responsible for the fact that materials modeling has so far played a minor role (if any) in industry are clarified. Nevertheless its use has increased over the years at all length and time scales. We refer the reader to a chapter in this same section (Xu et al., ▶ Chap. 7, "Titanium Alloys: From Properties Prediction to Performance Optimization") for a thorough description of how several different approaches can be relevant to the various stages of product design in industry. Payne also provides a thoughtful discussion of the needs that industries will nevertheless continue to have for modeling results, (acquisition of) expertise, and/or tools, depending, in particular, on the size of the company.

One should remark that many academy-industry collaborations exist worldwide – although not systematically – and also that research centers at IBM and at chemical industries such as BASF have constantly supported the development of advanced computational methods as well as their application in various domains of materials science, thus generating several patents (for the impact of ab initio simulations see, e.g., Andreoni et al. 2006, 2012, Afyfy et al. 2013, Laino and Weber 2019).

The two following chapters discuss materials of primary importance in contemporary technology: titanium dioxide (titania) and graphene. In both cases, several examples show how computer simulations have been able to complement and aid in understanding and also predicting the results of experimental investigation.

The chapter on titanium oxide (Selloni, ▶ Chap. 4, "Titania and Its Outstanding Properties: Insights from First Principles Calculations") illustrates how a series of ab initio calculations have been able to elucidate the complex nature of TiO_2 surfaces. For example, their physical properties and chemical reactivity depend on

the specific polymorph, on the specific surface and its structure, on the unavoidable presence of point defects such as oxygen vacancies, and on the exposure to water. It is also clear that the ab initio approach to different situations is not "standard" but special attention must be paid to the methodology to be applied. Indeed real-world applications primarily refer to reduced TiO_2. Thus, modeling imposes the need to go beyond the generalized gradient approximation (GGA) to the DFT exchange-correlation functional so as to reduce its intrinsic self-interaction error. Therefore PBE-DFT was augmented either with a (semiempirical) Hubbard (U) term or with a fraction of exact-exchange term as in the hybrid PBE0 exchange-correlation functional.

The chapter on graphene (Yazyev, ▶ Chap. 5, "Modeling Disordered and Nanostructured Graphene") focuses on the modeling of its peculiar characteristics and the calculation of their effects on a variety of physical properties that are crucial for any of the diverse applications of this material. In particular, an outstanding achievement of computations was the prediction of the structure of topological defects (dislocations and grain boundaries) prior to experiment that has confirmed those results. Key ingredients were the construction of ingenious structural models and subsequent ab initio calculations for the investigation of the electronic properties as well as for the determination of the most energetically favorable defect configurations. Calculations have also revealed the dependence of the electronic properties of nanostructured graphene on the nature of the edges, i.e., armchair or zigzag or chiral. We note that a complementary account of computational studies – also linked to experiment – of graphene nanoribbons can be found in ▶ Chap. 31, "Electronic Structure of Atomically Precise Graphene Nanoribbons," by Talirz and Pignedoli.

One of the greatest challenges for electronic structure calculations has long been that of approaching the physics of superconductors and determining in particular the critical temperature T_c. This "dream" has become reality – within the framework of density functional theory – only recently, after the seminal paper by Lüders et al. (2005). Boeri, ▶ Chap. 6, "Understanding Novel Superconductors with Ab Initio Calculations," guides the reader through the main steps leading to this achievement and subsequent successes of theory and computations for conventional superconductors. State-of-the-art calculations are not only able to contribute to the understanding of mechanisms leading to superconductivity but also to quantitatively characterize superconductive phases and predict the transition temperature. On the other hand, the case of unconventional superconductors like the high-T_c cuprates is not approachable yet with available methods.

The next chapter (Xu et al., ▶ Chap. 7, "Titanium Alloys: From Properties Prediction to Performance Optimization") presents an exhaustive and detailed review of the application of computational tools to the design and optimization of titanium-based alloys. The specific role of different methods at different length scales in this complex process is elucidated. They comprise electronic structure calculations and atomistic simulations for an early screening of fundamental material properties, mesoscale approaches for an understanding of microstructure evolution under specific processing and operational conditions, as well as finite element methods

for the simulation of various manufacturing processes and the final optimization of the material performance. Not only successful applications of the above protocol are described, but also difficulties are pointed out as well as important challenges for future progress. In particular, the aid of artificial intelligence strategies based on "big data" is called for. Connections of this chapter are evident with the *MTM* set of this handbook covering all the above-mentioned methodologies. Moreover, we remark that the example of titanium alloys is extremely interesting in view of their relevance in diverse industries and applications. Namely, depending on their composition, they are used in aerospace, marine, and biomedical applications.

The following articles are concerned with the study of specific and critical material processes via modeling at multiple scales.

The next chapter (Kamrin, ▶ Chap. 8, "Quantitative Rheological Model for Granular Materials: The Importance of Particle Size") deals with a class of complex systems – dense granular flows – whose understanding is still limited. This chapter provides an overview of granular material modeling at diverse levels and especially an insightful analysis of the constitutive behavior of dense granular matter through the development of a nonlocal continuum model of fluidity (the nonlocal granular fluidity (NGF) model). Several applications demonstrate, in particular, the importance of grain size effects on the flow properties. We remark that dense granular flows are common in countless natural processes (e.g., rock avalanches) and industrial applications (e.g., in the pharmaceutical and agricultural sectors). In the effort to describe, predict, and thus control their behavior, modeling and numerical approaches play a special role.

The section proceeds with a chapter (Masoero, ▶ Chap. 9, "Mesoscale Mechanisms of Cement Hydration: BNG Model and Particle Simulations") focused on cement – the most widely used man-made material in the world and in particular on the formation of its fundamental constitutive unit, calcium-silicate-hydrate (C-S-H). Linking the kinetics of the evolution of this process to the chemical (microscopic) characteristics of the original compounds, and eventually to the structural properties and mechanical behavior of the material, is one of the most critical and still largely open issues. This part of the handbook (*ACE*) dedicates specific chapters to the physical, chemical, and mechanical properties of cement, describing the results of computational methods and their performance relative to experiment (Ioannidou, ▶ Chaps. 71, "Mesoscale Structure and Mechanics of C-S-H"; Qomi et al., and ▶ 72, "Nanoscale Composition-Texture-Property Relation in Calcium-Silicate-Hydrates"). Masoero's plenary topic article can be considered a complement to those chapters. It deals with a relatively simple and very instructive scheme, the model of boundary nucleation and growth (BNG). It is an attempt to provide an approximate description of the early controlling step of hydration that is physically based and amenable to quantitative analysis. In particular, it shows how kinetic Monte Carlo simulations of nanoparticle aggregation can be exploited to improve on the parameters governing the BNG model and especially to provide an insight into their physical meaning.

The next chapter (Ciccotti and George, ▶ Chap. 10, "In Situ AFM Investigations and Fracture Mechanics Modeling of Slow Fracture Propagation in Oxide and Polymer Glasses") constitutes an exception because it mainly describes experimental

techniques and observations. It offers an insightful account of how slow fracture propagation can be quantitatively studied using the combination of in situ atomic force microscopy and phenomenological multiscale modeling. While it is focused on oxide and polymer glasses, the results discussed therein have significant and general implications for all time-dependent deformation processes in brittle materials. This detailed and extensive description of the experimental results should be seen as a source of inspiration for the development of computational methods for the simulation of fracture propagation and also other multiscale phenomena. We remark that the use of inserts to explain various key concepts in the subject matter offline from the main text is effective and it is hoped will be helpful to the reader.

Extended defects such as dislocations are ubiquitous in real crystalline materials and influence a variety of physical properties. Diverse examples of modeling and multiscale simulations are currently helping to understand, e.g., the origin of a dislocation or the way it propagates and affects the mechanical properties of a given material (see, e.g., the *MTM* section on "Crystal Plasticity: From the Atomic Scale to the Macroscale" – Editors: Cai and Ghosh). The next chapter (Fan and Cao, ▶ Chap. 11, "Long Time-Scale Atomistic Modeling and Simulation of Deformation and Flow in Solids") is concerned with the onset and propagation of dislocation-driven deformation and flow in crystals (specifically bcc metals) and glasses. These are processes that involve time scales well beyond the reach of classical molecular dynamics of atomistic models. In the *MTM* set of this handbook, we have presented methods allowing to access longer time scales, both those using direct acceleration techniques (see the section "Long Time-Scale Simulation Methods," Editors: Perez and Uberuaga) and those circumventing the time constraints with the aid of advanced enhanced sampling approaches (see the section "Atomistic Simulations," Editors: Andreoni and Yip). Application of the latter is not possible yet, given the complexity and size scale of the systems under investigation. A simpler but efficient alternative is applied instead, the autonomous basin climbing (ABC) algorithm, which allows one to explore and characterize the potential energy surface of a given system by coupling minimum energy pathway methods, kinetic Monte Carlo and transition state theory. We remark that the ABC method was originally employed to investigate viscous flows and the origin of the glass transition in supercooled liquids (Kushima et al. 2009). Note also that related issues are discussed in one of the chapters of the "Plenary Topics" part in *MTM* (Rizzo, Chap. 9, "Critical Phenomena in Glasses").

Dislocations are known to influence not only the mechanical response of a given material but also many other characteristics and functionalities. Understanding and predicting the effects of a dislocation on, e.g., electronic, optical, magnetic, or thermoelectric properties requires a quantum-based approach. To this end, a completely new full quantum theoretical framework has very recently been introduced by Li et al. (in a series of articles starting in 2017) and constitutes the subject of the last chapter of this section (Li et al., ▶ Chap. 12, "Quantized Dislocations for Functional and Quantum Materials"). In this theory a dislocation is quantized ("dislon" is the name of the basic quantum). Starting from a 1D approach to describe quantized single dislocations, these authors have shown that an extended (3D) scheme can be used to describe the electron and phonon interactions with multiple dislocations.

Several examples are presented. Benefits of this framework are claimed also for explicit calculations. Moreover, other implications can be envisaged.

3   Conclusions

A plenary topic is expected to be of general interest to a varied audience of readers. We think that each chapter of this section deserves to be classified in this way. In fact, for example, as we have pointed out, most of the chapters can be related to others across the parts of *MTM* or *ACE* and/or refer to themes (materials or processes) of very particular relevance.

The main messages arising from a careful reading concern, on the one hand, the value that state-of-the-art modeling and simulations have in tackling complex problems related to real materials and, on the other hand, the need to extend their reach through the development of multiscale methods and their integration in the design and optimization process.

In this respect, we underline Payne's contribution that is explicitly concerned with the possible impact of materials modeling in the realm of industry. Beyond that, we wish to call attention to the existence of a natural, cultural gap between the scientific community in the role of "producers" of modeling and simulation tools and society at large in the role of "users" of the science and technology. Bridging this gap is a long-standing challenge to both sides. The most promising solution is bound to involve hands-on collaboration among all the stakeholders. Still, we can conclude that – as illustrated also in this section – in the last decade, many opportunities have flourished for computational materials science to contribute at least to the understanding of problems of relevance to technology and societal needs.

References

Afyfy N, Andreoni W, Curioni A, Khomyakov P, Kim J, Sadana DK. Photovoltaic device with band-stop filter, US Patent Application Publication Number 2013-0312828, published on November 28, 2013

Andreoni W, Curioni A, Shevlin SA (2006) Dielectric materials. US Patent 7,057,244, issued on June 6, 2006

Andreoni W, Callegari AC, Cartier EA, Curioni A, D'Emic CP, Gousev E, Gribelyuk MA, Jamison PC, Jammy R, Lacey DL, McFeely FR, Narayanan V, Pignedoli CA, Shepard JF Jr, Zafar S (2012) Method of forming metal/high-κ gate stacks with high mobility. US Patent 8,153,514, issued on April 10, 2012

Geim AV, Novoselov KS (2007) The rise of graphene. Nature Mater 6:183

Kushima A, Lin X, Li J, Eapen J, Mauro JC, Qian X, Diep P, Yip S (2009) Computing the viscosity of supercooled liquids. J Chem Phys 130:224504

Laino T, Weber V (2019) Sodium ion solid-state conductors with sodium oxoferrate structure. US Patent Number 10,170,790, issued on January 1, 2019

Lüders M, Marques MAL, Lathiotakis NN, Floris A, Fast L, Continenza A, Massidda S, Gross EKU (2005) Ab-initio theory of superconductivity – I: density functional formalism and approximate functionals. Phys Rev B 72:024545

The Industrial Impact of Materials Modelling

3

Mike Payne

Contents

Abstract

It has been estimated that \$1Billion per year is spent on academic research on materials modelling. At the same time, there is increasing pressure on academics to realize and/or demonstrate the socioeconomic impact of their research. Naively, in attempting to reconcile these two facts, one might assume that there was a demonstrable economic impact of materials modelling amounting to multiple \$Billions per year. However, as will be discussed here, the situation is rather more complex than this naïve picture suggests and there are significant challenges to realizing the benefits of materials modelling.

1 Introduction

Academics tend to focus on the impact of their research in terms of the quantity of papers and the quality of the journals they publish in and the number of citations their work receives. In terms of the first metric, the annual output of papers based

M. Payne (✉)
Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK
e-mail: mcp1@cam.ac.uk

on Density Functional Theory (DFT) is so large that it has to be measured in kilopapers per year. In terms of the last metric, as Professor Nicola Marzari pointed out in 2013, DFT and electronic structure also reigns supreme with 18 the top 20 highest cited papers published by the American Physical Society. Therefore, there is little doubt about the academic impact of this field. However, here we focus on the economic impact of materials modelling, and so, from this point on in this chapter, "impact" will be used to refer exclusively to economic impact. Professor Marzari has estimated the global cost of academic research in materials involving DFT to be of the order of $1 Billion per year. This is based on an estimate of 10,000 for the number of DFT papers published per year which focus on materials and an estimate of $100,000 for the cost of the staff and computational resources required to generate each paper. If this is the case, then one would have expected that there would be a demonstrable economic impact of DFT modelling of many times this sum.

There are, of course, many simple and routinely used methods that academics have perfected to distract any detailed scrutiny of the wider impact of their work. One of the most common ones is to simply assert that the return on investment (RoI) on money spent on academic research is 19, sometimes this is even presented as the return on investment each year. It is a commonly quoted figure, and the methodology and, indeed, the time at which this figure was calculated are somewhat unclear. Nevertheless, this figure is widely used to justify investment in academic research. Notwithstanding any concerns about the figure itself, the obvious issue is that this result cannot extend to all levels of investment in academic research. If a country decides to spend its entire GDP on research 1 year, I find it doubtful that the research will generate 19 times that sum in total future economic benefits, let alone every year. If this were the case, then all the world's economic problems could be trivially solved by investing vast sums of money in academic research.

Over the last two decades, the UK has become increasingly focussed on the socioeconomic benefits of the research it funds. Initially, in the case of research grant applications to UK Research Councils, this involved little more than asking academics to consider whether their work might have impact. This question was usually addressed by including a completely unjustified statement about the £millions of economic benefit that the research would generate. Over the years, this requirement to consider impact has become more demanding and now every grant application to UK Research Councils has to include a (maximum) two page "Pathways to Impact" document explaining how the impact of the research will be realized. However, there are a significant number of weaknesses in the current UK system. For a start, the vast majority of these grant applications are refereed by other UK academics who may have limited, if any, experience of generating socioeconomic impact. However, the most significant flaw in the system is that it is entirely focused on the "input" side – the grant application – and there is currently no monitoring to see whether the claimed long-term socioeconomic impact has been achieved, let alone applying any sanctions on people who serially fail to deliver their promised socioeconomic benefit. Perhaps not surprisingly, there is a very varied level of engagement in generating impact within the UK research community.

The solution to the challenge of generating impact is, of course, trivial. It can be achieved simply by focussing on the "output" side of research and rewarding researchers who genuinely generate impact. If the easiest way to get the funding for a postdoctoral researcher is by generating impact, then I can guarantee that virtually every academic would focus on this challenge. If other countries wish to follow UK down this direction of expecting socioeconomic benefit from academic research, they are advised to create an output-based mechanism for ensuring this.

The importance of impact in the UK has now extended considerably beyond research grant applications. About once every 7 years, there is an assessment of the quality of all UK university research. Initially, if somewhat unimaginatively, this process was called the Research Assessment Exercise (RAE). During this era, it focused almost exclusively on academic measures of the quality of research. The RAE has now been replaced by the Research Excellence Framework (REF), and the first operation of REF took place in 2014. Under REF, the socioeconomic impact of research determines how £200 million of annual government funding is distributed to English universities by HEFCE (Higher Education Funding Council for England – similar mechanisms operate in the rest of the UK). The mechanism for evaluating impact under REF is through so-called Impact Case Studies, examples of which can be found at http://impact.ref.ac.uk/CaseStudies/. In addition to directly rewarding successful examples of impact, which is laudable, it should be noted that the last REF exercise only required one Impact Case Study for every ten academics entered into REF – thus recognizing that not every academic could be expected to generate socioeconomic benefit from their work during every REF reporting period. REF demanded high levels of auditable evidence of socioeconomic impact. For instance – though it does not relate to the current subject area – if a historian had made or contributed to a major television program – which one might have thought would have been an excellent example of social impact – an Impact Case Study would only have been admissible if there was auditable data for the size of the audience. Thus, many examples of academic research that have had significant socioeconomic benefit are not eligible for REF, and it is easy to envisage such examples in the field of materials modelling. If someone from industry finds the solution to a problem in the scientific literature, it is very unlikely that the consequent economic benefit would be associated with the original academic research as, in most cases, the academic would not be aware of this benefit.

Another example of impact playing a larger role in the funding of UK universities is provided by the change over the years in the mechanism for providing government funding for infrastructure. In the past, large amounts of infrastructure funding, say for buildings or major research facilities, were provided based on purely academic arguments. Now these schemes have been replaced by the UK Research Partnership Investment Fund (http://www.hefce.ac.uk/rsrch/ukrpif/). The UKRPIF provides significant investments into UK universities but only *after* those universities have brought in at least twice that amount of money in industrial (or other non-government) funding.

While these examples are taken exclusively from the UK, it does seem that discussions about the importance of generating socioeconomic impact from publicly

funded research are becoming widespread. The materials modelling community has, at least in part, justified its research funding for decades on the basis that this was an investment that would give us the ability to design materials which would obviously provide significant impact. However, while many industries have moved to "virtual" (i.e., entirely computational) design for all or some of their operations, the materials community is faced with the reality that virtual materials design is a long way from becoming the entirety of the product design cycle. This is not to say that the various materials modelling methods described in this handbook will not have a role in virtual materials design. It is just a statement of the fact that, whatever information we can determine about a material, we cannot currently guarantee, based on modelling alone, that the material will not deteriorate or fail over the lifetime of a product. As a community, we know that materials are complex on many levels, in their compositions, microstructures, and properties, and that materials properties range over enormous ranges of lenthscales and timescales. It is not clear that funders will continue to be sympathetic to these issues and may increasingly ask our community the question "If I can design a car or an airplane on the computer and go away and make it without creating a prototype, why can't I design a material on the computer and then go away and make it in exactly the same way." The first step of this process, namely, designing a material on the computer, has been an active research field for many years with numerous examples, such as the numerous predictions of superhard materials (see for instance, Mattesini and Matar 2001). These are materials that have elastic moduli that are larger than those of diamond. However, none of the predicted superhard materials have been developed into real-world products. There can be many reasons for this. One common problem is the failure to fabricate a material with the same composition and structure as the theoretically designed material. In the case of superhard materials, there may be failure mechanisms operating on larger lengthscales that cause the material to fail under loads that are smaller than the ones that can be accommodated by diamond so that the material is harder than diamond on the atomic scale but weaker on longer lengthscales. While we may struggle to point to examples of purely computational design of materials, materials modelling methods have been used by many companies for decades, so, presumably, there must be some benefit gained from their use. It is, however, quite difficult to determine this benefit as companies tend not to be very forthcoming about their use of materials modelling methods. Swenson et al. researched this issue for a report entitled "Modeling and Simulation: The Return on Investment in Materials Science" published by International Data Corporation (IDC) in 2004. More recently, Gerhard Goldbeck of Goldbeck Consulting has published a series of extensive reports on this topic (Goldbeck 2012, 2014; Goldbeck and Court 2016) and readers who wish to gain a deep understanding of the impact of materials modelling should read these reports. The remainder of this chapter will provide a summary of the most important findings presented in these reports interspersed with the author's own experiences and other people's accounts of their experiences of the impact of materials modelling – a number of these were covered in an E-CAM Industry Workshop on materials design "From the Atom to the Material".

2   Industrial Requirements for Materials Modelling

The first thing that needs to be appreciated by any academic who wishes to engage with industry is the enormous variation in expertise, work practices, and requirements for materials modelling from one company to the next. At the E-CAM Workshop, this was summed up by the phrase "There is no such thing as industry!" The talks and comments from industry-based participants reflected this broad range of requirements from industry for materials modelling, in terms of types of material, properties of interest, and, interestingly, the level of accuracy required. This ranged from one extreme of a requirement for quantitatively accurate predictive simulations to, in the middle, not being worried about the absolute value of a predicted quantity but only about trends under change of composition to the other extreme of only using modelling and simulation to provide insight into complex physical problems without requiring quantitative accuracy. It became apparent that in most cases, even in large companies, materials modelling is a very small-scale activity at present, and in many companies, it does not exist at all. It is also not unusual to have no company-wide view of modelling activities, so that there is no information exchange between different divisions of the same organization. A further important conclusion from the E-CAM Workshop was that the value of materials modelling was often only realized when these tools were integrated into the entire product design process. Materials operate in systems and any benefit in a material property is only realized in the real world if it can be incorporated into a product. Hence, the one story universally heard from industry is that materials modelling plays a role in product development but only a role – and the failure to address the rest of the ecosystem will usually prevent successful deployment of a material. This was clearly demonstrated in Ford's Virtual Aluminum Casting project (Allison et al. 2006) and it is referred to as Integrated Computational Materials Engineering (ICME, Pollack et al. 2008). Goldbeck (2016) provides further examples of ICME generating significant RoI including:

GE Aviation utilized an ICME approach to achieve a reduction in the amount of rhenium (a rare and expensive element) in superalloys for aircraft engine turbine airfoil components which resulted in the introduction of a new alloy in 2 years rather than the typical 6 years historically required.
QuesTek Innovations led a project that resulted in the development of the corrosion-resistant Ferrium S53 advanced high-strength steel alloy for landing gear and other applications. The ICME approach led to significant reductions in alloy development time and an estimated development cost savings of nearly $50 M.

ICME is something that academics cannot do alone, and, hence, there is a significant challenge to making a real economic impact from academic research unless it leaves the academic sphere. The level of investment needed to implement ICME represents a barrier to adoption which can make it extremely difficult to persuade a company of the potential benefits of materials modelling. Furthermore, to be incorporated into ICME, modelling methods need to be easy to use and

not dependent on high levels of expertise and experience. It should be noted that the VAC project outsourced a significant fraction of its materials modelling requirements to academia and this does provide a mechanism for circumventing this problem of accessibility. Interestingly, Goldbeck (2014) identifies a similar phenomenon among the network of Psi-k researchers (http://psi-k.net/), many of whom were found to have industry connections.

The various impact reports point to the high level of trust and, hence, widespread adoption of density functional theory methods by industry compared to the lower level of trust and, hence, more restricted adoption of empirical and coarse-grained methods. Golbeck (2016) emphasizes industry demands for improved guidance to help assess the accuracy of any computational method for a particular materials problem. As discussed earlier, industries may not be too worried about a high level of accuracy in any particular case but do need guidance about the expected level of error in any prediction of a materials property.

3 The Role of the Translator and the Level of Materials Modelling Activity

Many of us will know people or may even be the person who has an uncanny ability to relate a complex materials process to accessible atomic scale calculations. The E-CAM Industry Workshop highlighted their importance as every example of materials modelling significantly driving forward progress in an industrial process involved such an individual. Indeed, software vendors in the materials space probably spend as much time carrying out this task as developing their software. Given the importance of this task, it is encouraging that the European Materials Modelling Council has recognized this role and given it the title of "translator" – the individual who can connect (in both directions) a complex product challenge to a set of feasible materials modelling tasks. There is also direct numerical evidence in the IDC report that increasing levels of experience and expertise produce a greater level of impact of materials modelling – essentially, the more materials modelling a company carries out, the higher the rate of return on investment. Disappointingly, it seems that neither the IDC report nor, in many cases, a company's own experience of the value of materials modelling have led to increases in the amount of materials modelling a company does.

4 What Materials Modelling Can Do for Industry

Once one moves away from the academic aspiration of designing materials entirely on the computer to the real-world industrial use of materials modelling, perhaps as a component of ICME, then there are many ways that modelling can add value to a project. These include:

More efficient experimentation – for instance, using modelling to determine which direction to move in composition space to create a material with better properties.

Broader exploration and deeper understanding – in the case where modelling is cheaper than experiment and accurate enough using modelling to perform a more extensive parameter search than could have been afforded using experiment alone.

Saving a product development project and/or accelerated product development – for instance, modelling providing a solution to a problem that had not been solved using the tools previously available for development.

Risk management through safety testing.

It should be noted that none of these ways involve the entire design of a material but instead are contributions that modelling can make to the materials design process. In their IDC report, Swenson et al. developed quantitative models to illustrate the potential RoI of each scenario. These models were reanalyzed with updated financial figures in Goldbeck (2012). These analyses suggest a more credible value for the RoI for materials modelling in the range of 3 to 9 – with the higher value associated with a higher use of modelling suggesting that the more resource a company invests in materials modelling, the higher their rate of return on this investment.

5 Materials Modelling Marketplaces

The traditional model of operation of software companies creates a significant barrier to adoption of materials modelling methods, particularly for small- and medium-sized companies – usually referred to as "small and medium enterprises" (SMEs). The traditional model for software companies is selling licenses for software products which allow the software to be installed and used on the client's hardware. This model works well in the case of large companies who have suitable computer hardware and IT expertise to support the necessary hardware and software. However, SMEs often do not possess large computational resources or significant levels of IT support, In this case, a move into materials modelling may require substantial investment in computer hardware and increased levels of IT support in addition to the cost of the software license. Without an absolute guarantee of a successful outcome from the materials modelling and a consequent economic benefit to the company, one can understand why many companies are cautious about setting up a substantial materials modelling effort. This concern could be allayed if there were more information generally available about the successes of materials modelling, but there is an understandable reluctance for companies to reveal such information as their competitive advantage gained from the modelling could be revealed to their competitors. Sometimes this information is eventually divulged but usually only after a significant period of time so that the competitive advantage has elapsed and disclosure no longer risks the competitive advantage. Furthermore, the cost of software licenses is predicated on the assumption that they

are used continuously. For many SMEs, this is not a realistic model as it is likely that only a small fraction of staff time could be expended on materials modelling. The high start-up cost represents a significant barrier to entry for materials modelling. In various other market sectors, this issue has been addressed by moving to a "Cloud" model of computing and a charge per use approach to software – usually referred to as "Software as a Service." The US nanoHUB (https://nanohub.org/) provides an excellent example of a Cloud-based materials modelling service which provides hundreds of modelling tools and has over a million users who do not have to pay to use the resources provided by the site – even the computational resources needed to run the software. This clearly limits the range of software provided as this model precludes the participation of materials modelling software vendors and hence the lack of particular tools and techniques. Interestingly, the European Materials Modelling Council (EMMC https://emmc.info/) is creating a materials modelling marketplace that, in principle, could address this problem as it would allow any user to access both free and commercial modelling tools and purchase the necessary computational resources to for the calculations they wish to carry out – paying on a per-use basis for the actual work done rather than requiring a large initial investment. It will be interesting to see how this initiative develops in the longer term. I hope that it will lead to a much wider adoption of the "Software as a Service" model for materials modelling which will be crucial for the materials modelling community to see widespread use of its methods across industry.

References

Allison J, Li M, Wolverton C, Su X (2006) Virtual aluminum castings: an industrial application of ICME. JOM 58(11):28–35
Goldbeck G (2012) The economic impact of molecular modelling of chemicals and materials. https://gerhardgoldbeck.files.wordpress.com/2014/01/the-economic-impact-of-modelling.pdf
Goldbeck G (2014) Industry interactions of the electronic structure research community in Europe. https://gerhardgoldbeck.files.wordpress.com/2014/02/psik-industry-interactions-report1.pdf
Goldbeck G, Court C (2016) The economic impact of materials modelling. https://zenodo.org/record/44780#.Wdkc4Hbrv4g
Mattesini M, Matar SF (2001) Int J Inorg Mat 3(7):943–957
Pollock TM, Allison JE, Backman DG, Boyce MC, Gersh M, Holm EA, LeSar R, Long M, Powell IC IV, Schirra JJ, Whitis DD, Woodward C (2008) Integrated computational materials engineering: a transformational discipline for improved competitiveness and National Security. The National Academies Press, Washington DC
Swenson M, Languell M, Golden J (2004) Modeling and simulation: the return on investment in materials science. IDC White Paper, June 2004

Titania and Its Outstanding Properties: Insights from First Principles Calculations

4

Annabella Selloni

Contents

Abstract

Titanium dioxide (TiO$_2$) is one of the most widely used photocatalytic materials due its abundance, non-toxicity, and high stability in different environments and conditions. TiO$_2$ is also considered a prototypical metal oxide and a model system for experimental and theoretical studies of materials properties and surface reactions. TiO$_2$ has several polymorphs, rutile and anatase being the most common ones. Rutile is the thermodynamically most stable bulk phase; anatase is stable in nanoparticles and shows higher activity in many photocatalytic reactions, making it the most interesting phase for use in high surface area photocatalytic and photovoltaic devices. While insulating when stoichiometric, TiO$_2$ is most frequently reduced and electrically conducting due to the presence of intrinsic defects and their induced excess electrons, which transform it into an *n*-type semiconductor with many technologically relevant properties. As surfaces have a prominent role in most of TiO$_2$'s applications, intense efforts

A. Selloni (✉)
Department of Chemistry, Princeton University, Princeton, NJ, USA
e-mail: aselloni@princeton.edu

have been devoted to the characterization of TiO_2 surfaces and their interactions with various molecular species and environments. In this chapter, we present an overview of recent theoretical/computational studies on TiO_2, focusing mainly on reduced anatase surfaces and their interactions with molecular oxygen and water, which have a central role in various fields, from photocatalysis and nanomaterial synthesis to geochemistry and environmental chemistry. These studies highlight the effectiveness of combined experimental and theoretical approaches as well as the important role of simulations in bridging the gap between experiments under vacuum conditions and in realistic environments.

1   Introduction

Titania (TiO_2) is one of the most chemically stable, environmentally compatible, and functionally versatile oxide materials. While huge amounts ($\sim$7 million tons) of TiO_2 are produced every year for use as a white pigment in paints and many other common products, TiO_2 is also a prime material for advanced applications such as solar energy harvesting and hydrogen generation from water (Chen et al. 2010), photocatalytic decomposition and synthesis of organic chemicals (Schneider et al. 2014), environmental remediation (Hoffmann et al. 1995), reduction of CO_2 to chemical fuels (Ma et al. 2014), electron conductor in dye-sensitized solar cells (Hagfeldt et al. 2010), rechargeable batteries, super-capacitors, sensors (Bai and Zhou 2014), and biomedical devices (Rajh et al. 2014). In fact, TiO_2 is currently almost the only material that is suitable for photocatalytic applications on the industrial scale.

A fundamental characteristic of TiO_2 of great importance in photocatalysis is the energy of its valence band maximum (VBM) and conduction band minimum (CBM) (Graetzel 2001). Figure 1 shows the band edges of various photocatalysts, including TiO_2, compared to the water oxidation and reduction potentials. A photoexcited electron (e^-) can be transferred to an acceptor molecule (A) if the CBM of the material is higher than the reduction potential of A. Similarly, a photoexcited hole (h^+) can be filled by donor (D) electrons if the VBM of the material is located lower than the oxidation potential of D. TiO_2 has the appropriate band alignments for water splitting, which is one of the main reasons of the scientific interest in this material for over 40 years. The low energy of TiO_2's VBM is an additional favorable feature, which can largely explain why TiO_2 is an efficient photocatalyst for the oxidative degradation of organic pollutants.

TiO_2 has several crystalline forms, rutile and anatase being the most common ones. Rutile is the thermodynamically most stable bulk phase, while anatase is very common and stable in nanomaterials, due to the relatively low formation energies of its surfaces (Zhang and Banfield 1998). The anatase phase shows also the highest photocatalytic activity, which makes it the most interesting phase for use in high surface area photocatalytic and photovoltaic devices (Kavan et al. 1996). The structures of these two TiO_2 polymorphs are shown in Fig. 2. Both rutile (tetragonal, P_{42}/mnm, $a = b = 4.594$ Å, $c = 2.959$ Å) and anatase (tetragonal,
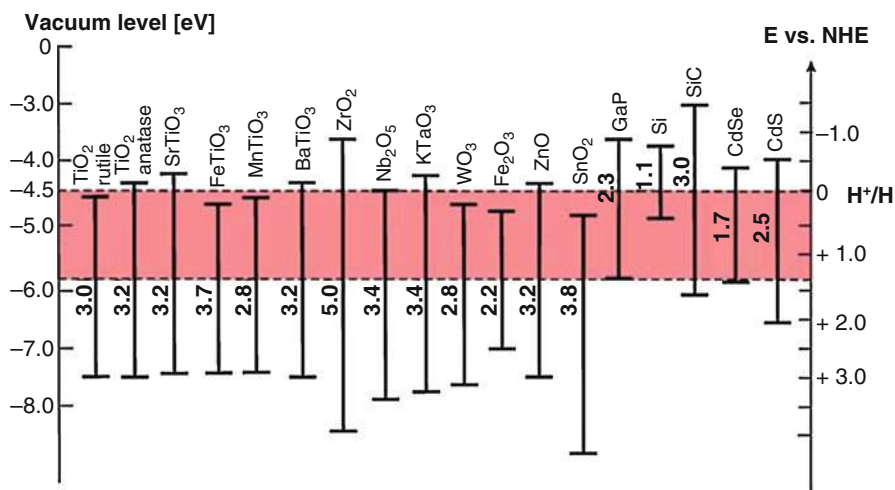
Fig. 1 Band edge positions of various semiconductor photocatalysts relative to vacuum level compared to the water reduction and oxidation potentials. (Reproduced from Batzill 2011 with permission of the Royal Society of Chemistry)

Fig. 2 Bulk structures of (**a**) anatase and (**b**) rutile. The tetragonal unit cell of rutile has dimensions $a = b = 4.587$ Å, $c = 2.957$ Å, and the one of anatase $a = b = 3.782$ Å, $c = 9.502$ Å
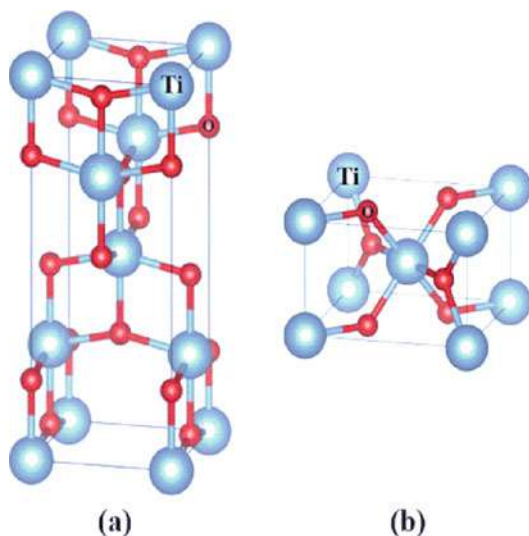
$I4_1/\text{amd}$, $a = b = 3.782$ Å, $c = 9.502$ Å) are built from chains of distorted TiO_6 octahedra, with Ti atoms coordinated by six oxygens and each oxygen atom surrounded by three planar Ti atoms. The distortion of the octahedral field is small in rutile, where the two apical Ti-O bonds are $\sim$1.98 Å and the four equatorial Ti-O bonds form a square with bond length $\sim$1.95 Å. The distortion is more pronounced in anatase, where the mirror plane symmetry perpendicular to the apical

bonds ($\sim$1.973 Å) is lost, and the four equatorial Ti-O bonds ($\sim$1.93 Å) are not in a plane. Interestingly, first principles electronic structure calculations have for long time been unable to reproduce the experimentally observed higher stability of rutile relative to anatase. Only very recently highly accurate quantum Monte Carlo calculations have shown that the energetic contribution arising from thermal nuclear motion is crucial in determining the phase stability of these two TiO_2 polymorphs (Trail et al. 2017).

Because surfaces play a major role in virtually all TiO_2 applications, it is of great interest to understand the effect of surface structure on various TiO_2's functional properties. Therefore, intense efforts have been devoted to characterizing the atomic geometries, electronic structures, and reactivities of different anatase and rutile surface terminations (Diebold 2003; Pang et al. 2013; De Angelis et al. 2014). Figure 3 shows the equilibrium crystal shapes of rutile and anatase derived from first principles calculations of the surface energies for different terminations. For rutile, the computed Wulff shape is mainly terminated by (110) and (011)/(101) surfaces (Ramamoorthy et al. 1994), whereas for anatase it consists of a truncated tetragonal bipyramid exposing majority (101) and minority (001) facets (Lazzeri et al. 2001). It is quite remarkable that these shapes, obtained assuming perfectly clean, stoichiometric, and defect-free bulk-terminated surfaces, agree well with the typical shapes of many synthetic and natural samples, which actually contain a large variety of imperfections, defects, and adsorbates.

In this Chapter, we present an overview of recent investigations on TiO_2, with particular focus on theoretical studies based on first principles total energy calculations and molecular dynamics simulations. The aim is to highlight the new physical/chemical insights provided by these studies as well as the advances that can be made by a combined theoretical and experimental approach. Most theoretical studies on TiO_2 are based on density functional theory (DFT), which has generally proven quite successful at predicting TiO_2 properties and explaining experimental observations over the last 20 years (De Angelis et al. 2014). However, the standard local density approximation (LDA) and generalized gradient approximation (GGA) of DFT fail in describing excited state properties as well as the behavior of the excess

Fig. 3 Equilibrium shapes of rutile (left) and anatase (right) obtained using the Wulff construction and the surface energies computed in Ramamoorthy et al. (1994) and Lazzeri et al. (2001), respectively
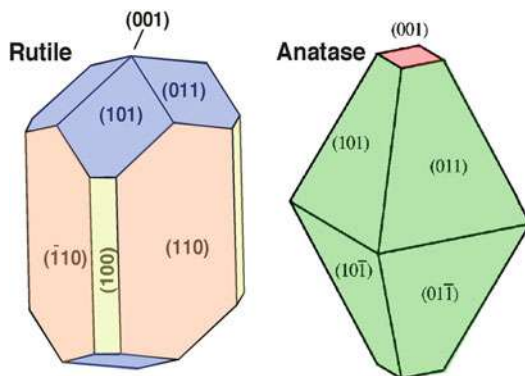
electrons of reduced TiO_2 (see below) (Di Valentin et al. 2009). This limitation originates from the self-interaction error (Cohen et al. 2008), causing problems such as the underestimation of the band gap and the over-delocalization of electrons in defects or impurity states. Green's function-based methods (Onida et al. 2002) are in general of great value to deal with problems such as calculations of excitation and ionization energies, transition matrix elements, absorption coefficients, and so on but are extremely expensive for studying systems of large size, such as surfaces and interfaces. In practice, more pragmatic approaches, notably the DFT+U method (Anisimov et al. 1991), explicitly penalizing the electron delocalization, and hybrid DFT functionals (Becke 1993; Perdew et al. 1996), including a fraction of Hartree-Fock exact exchange, are typically used to mitigate the self-interaction error. However, both methods are rather arbitrary in the amount of "correction" they include. As a result, the U value or fraction of exchange that works well for a particular application often does not produce similarly good results for another. Hence, much care is needed not only in applying these methods but also in comparing their results to experiment.

2   Reduced TiO_2

In stoichiometric form, rutile and anatase TiO_2 are insulators with band gaps of ~3.05 and 3.20 eV and formal oxidation states O^{2-} and Ti^{4+} for the oxygen and titanium ions, respectively. Most frequently, however, TiO_2 samples are reduced (i.e., substoichiometric with a reduced oxygen content) and electrically conducting, due to the presence of intrinsic defects, notably oxygen vacancies (V_O's) and titanium interstitials (Ti_{int}'s) (Di Valentin et al. 2009). These defects induce excess electrons, e.g., the removal of a neutral oxygen atom leaves two excess electrons in the lattice, thus transforming TiO_2 into an *n*-type semiconductor.

TiO_2's reduction is also associated with the formation of Ti^{3+} species, for which there are various experimental evidences, such as a g-tensor typical of a Ti $3d^1$ state measured in electron paramagnetic resonance (EPR) experiments (Berger et al. 2005; Livraghi et al. 2011, 2014), and the presence of a gap state at ≈ 2 eV above the valence band maximum and ≈ 1 eV below the conduction band minimum, which is observed by photoelectron spectroscopy (PES) and electron energy loss spectroscopy (EELS) (Diebold 2003; Thomas et al. 2007). This deep level indicates that V_O's excess electrons occupy well-localized states characteristic of small polaron behavior. At the same time, however, the occurrence of *n*-type conductivity suggests that these defects behave as shallow donors (Forro et al. 1994). For anatase, this behavior was confirmed by ARPES measurements (Moser et al. 2013), which showed a shallow donor level, ~40 meV below the CB, in samples containing a large concentration of O-vacancies.

These seemingly conflicting results have motivated numerous investigations of intrinsic defects and their excess electrons in reduced TiO_2 in the last decades, and several general features of their states have now become clearer. In particular, there is now convincing evidence that excess electrons in bulk rutile tend to form
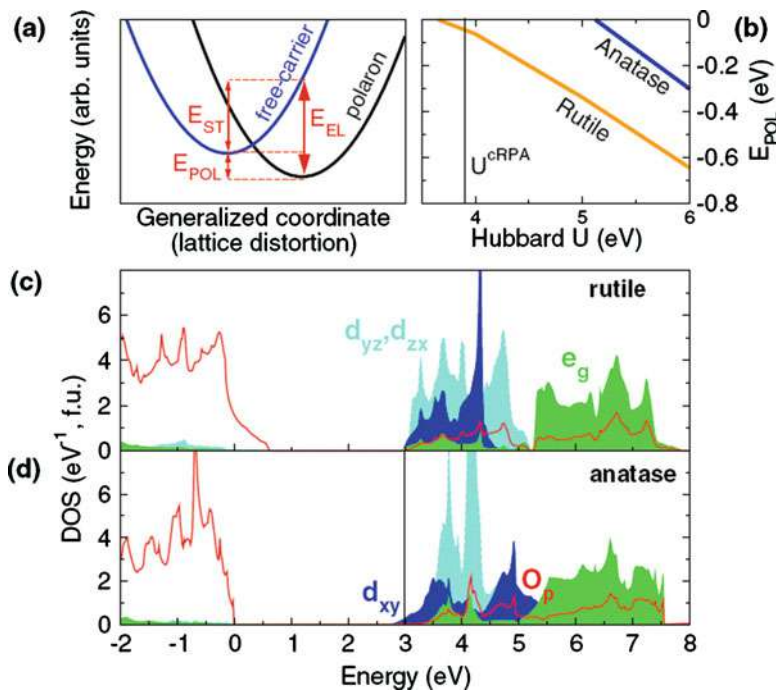
Fig. 4 (**a**) Schematic configuration coordinate diagram showing the polaronic formation energy (E_{POL}), lattice deformation energy (E_{ST}), and electronic localization energy (E_{EL}) as a function of lattice distortion for the polaronic and delocalized states. E_{POL} and E_{EL} represent the adiabatic and vertical excitation energies of the electron from the localized polaronic state to the "free" conduction band state and can be probed experimentally by, e.g., transport measurements and PES, respectively. (**b**) E_{POL} as a function of Hubbard U in bulk rutile (orange) and anatase (blue) shows that electron localization is "easier" (i.e., requires a smaller U value) in rutile than in anatase. The vertical line indicates the value U^{cRPA} determined using advanced RPA calculations. Orbitally decomposed density of states (DOS) in rutile (**c**) and anatase (**d**), aligned with respect to the Ti core levels. (Reproduced with permission from Setvin et al. 2014a)

small polarons largely localized at single Ti lattice sites, whereas electrons prefer to form large polarons in anatase (Setvin et al. 2014b) (Fig. 4). Electron transport thus occurs mainly via hopping in rutile (Deskins and Dupuis 2007), while a band-like mechanism becomes accessible at relatively low temperature in anatase.

Another well-known characteristic of reduced TiO_2 is its blue color, whose intensity increases with the degree of reduction (Fig. 5). It was proposed earlier on that electronic $d \rightarrow d$ transitions originating from the Ti^{3+}- induced states in the band gap are responsible for the blue color (Khomenko et al. 1998), but the final states of these transitions were not known in detail. More recently, two-photon photoemission (2PPE) measurements identified an excited resonant state centered at ~ 2.5 eV above the Fermi energy (E_F) associated with Ti^{3+} species in reduced TiO_2 (Onda et al. 2004; Argondizzo et al. 2015; Wang et al. 2015; Zhang et al. 2015).
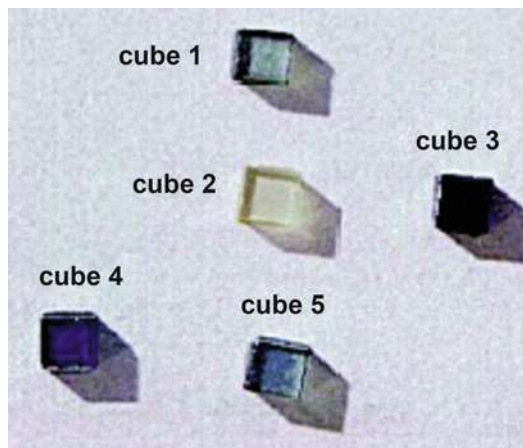
DFT calculations showed that this excited state is closely related to the band gap state, as they both result from the Jahn-Teller-induced splitting of the $3d$ orbitals of Ti^{3+} ions (Wang et al. 2015). The $d \rightarrow d$ transitions from the Ti^{3+} related occupied band gap state at $\sim$ -(0.8–1.0) eV to the empty excited state at $\sim$2.5 eV increase the photoabsorption in the ultraviolet (UV) region. At the same time, the wide distribution of these two states can extend the absorbance to the visible range. This extended photoabsorption together with the near infrared absorption produces an absorption minimum in the blue, which is the reason why reduced TiO_2 samples appear as blue.

3   Structure and Reactivity of Anatase TiO_2 Surfaces

While surface science studies of rutile date back to the late 1970s (Henrich et al. 1976), experimental characterization of well-defined anatase surfaces has been for long time challenging due to difficulties in growing sufficiently large single crystals for surface science measurements. For this reason, a great deal of information on anatase surfaces has initially come from theoretical studies (De Angelis et al. 2014). Anatase crystals most frequently expose (101) and (100)/(010) surfaces, with a smaller contribution of (001) facets (Fig. 3). The atomic geometries of (101) and (001) surfaces are shown in Fig. 6. From low energy electron diffraction (LEED) experiments, it is known that the majority (101) surface has the same (1×1) periodicity of the bulk termination. This surface is quite corrugated and exposes both undercoordinated fivefold Ti (Ti_{5c}) and twofold O (O_{2c}) atoms, as well as fully coordinated sixfold Ti (Ti_{6c}) and threefold O (O_{3c}) atoms (Fig. 6a). DFT studies have provided a detailed picture of the atomic relaxations on this surface, and the computed pattern has been found to agree well with the results of surface X-ray diffraction (SXRD) measurements in a recent study (Treacy et al. 2017).
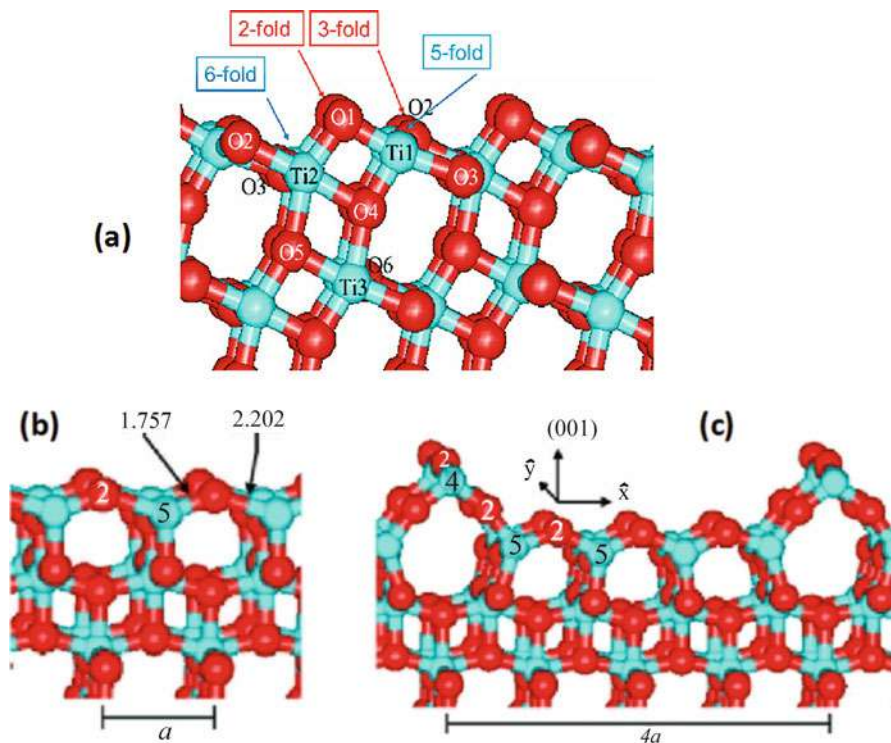
Fig. 6 Ball and stick models (side views) of (**a**) anatase TiO_2 (101); (**b**) relaxed anatase TiO_2 (001) – (1 × 1); (**c**) reconstructed anatase TiO_2 (001) – (1 × 4) according to the ADM model (Lazzeri and Selloni 2001)

Unlike anatase (101), the bulk-terminated anatase TiO_2(001) surface exposes only coordinatively unsaturated Ti atoms (Ti_{5c}) together with O_{2c} and O_{3c} atoms (Fig. 6b). However, this structure is only metastable, and experiments have revealed that the anatase (001) surface is actually reconstructed with a (1 × 4) periodicity with respect to the bulk (Herman et al. 2000). The model that is widely accepted to represent the structure of the (1 × 4) reconstructed surface is the so-called "ad-molecule" (ADM) model (Fig. 6c), which was originally proposed on the basis of DFT calculations (Lazzeri and Selloni 2001). The ADM model is constructed by periodically replacing [100] rows of bridging O_{2c} atoms of the unreconstructed 1 × 1 surface with rows of TiO_3 species. This results in "ridges" on which the Ti atoms are fourfold coordinated. The computed surface energy of the 1 × 4 ADM structure is nearly half that of the unreconstructed surface, indicating that this reconstruction stabilizes the surface quite effectively. The primary mechanism of this stabilization is the relief of the surface stress (Lazzeri and Selloni 2001). On the bulk-terminated surface, the large Ti_{5c}-O_{2c}-Ti_{5c} bond angles (∼156°) and Ti_{5c}-O_{2c} bond lengths (1.96 Å) give rise to a significant tensile surface stress. In the 1 × 4 ADM structure, the insertion of an extra TiO_2 row causes a compression which

makes the surface bonds shorter, 1.80–1.85 Å, and the bond angles smaller, $\sim 120°$; as a result, the tensile stress is also strongly reduced.

3.1 Reduced Anatase TiO$_2$(101) Surface

A well-known result of surface science studies on TiO$_2$ is that the rutile TiO$_2$(110) surface exhibits a large (5–10%) concentration of surface oxygen vacancies under the typical preparation conditions used for those experiments (Diebold 2003). In contrast, STM measurements on a freshly cleaved anatase TiO$_2$(101) sample revealed an almost perfect surface with very few subsurface impurities and adsorbates and no oxygen vacancies (He et al. 2009a). DFT calculations showed that oxygen vacancies have indeed lower formation energy in the subsurface and deeper layers than at the very surface (Cheng and Selloni 2009a). DFT studies also predicted that an O-vacancy could easily migrate from the surface to the subsurface with a barrier of about 0.75 eV (Cheng and Selloni 2009a). Experimentally, surface oxygen vacancies are not naturally present on anatase (101) but could be induced by electron bombardment and found to migrate to subsurface sites at temperature larger than 200 K, in agreement with the theoretical prediction (Scheiber et al. 2012).

The different characteristics of oxygen vacancies and their accompanying excess electrons on the rutile TiO$_2$ (110) and anatase TiO$_2$ (101) surfaces are illustrated in Fig. 7. On rutile, V$_O$s prefer to form at surface bridging oxygen sites,

(a)   (b)
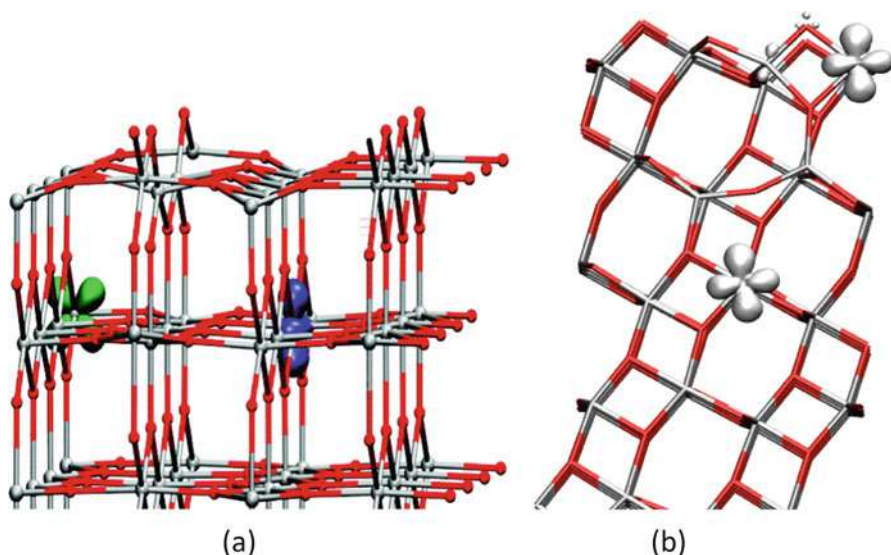
Fig. 7 Spatial distribution of the two excess electrons from an oxygen vacancy at the reduced rutile (110) and anatase (101) surfaces. (**a**) Surface V$_O$ in reduced rutile TiO$_2$ (110) (Deskins et al. 2011) and (**b**) subsurface V$_O$ in reduced anatase TiO$_2$ (101) (Selçuk and Selloni 2014), both from DFT+U calculations

and their excess electrons are mainly located away from the vacant site in the subsurface, where they form small polarons (Ti^{3+} species) that can hop rapidly among different sites (Deskins et al. 2011) (Fig. 7a). While the distribution of the excess electrons of a subsurface V_O on anatase (101) is not known as well as in the case of a surface V_O on rutile (110), several theoretical studies have predicted a distribution with one of the two electrons localized at a subsurface Ti site close to the vacancy and the other more or less (depending on the U value) delocalized on surface Ti sites (Selçuk and Selloni 2014); an example where both electrons form two well-localized Ti^{3+} species is shown in Fig. 7b.

3.2 The Interaction of Molecular Oxygen with Reduced Anatase $TiO_2(101)$

The interaction of molecular oxygen with TiO_2 surfaces plays a key role in many technologically important processes such as catalytic oxidation reactions, chemical sensing, and photocatalysis. For example, O_2 adsorbed on TiO_2 surfaces acts as an electron scavenger and is often used to suppress electron-hole recombination, which increases the lifetime of the excited state and thus the yield of the photocatalytic reaction (Berger et al. 2005). Molecular oxygen does not adsorb on stoichiometric TiO_2; excess electrons, i.e., reduced surfaces, are required. In fact, oxygen adsorption involves the transfer of one or more of these excess electrons from the TiO_2 surface to the O_2 molecule. Oxygen adsorption is also the first step toward oxygen incorporation, a fundamental reaction that strongly affects the chemical properties and charge-carrier densities, as it can transform the material from an *n*-type semiconductor to a poor electronic conductor.

An important aspect of the interaction of O_2 with reduced anatase TiO_2 concerns the charge state of adsorbed O_2. While DFT calculations generally predict that O_2 adsorbs as a peroxide (O_2^{2-}) (Aschauer et al. 2010), EPR measurements on anatase powders have identified stable superoxide (O_2^-) species (Berger et al. 2005). This difficulty has motivated hybrid functional calculations of the charge transfer reaction between reduced anatase (101) and molecular O_2 (Li and Selloni 2013). It was found that although the peroxo O_2^{2-} species is energetically more stable than O_2^-, there is a significant barrier, $\sim$0.3–0.4 eV, that must be overcome to transfer a second electron to the adsorbed O_2^- (whose formation is barrierless) and transform it to a peroxide. The existence of this barrier can thus explain why experimentally superoxo species are often observed.

Although O-vacancies are located subsurface on anatase (101), they can actively take part in the reaction with molecular O_2. In FPMD simulations it was indeed observed that an O_2 molecule adsorbed at a surface Ti_{5c} site close to a subsurface V_O can further react with the subsurface V_O according to the sequence **b-e** shown in Fig. 8 (Setvin et al. 2013). In the resulting structure, denoted $(O_2)_O$, the O_2 takes the position of a twofold coordinated O (O_{2c}) atom in a bridging configuration.
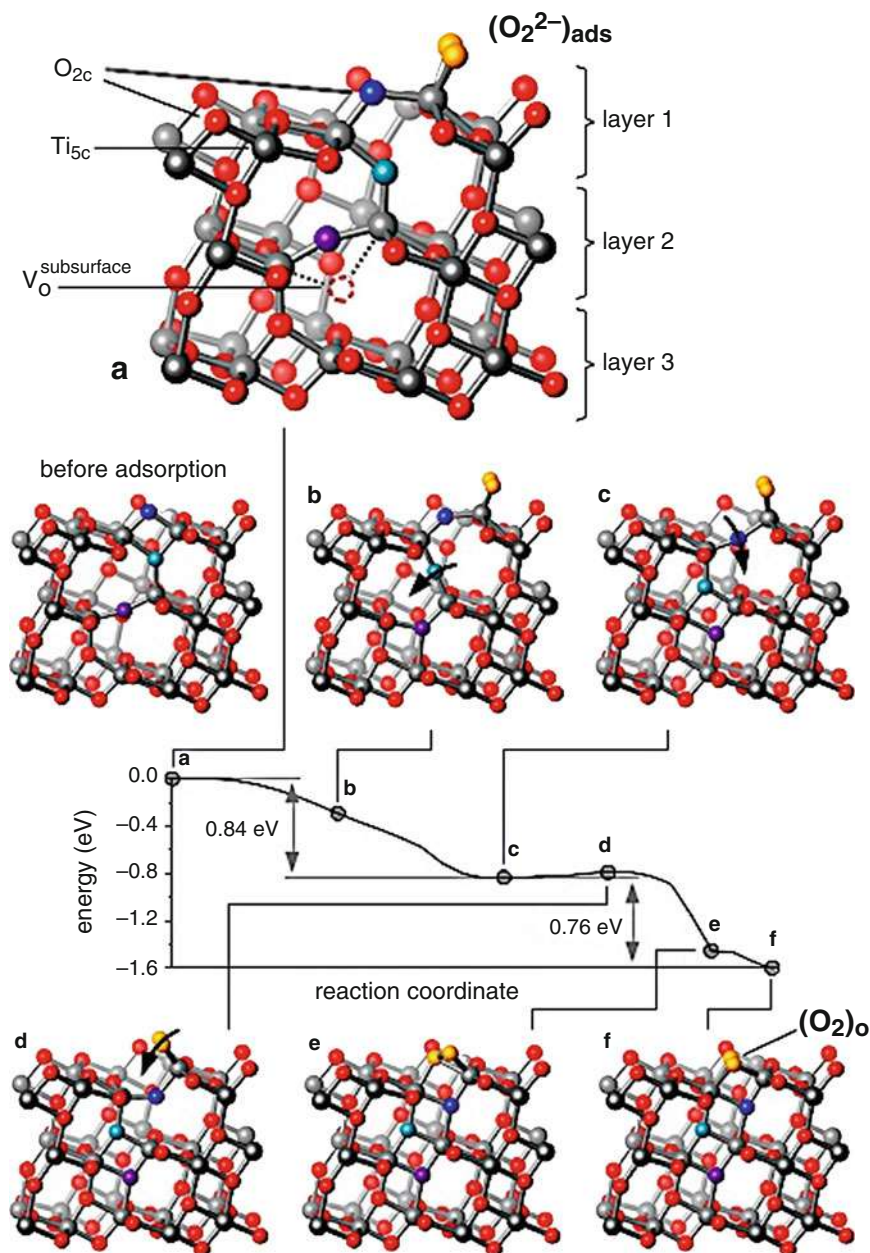
Fig. 8 Potential energy profile and selected structure snapshots along the pathway leading from an adsorbed O_2 atop a subsurface oxygen vacancy to the bridging dimer state at the anatase TiO_2 (101) surface. (Adapted from Setvin et al. 2013)

The $(O_2)_O$ retains a bond length of 1.46 Å, characteristic of peroxide. This species (often referred to as "interstitial" or "bridging" O_2) had been previously proposed to be an important intermediate in the photocatalytic splitting of H_2O on anatase (101) (Li et al. 2010).

Experiments were performed to verify whether such an $(O_2)_O$ species really exists (Setvin et al. 2013). Since $(O_2)_O$ should also form when an O_2 directly reacts with a surface V_O, surface V_O's were created on anatase (101) via electron bombardment and subsequently the surface was dosed with O_2. STM measurements showed that some of the V_O's were replaced by double spots located at O_{2c} sites, whose appearance agrees well with calculated STM images of the $(O_2)_O$ configuration. Without such artificially created surface V_O's, the same $(O_2)_O$ were observed quite rarely, most likely because O_2 does not often adsorb just on top of a subsurface oxygen vacancy.

4   Structure of TiO$_2$ Aqueous Interfaces

The structure of the TiO_2 – water interface – is relevant to a wide range of applications, from photocatalysis and biocompatible devices to more traditional applications like pigments and coatings (Sun et al. 2010). It was shown theoretically (Vittadini et al. 1998), and later confirmed by STM measurements (He et al. 2009a), that isolated water molecules adsorb in undissociated form on the anatase TiO_2 (101) surface, with the water oxygen (O_w) forming a dative bond with the Ti_{5c}, and the water hydrogen atoms (H_w) forming H-bonds with two O_{2c} atoms on the two adjacent rows (Fig. 9). The absence of water dissociation indicates that the (101) surface is little reactive, as one generally expects for very stable (i.e., low energy) surfaces. Molecular adsorption was theoretically predicted to remain favorable also at high water coverages, e.g., for one, two, and three adsorbed water monolayers (Tilocca and Selloni 2004), in accord with temperature-programmed desorption (TPD) experiments (Herman et al. 2003).

Several first principles molecular dynamics (FPMD) simulations have examined the structure of the interface between anatase (101) and liquid water (Cheng and Selloni 2010; Sumita et al. 2010; Hosseinpour et al. 2017). A first layer of water molecules adsorbed molecularly at Ti_{5c} sites and a second layer forming strong H-bonds with the O_{2c} surface atoms characterize the structure of this interface. At variance with the case of submonolayer water adsorption, the water hydrogen atoms of the molecules adsorbed at Ti_{5c} sites form H-bonds with the molecules of the second layer rather than with surface O_{2c} atoms. These two layers form a stable bilayer due to the presence of short H-bonds between first and second layer water molecules (Fig. 10a).

Only a few studies have examined the interface of water with anatase $TiO_2(001)$, and it is still unclear whether this surface maintains the (1×4) structure found in UHV when in contact with water. For both the bulk-terminated and reconstructed
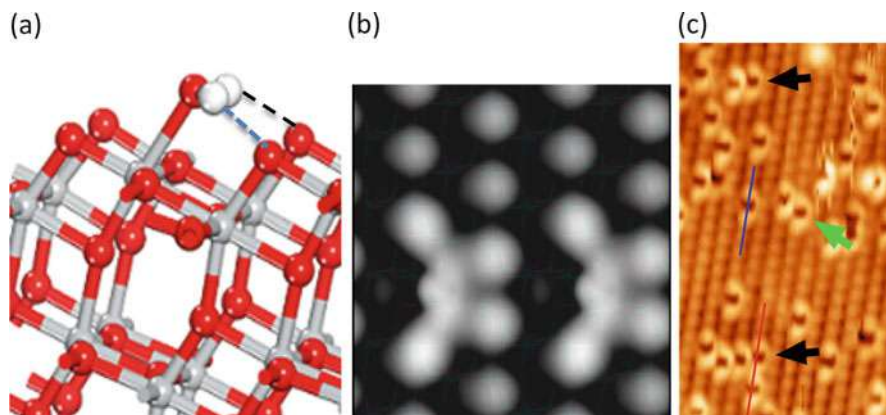
Fig. 9 Water on anatase TiO$_2$ (101): (**a**) computed adsorption geometry of an isolated molecule, showing the H-bonds formed by the water hydrogens with the surface bridging oxygen atoms; (**b**) simulated constant density image, determined from the integrated local density of states in an energy window of 2.75 eV from the conduction band minimum; the image is periodically repeated. (**c**) Experimental empty state STM image taken at a sample temperature of 190 K (He et al. 2009b). The green and black arrows indicate ordered water clusters. Note that the image of a single water molecule extends over five lattice sites

surfaces, however, theoretical studies have predicted that water adsorbs dissociatively in the dilute limit and in mixed molecular-dissociative mode at high coverage, the fraction of dissociated molecules being of the order 25–30% (Selçuk and Selloni 2013). A similar amount of dissociated water molecules has been reported in first principles molecular dynamics simulations of the aqueous (001) interface (Fig. 10b) (Sumita et al. 2010). When one water molecule dissociates, one of the two Ti-O$_{2c}$ bonds of the Ti binding site breaks, causing a significant local restructuring.

Another well-studied aqueous interface is that with rutile TiO$_2$(011). This is the second most frequent surface in the equilibrium shape of rutile crystals (Fig. 3) and has attracted attention because its photocatalytic activity was reported to be significantly different from that of the most stable rutile (110) surface (Ohno et al. 2002). After surface science experiments discovered that the rutile TiO$_2$(011) surface is reconstructed with a (2 × 1) periodicity (Beck et al. 2004), extensive search based on DFT calculations identified a structural model with much lower surface energy in comparison to the bulk-truncated surface (Gong et al. 2009), which could also well reproduce the results of surface X-ray diffraction (SXRD)(Torrelles et al. 2008) and voltage-dependent STM measurements (Gong et al. 2009) (Fig. 11a).

To understand what happens when the rutile TiO$_2$ (011) surface structure observed under vacuum conditions is put in contact with the aqueous environment typical of TiO$_2$ photocatalysis, it is useful to first investigate the interaction of the surface with both isolated water molecules and a water monolayer. Using DFT calculations, it was found that the adsorption energy of an isolated water molecule on the non-reconstructed (1 × 1) surface is about three times larger than on the
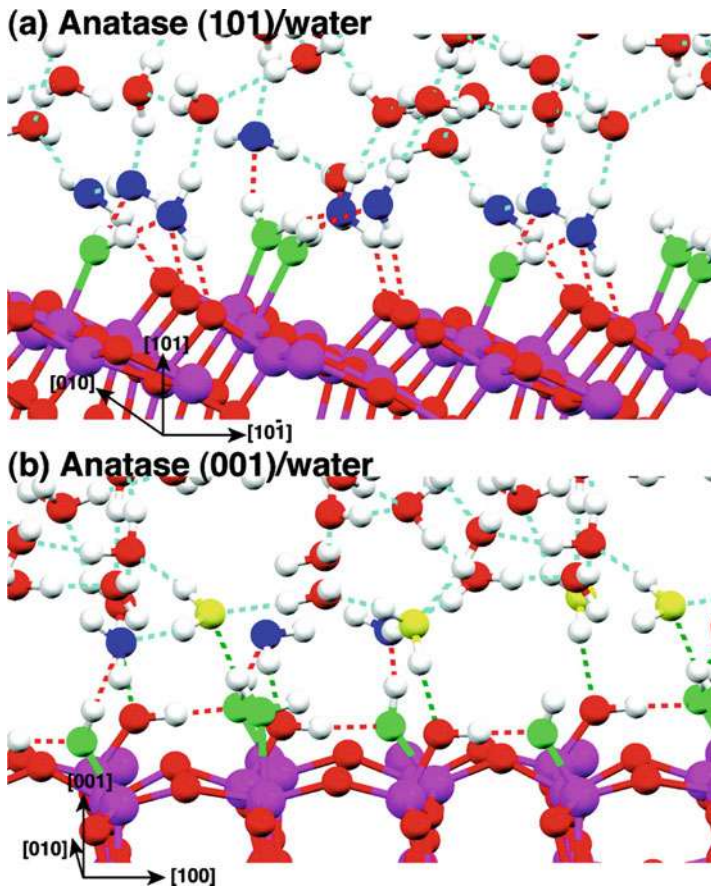
Fig. 10 Snapshots from FPMD simulations of the interface of bulk water with (**a**) anatase TiO_2 (101) and (**b**) anatase TiO_2 (001). Water molecules or hydroxyl groups with green oxygen atoms adsorb onto five-coordinated Ti (Ti_{5c}). Water molecules with blue oxygen adsorb onto surface two-coordinated O_{2c} or adsorbed water molecules via strong H-bonds (red dashed lines). Water molecules with yellow oxygen adsorb to H_2O or OH that are already adsorbed onto Ti_{5c} via weak H-bonds (green dashed lines). Cyan dashed lines show ordinary H-bonds. (From Sumita et al. 2010)

reconstructed one, i.e., $E_{ads} = -1.27$ vs -0.42 eV (Aschauer and Selloni 2011). In other words, the bulk-terminated surface is much more reactive than the reconstructed one, consistent with its higher surface energy. Similarly, the adsorption of one water monolayer was found to be highly favorable ($E_{ads} = -1.04$ eV per molecule) on the non-reconstructed surface, whereas only a half-monolayer of water could bind directly to the surface Ti sites on the reconstructed surface (Aschauer and Selloni 2011). The formation of a tightly bound monolayer of adsorbed water molecules strongly stabilizes the non-reconstructed surface with respect to the reconstructed one, thus suggesting that the latter may switch back to
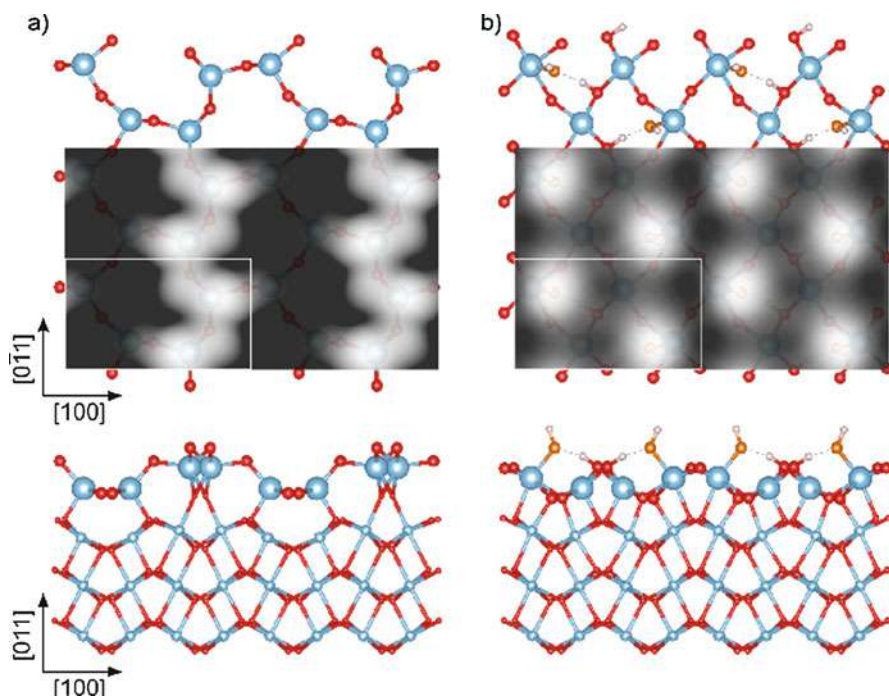
Fig. 11 Computed structures (top and side views) and STM images of (**a**) the reconstructed rutile TiO$_2$ (011), (2 × 1) surface; (**b**) the bulk-terminated rutile TiO$_2$ (011), (1x1) surface with two dissociatively adsorbed water molecules per (2 × 1) unit cell. Oxygen atoms belonging to terminal OH groups are highlighted in orange, and the white rectangle in the STM images shows the unit cell used for the calculations. Constant-density STM images are computed for a 0.5 eV bias from the conduction band minimum and are obtained by averaging images of snapshots taken from a FPMD simulation at 300 K (Balajka et al. 2017)

the bulk-terminated structure (i.e., deconstruct) when put in contact with an aqueous environment.

Recently experiments were carried out to verify this suggestion (Balajka et al. 2017). After exposure of the (2 × 1)-reconstructed surface to liquid water, the sample was transferred back to UHV where an overlayer with an apparent "c(2 × 1)" symmetry was observed in scanning tunneling microscopy (STM) images. On the basis of STM analysis and DFT calculations, this overlayer could be attributed to dissociated water on top of the unreconstructed (1 × 1) surface, confirming that the (2×1) reconstruction of the UHV-prepared surface is lifted when the surface is in contact with liquid water, as theoretically predicted. The deconstruction is achieved by rearranging only the surface atoms, and there is no mass transport involved. After evacuation of the liquid water, dissociated water hydroxyls remain adsorbed on undercoordinated Ti$_{5c}$ atoms of the (1 × 1) surface, forming an ordered overlayer with (2 × 1) symmetry and two hydroxyl groups per unit cell (Fig. 11b).

5   Excess and Photoexcited Electrons at Anatase TiO$_2$ Surfaces and Aqueous Interfaces

As for excess electrons in reduced TiO$_2$, the picture of photoexcited electrons in anatase is based on the polaron model, where the electron forms a localized state that is stabilized by an accompanying lattice distortion (the EPR signal of photoexcited electrons in anatase is indeed very similar to that of excess electrons from intrinsic defects or donor impurities). An important question to ask at this point concerns the degree of localization of the polaronic state.

As mentioned in Sect. 2, some experimental studies have identified large polaronic states formed by excess electrons in reduced and *n*-doped anatase. In particular, evidence of large polaronic states has been provided by angle-resolved PES (ARPES) measurements, which show a shallow dispersive quasiparticle band at about 40 meV below the CBM in reduced anatase (001) films (Moser et al. 2013). However, different results were obtained in other experiments. A prominent gap state 1 eV below the Fermi level was observed in PES experiments on single crystal anatase (101), whereas the same feature was drastically less pronounced on the anatase (001) surface (Thomas et al. 2007). The deep gap level on anatase (101) is similar to that observed on the rutile (110) surface, for which the formation of small polarons is well established. Finally, both deep and shallow polaronic states on anatase (101) were observed by Setvin et al. (2014), who used an STM tip to create V$_O$s on the surface (normally devoid of V$_O$s) and probed the local density of states around both these vacancies and subsurface impurities using scanning tunneling spectroscopy (STS).

In addition to the above studies carried out on well-defined surfaces under vacuum conditions, experiments on disperse anatase samples have revealed a remarkable dependence of the photocatalytic activity of anatase on the crystal surface, with oxidation and reduction reactions occurring preferentially on {001} and {101} facets, respectively (Tachikawa et al. 2011; Yu et al. 2014). Since these reactions involve photoexcited carriers, it is relevant to ask whether the facet dependence of the photocatalytic activity is related to the diverse behaviors of excess electrons observed in the surface science experiments or is rather due to the different reactivity of the two surfaces. Answering this question requires understanding how the behavior of excess electrons that is observed under vacuum conditions is modified in a humid or aqueous environment, as typically used in photocatalysis. This understanding is not yet available from experiment, yet interesting insights could be obtained from theory and simulations.

A recent study examined the behavior of excess electrons in proximity of the three most important (101), (001), and (100) anatase surfaces (denoted a-101, a-001, and a-100, respectively) and their aqueous interfaces, using both DFT+U static calculations and DFT+U-based first principles molecular dynamics (FPMD) simulations (Selcuk and Selloni 2016). The investigated slab models contained a single photoexcited electron or an excess electron provided either by an adsorbed H at a surface undercoordinated oxygen resulting in a surface bridging hydroxyl (O$_{br}$H) or by a bulk dopant, such as an interstitial H atom or a Nb impurity substituting a Ti atom inside the slab.

Results for a-101 and a-100 show that electron trapping does not occur on and near the *defect-free* surfaces, but can be induced by a surface hydroxyl, and possibly other types of defects. The stability of the $O_{br}H$ complex is at the origin of an interesting and rather unexpected effect that is electron-induced water dissociation. The typical mechanism of water dissociation involves two water molecules: one of the two molecules transfers a H to the O_{br}, and the other transfers a H to the first water before being adsorbed as a OH on a Ti_{5c} (Fig. 12).

To further explore the effects of the water environment on the localization of the excess electron, H-doped and Nb-doped (101) slab models were considered, in which the positively charged impurity, acting as an attractive potential for the electron, was located in the innermost layers of the slab, thus reducing the tendency of the electron to migrate to the surface. The simulations show that the excess electron is very mobile and spreads over several Ti sites when the slab is in vacuo, whereas it can rapidly reach the surface and becomes trapped in the form of a Ti^{3+} species close to an $O_{br}H$ when the system is solvated. Clearly, the stabilization provided by the formation of the Ti^{3+}-$O_{br}H$ complex makes the surface much more favorable for the electron in water/humid environment than in vacuo, despite the competition of water for the occupation of the surface Ti_{5c} sites.

A very different behavior is observed for excess electrons near the a-001 surface (Fig. 13). Unlike on a-101, FPMD simulations show that the excess electron avoids the a-001 surface and remains in the subsurface and inner layers throughout the simulation. At each instant, the electron forms a two-dimensional state in a (001) plane, but switches easily from one plane to another, which could be the reason of the three-dimensional dispersion observed in experiment. Due to the limited size of the model, the two-dimensional state appears as completely delocalized rather than a large polaron, whose radius is estimated to be around 10–20 Å (Moser et al. 2013).

A simple explanation of the reason why the electron tends to avoid the surface layers is given by the layer-resolved density of states (LRDOS) of an a-001 slab in the absence of excess electrons (Fig. 13d). While all inner layers of the slab contribute similarly to the DOS, both the conduction and valence band edges of the outermost layer are shifted upward by about 0.5 eV in comparison to the other layers. Since the 3d orbitals of the surface Ti_{5c} atoms are located at a higher energy compared to those in the subsurface and bulk layers, excess electrons prefer to remain inside the slab rather than at the surface. By contrast, holes will likely prefer to move to the surface, where they can be readily available for oxidation reactions.

Besides showing clear differences in the behaviors of electrons at the a-101 and a-001 terminations, the results of this study (Selcuk and Selloni 2016) further illustrate the crucial role of the environment on the structure and reactivity of TiO_2 surfaces. Whereas no trapping of electrons occurs on the a-101 surface in vacuum, an excess electron at the aqueous a-101 interface tends to become trapped into a stable Ti^{3+}-$O_{br}H$ complex through dissociation of an adsorbed water molecule or reaction with an aqueous proton. This trapping favors the interaction of electrons with adsorbates, so that reduction reactions should be facilitated on the (101) surface. On the other hand, the a-001 surface is strongly repulsive for electrons and attractive for holes and is thus predicted to have a significant activity for oxidation reactions. This difference between the two most common terminations of anatase agrees well with
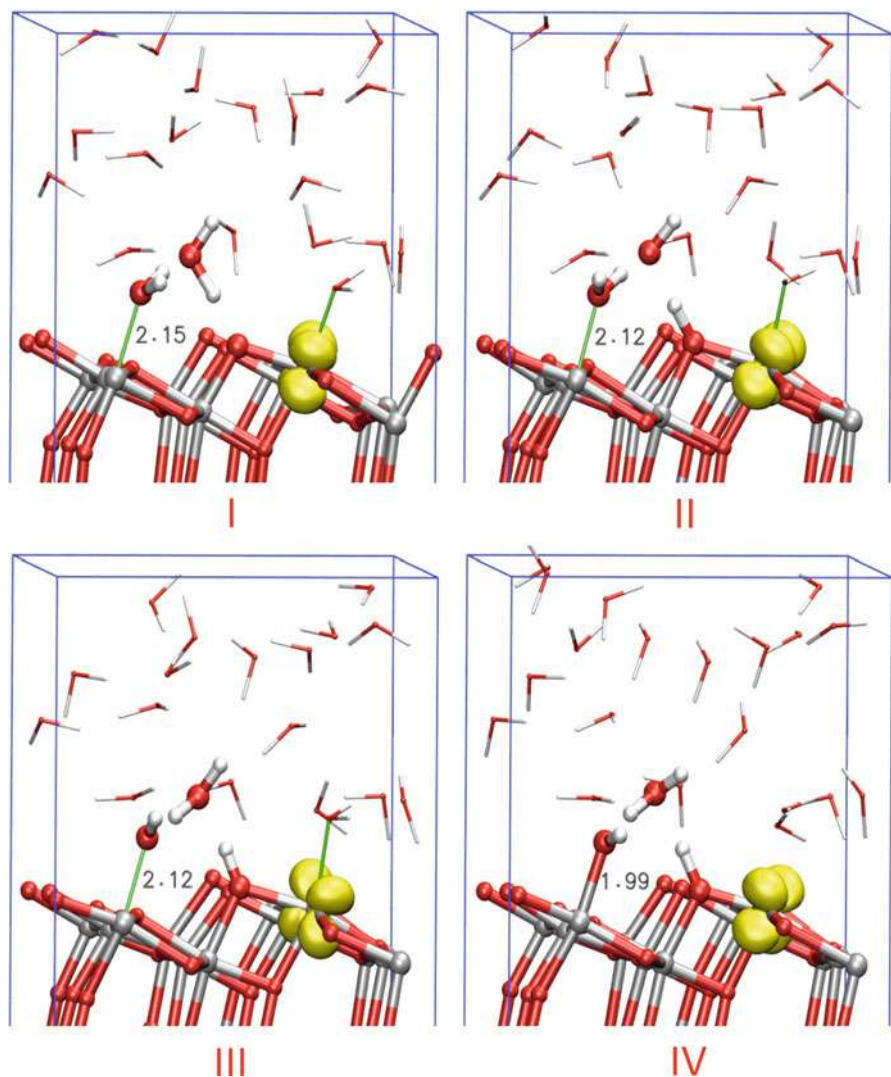
Fig. 12 Typical mechanism of water dissociation induced by a localized electron at the anatase (101) – water interface; the sequence of four snapshots has a duration of ∼0.1 ps. Images are obtained from the simulation of a photoexcited electron using DFT+U with U = 3.9 eV. In the absence of the excess electron, H^+ of $O_{br}H$ and the solvated OH^- in the second snapshot recombine very rapidly

photoemission spectroscopy experiments, explains the observed facet dependence of photocatalytic reactions on TiO_2, and supports the suggestion that optimization of the ratio of a-101 and a-001 facets could provide a way to enhance the photocatalytic activity of anatase.
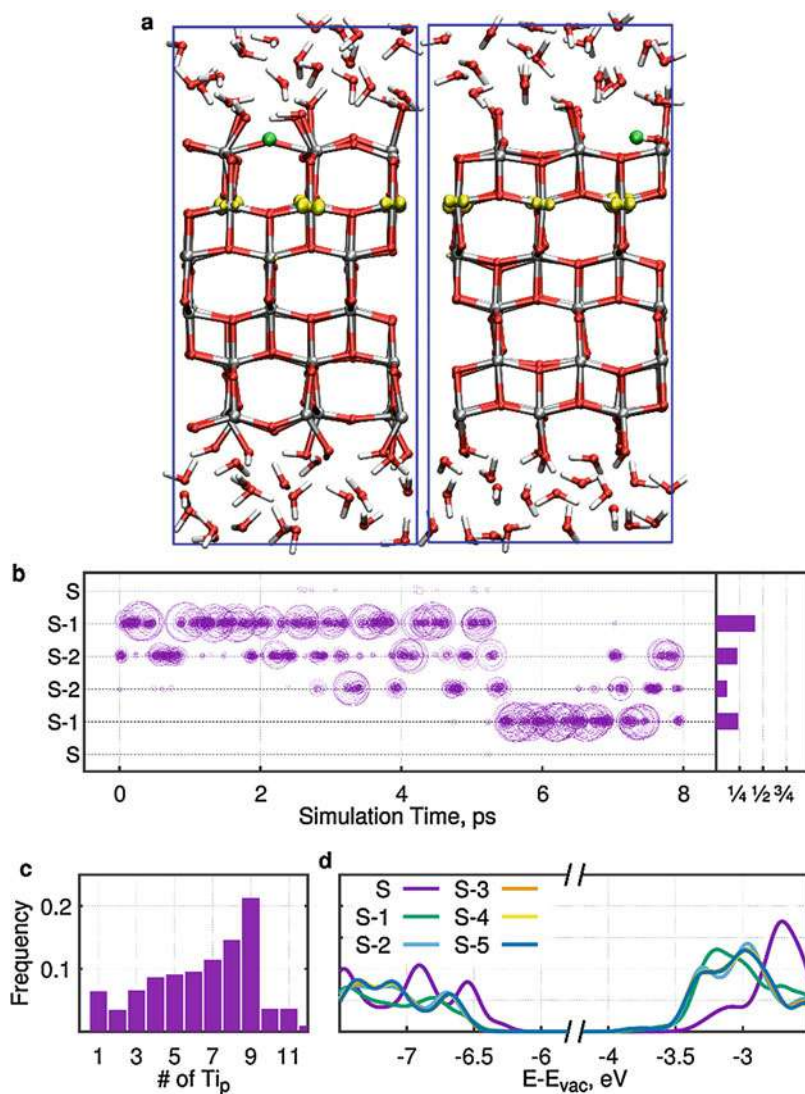
Fig. 13 (**a**) Side views of a six-layer hydroxylated (001) slab in water, from two different perspectives. The hydrogen on the bridging hydroxyl is shown in green. (**b**) Location of the e^-_{ex} in the direction perpendicular to the surface during a simulation with U = 3.9 eV and T = 400 K and corresponding distribution. The e^-_{ex} clearly avoids the (001) surface but switches frequently between different bulk layers. (**c**) Probability distribution of the number of Ti_p sites; note that the number of Ti atoms in each (001) layer of the slab is 9. (**d**) Layer-resolved density of states for an 11-layer a-001 slab. The labels S, S-1, S-2, etc. indicate the surface layer and the subsequent layers below the surface. Both the VB and the CB edges are shifted to significantly higher energies at the surface layer in comparison to the inner layers. The computed CB edge is located at −4.03 eV below the vacuum level, and the band gap is 2.09 eV. (From Selcuk and Selloni 2016)

6   Summary and Outlook

Despite its relatively low efficiency, TiO_2 remains one of the most widely used materials in photocatalysis as well as a benchmark for the development of new functional materials and a model system for the study of reaction mechanisms. At the same time, new applications as well as new areas of fundamental research on TiO_2 are emerging, which keep the interest in this material very lively. For example, TiO_2 is considered the most promising material for use as a thin film protective coating of semiconducting photoelectrodes in solar-fuel generators (Hu et al. 2015), while the study of 2D electron gas behavior at the anatase surface (Rödel et al. 2015) is an exciting new direction that has also the potential of novel interesting applications in electronic devices.

Theoretical studies and simulations have contributed substantially to the understanding of the fundamental physical and chemical properties of TiO_2 and related metal oxide materials over the last decades. In this chapter, we have presented an overview of a number of these contributions, ranging from the bulk electronic and optical properties to the polaronic character of defect states and charge carriers, the surface reactivity, the structure of the aqueous interface, and the behavior of electrons at this interface. Besides providing insight into existing experimental result, theory and simulations have proven capable to predict new properties subsequently verified by experiments and also to bridge the gap between experiments under carefully controlled conditions and experiments in more realistic environments.

References

Anisimov VI, Zaanen J, Andersen OK (1991) Band theory and Mott insulators: Hubbard U instead of Stoner I. Phys Rev B 44(3):943

Argondizzo A, Cui X, Wang C, Sun H, Shang H, Zhao J, Petek H (2015) Ultrafast multiphoton pump-probe photoemission excitation pathways in rutile TiO_2(110). Phys Rev B 91(15):155429

Aschauer U, Selloni A (2011) Structure of the rutile TiO_2 (011) surface in an aqueous environment. Phys Rev Lett 106(16):166102

Aschauer U, Chen J, Selloni A (2010) Peroxide and superoxide states of adsorbed O_2 on anatase TiO_2 (101) with subsurface defects. Phys Chem Chem Phys 12(40):12956–12960

Bai J, Zhou B (2014) Titanium dioxide nanomaterials for sensor applications. Chem Rev 114(19):10131–10176

Balajka J, Aschauer U, Mertens SFL, Selloni A, Schmid M, Diebold U (2017) Surface structure of TiO_2 rutile (011) exposed to liquid water. J Phys Chem C 121(47):26424–26431

Batzill M (2011) Fundamental aspects of surface engineering of transition metal oxide photocatalysts. Energy Environ Sci 4(9):3275–3286

Beck TJ, Klust A, Batzill M, Diebold U, Di Valentin C, Selloni A (2004) Surface structure of TiO_2(011)-(2×1). Phys Rev Lett 93(3):036104/036101–036104/036104

Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98(7):5648–5652

Berger T, Sterrer M, Diwald O, Knözinger E, Panayotov D, Thompson TL, Yates JT Jr (2005) Light-induced charge separation in anatase TiO_2 particles. J Phys Chem B 109(13):6061–6068

Chen X, Shen S, Guo L, Mao SS (2010) Semiconductor-based photocatalytic hydrogen generation. Chem Rev 110(11):6503–6570

Cheng H, Selloni A (2009a) Energetics and diffusion of intrinsic surface and subsurface defects on anatase TiO2(101). J Chem Phys 131(5):054703

Cheng H, Selloni A (2009b) Surface and subsurface oxygen vacancies in anatase TiO_2 and differences with rutile. Phys Rev B Condens Matter Mater Phys 79(9):092101

Cheng H, Selloni A (2010) Hydroxide ions at the water/anatase TiO_2(101) interface: structure and electronic states from first principles molecular dynamics. Langmuir 26(13):11518–11525

Cohen AJ, Mori-Sanchez P, Yang WT (2008) Insights into current limitations of density functional theory. Science 321(5890):792–794

De Angelis F, Di Valentin C, Fantacci S, Vittadini A, Selloni A (2014) Theoretical studies on anatase and less common TiO_2 phases: bulk, surfaces, and nanomaterials. Chem Rev 114:9708

Deskins NA, Dupuis M (2007) Electron transport via polaron hopping in bulk TiO_2: a density functional theory characterization. Phys Rev B 75(19):195212

Deskins NA, Rousseau R, Dupuis M (2011) Distribution of Ti^{3+} surface sites in reduced TiO_2. J Phys Chem C 115(15):7562–7572

Di Valentin C, Pacchioni G, Selloni A (2009) Reduced and n-type doped TiO_2: nature of Ti^{3+} species. J Phys Chem C 113(48):20543

Diebold U (2003) The surface science of titanium dioxide. Surf Sci Rep 48(5–8):53–229

Forro L, Chauvet O, Emin D, Zuppiroli L, Berger H, Lévy F (1994) High mobility n-type charge carriers in large single crystals of anatase (TiO_2). J Appl Phys 75(1):633–635

Gong X-Q, Khorshidi N, Stierle A, Vonk V, Ellinger C, Dosch H, Cheng H, Selloni A, He Y, Dulub O, Diebold U (2009) The 2×1 reconstruction of the rutile TiO_2(011) surface: a combined density functional theory, X-ray diffraction, and scanning tunneling microscopy study. Surf Sci 603(1):138–144

Graetzel M (2001) Photoelectrochemical cells. Nature 414(6861):338–344

Hagfeldt A, Boschloo G, Sun L, Kloo L, Pettersson H (2010) Dye-sensitized solar cells. Chem Rev 110(11):6595–6663

He Y, Dulub O, Cheng H, Selloni A, Diebold U (2009a) Evidence for the predominance of subsurface defects on reduced anatase TiO_2(101). Phys Rev Lett 102(10):106105

He Y, Tilocca A, Dulub O, Selloni A, Diebold U (2009b) Local ordering and electronic signatures of submonolayer water on anatase TiO_2(101). Nat Mater 8(7):585–589

Henrich VE, Dresselhaus G, Zeiger HJ (1976) Observation of two-dimensional phases associated with defect states on the surface of TiO_2. Phys Rev Lett 36(22):1335–1339

Herman GS, Sievers MR, Gao Y (2000) Structure determination of the two-domain (1×4) Anatase TiO_2(001) surface. Phys Rev Lett 84(15):3354–3357

Herman GS, Dohnalek Z, Ruzycki N, Diebold U (2003) Experimental investigation of the interaction of water and methanol with anatase-TiO_2(101). J Phys Chem B 107(12):2788–2795

Hoffmann MR, Martin ST, Choi WY, Bahnemann DW (1995) Environmental applications of semiconductor photocatalysis. Chem Rev 95(1):69–96

Hosseinpour S, Tang F, Wang F, Livingstone RA, Schlegel SJ, Ohto T, Bonn M, Nagata Y, Backus EHG (2017) Chemisorbed and physisorbed water at the TiO_2/water interface. J Phys Chem Lett 8(10):2195–2199

Hu S, Lewis NS, Ager JW, Yang J, McKone JR, Strandwitz NC (2015) Thin-film materials for the protection of semiconducting photoelectrodes in solar-fuel generators. J Phys Chem C 119(43):24201–24228

Kavan L, Graetzel M, Gilbert SE, Klemenz C, Scheel HJ (1996) Electrochemical and photoelectrochemical investigation of single-crystal anatase. J Am Chem Soc 118(28):6716

Khomenko VM, Langer K, Rager H, Fett A (1998) Electronic absorption by Ti^{3+} ions and electron delocalization in synthetic blue rutile. Phys Chem Miner 25(5):338–346

Lazzeri M, Selloni A (2001) Stress-driven reconstruction of an oxide surface. The anatase TiO_2(001)-(1 × 4) surface. Phys Rev Lett 87(26):266105/266101–266105/266104

Lazzeri M, Vittadini A, Selloni A (2001) Structure and energetics of stoichiometric TiO$_2$ anatase surfaces. Phys Rev B Condens Matter Mater Phys 63(15):155409/155401–155409/155409

Li Y-F, Selloni A (2013) Theoretical study of interfacial electron transfer from reduced anatase TiO$_2$(101) to adsorbed O$_2$. J Am Chem Soc 135(24):9195–9199

Li M, Hebenstreit W, Diebold U, Tyryshkin AM, Bowman MK, Dunham GG, Henderson MA (2000) The influence of the bulk reduction state on the surface structure and morphology of rutile TiO$_2$(110) single crystals. J Phys Chem B 104(20):4944–4950

Li Y-F, Liu Z-P, Liu L, Gao W (2010) Mechanism and activity of photocatalytic oxygen evolution on Titania Anatase in aqueous surroundings. J Am Chem Soc 132(37):13008–13015

Livraghi S, Chiesa M, Paganini MC, Giamello E (2011) On the nature of reduced states in titanium dioxide as monitored by electron paramagnetic resonance. I: the anatase case. J Phys Chem C 115(51):25413–25421

Livraghi S, Rolando M, Maurelli S, Chiesa M, Paganini MC, Giamello E (2014) Nature of reduced states in titanium dioxide as monitored by electron paramagnetic resonance. II: Rutile and Brookite cases. J Phys Chem C 118(38):22141–22148

Ma Y, Wang X, Jia Y, Chen X, Han H, Li C (2014) Titanium dioxide-based nanomaterials for photocatalytic fuel generations. Chem Rev 114(19):9987–10043

Moser S, Moreschini L, Jacimovic J, Barisic OS, Berger H, Magrez A, Chang YJ, Kim KS, Bostwick A, Rotenberg E, Forro L, Grioni M (2013) Tunable polaronic conduction in anatase TiO$_2$. Phys Rev Lett 110(19):196403

Ohno T, Sarukawa K, Matsumura M (2002) Crystal faces of rutile and anatase TiO$_2$ particles and their roles in photocatalytic reactions. New J Chem 26(9):1167–1170

Onda K, Li B, Petek H (2004) Two-photon photoemission spectroscopy of TiO$_2$(110) surfaces modified by defects and O$_2$ or H$_2$O adsorbates. Phys Rev B 70(4):045415

Onida G, Reining L, Rubio A (2002) Electronic excitations: density-functional versus many-body Green's-function approaches. Rev Mod Phys 74(2):601–659

Pang CL, Lindsay R, Thornton G (2013) Structure of clean and adsorbate-covered single-crystal rutile TiO$_2$ surfaces. Chem Rev 113(6):3887–3948

Perdew JP, Ernzerhof M, Burke K (1996) Rationale for mixing exact exchange with density functional approximations. J Chem Phys 105(22):9982–9985

Rajh T, Dimitrijevic NM, Bissonnette M, Koritarov T, Konda V (2014) Titanium dioxide in the service of the biomedical revolution. Chem Rev 114(19):10177–10216

Ramamoorthy M, Vanderbilt D, King-Smith RD (1994) First-principles calculations of the energetics of stoichiometric TiO$_2$ surfaces. Phys Rev B 49(23):16721–16727

Rödel TC, Fortuna F, Bertran F, Gabay M, Rozenberg MJ, Santander-Syro AF, Le Fèvre P (2015) Engineering two-dimensional electron gases at the (001) and (101) surfaces of TiO$_2$ anatase using light. Phys Rev B 92(4):041106(R)

Scheiber P, Fidler M, Dulub O, Schmid M, Diebold U, Hou W, Aschauer U, Selloni A (2012) (Sub)surface mobility of oxygen vacancies at the TiO$_2$ anatase (101) surface. Phys Rev Lett 109(13):136103

Schneider J, Matsuoka M, Takeuchi M, Zhang J, Horiuchi Y, Anpo M, Bahnemann DW (2014) Understanding TiO$_2$ photocatalysis: mechanisms and materials. Chem Rev 114(19):9919–9986

Selçuk S, Selloni A (2013) Surface structure and reactivity of anatase TiO$_2$ crystals with dominant {001} facets. J Phys Chem C 117(12):6358–6362

Selçuk S, Selloni A (2014) Influence of external electric fields on oxygen vacancies at the anatase (101) surface. J Chem Phys 141:084705

Selcuk S, Selloni A (2016) Facet-dependent trapping and dynamics of excess electrons at anatase TiO$_2$ surfaces and aqueous interfaces. Nat Mater 15(10):1107–1112

Setvin M, Aschauer U, Scheiber P, Li Y-F, Hou W, Schmid M, Selloni A, Diebold U (2013) Reaction of O$_2$ with subsurface oxygen vacancies on TiO$_2$ anatase (101). Science 341(6149):988–991

Setvin M, Franchini C, Hao X, Schmid M, Janotti A, Kaltak M, Van de Walle CG, Kresse G, Diebold U (2014a) Direct view at excess electrons in TiO$_2$ rutile and anatase. Phys Rev Lett 113(8):086402

Setvin M, Hao X, Daniel B, Pavelec J, Novotny Z, Parkinson GS, Schmid M, Kresse G, Franchini C, Diebold U (2014b) Charge trapping at the step edges of TiO$_2$ anatase (101). Angew Chem Int Ed 53(18):4714–4716

Sumita M, Hu C, Tateyama Y (2010) Interface water on TiO$_2$ anatase (101) and (001) surfaces: first-principles study with TiO$_2$ slabs dipped in bulk water. J Phys Chem C 114(43):18529–18537

Sun C, Liu L-M, Selloni A, Lu GQ, Smith SC (2010) Titania-water interactions: a review of theoretical studies. J Mater Chem 20(46):10319–10334

Tachikawa T, Yamashita S, Majima T (2011) Evidence for crystal-face-dependent TiO$_2$ photocatalysis from single-molecule imaging and kinetic analysis. J Am Chem Soc 133(18):7197–7204

Thomas AG, Flavell WR, Mallick AK, Kumarasinghe AR, Tsoutsou D, Khan N, Chatwin C, Rayner S, Smith GC, Stockbauer RL, Warren S, Johal TK, Patel S, Holland D, Taleb A, Wiame F (2007) Comparison of the electronic structure of anatase and rutile TiO$_2$ single-crystal surfaces using resonant photoemission and x-ray absorption spectroscopy. Phys Rev B 75(3):035105

Tilocca A, Selloni A (2004) Vertical and lateral order in adsorbed water layers on anatase TiO$_2$(101). Langmuir 20(19):8379–8384

Torrelles X, Cabailh G, Lindsay R, Bikondoa O, Roy J, Zegenhagen J, Teobaldi G, Hofer WA, Thornton G (2008) Geometric structure of TiO$_2$(011)(2x1). Phys Rev Lett 101(18):185501

Trail J, Monserrat B, López Ríos P, Maezono R, Needs RJ (2017) Quantum Monte Carlo study of the energetics of the rutile, anatase, brookite, and columbite TiO$_2$ polymorphs. Phys Rev B 95(12):121108(R)

Treacy JPW, Hussain H, Torrelles X, Grinter DC, Cabailh G, Bikondoa O, Nicklin C, Selcuk S, Selloni A, Lindsay R, Thornton G (2017) Geometric structure of anatase TIO$_2$(101). Phys Rev B 95(7):075416

Vittadini A, Selloni A, Rotzinger FP, Graetzel M (1998) Structure and energetics of water adsorbed at TiO$_2$ anatase (101) and (001) surfaces. Phys Rev Lett 81(14):2954–2957

Wang Z, Wen B, Hao Q, Liu L-M, Zhou C, Mao X, Lang X, Yin W-J, Dai D, Selloni A, Yang X (2015) Localized excitation of Ti^{3+} ions in the photoabsorption and photocatalytic activity of reduced rutile TiO$_2$. J Am Chem Soc 137(28):9146–9152

Yu J, Low J, Xiao W, Zhou P, Jaroniec M (2014) Enhanced photocatalytic CO$_2$-reduction activity of anatase TiO$_2$ by coexposed {001} and {101} facets. J Am Chem Soc 136(25):8839–8842

Zhang HZ, Banfield JF (1998) Thermodynamic analysis of phase stability of nanocrystalline titania. J Mater Chem 8(9):2073–2076

Zhang Y, Payne DT, Pang CL, Fielding HH, Thornton G (2015) Non-band-gap photoexcitation of hydroxylated TiO$_2$. J Phys Chem Lett 6(17):3391–3395

Modeling Disordered and Nanostructured Graphene

5

Oleg V. Yazyev

Contents

Abstract

Being isolated for the first time in 2004, graphene has started an entire new field of two-dimensional (2D) materials driven by novel properties and phenomena observed in these materials as well as their potential technological applications. Due to their low-dimensional nature, the effects of lattice disorder and boundaries in 2D materials are enhanced in comparison to bulk materials. This chapter covers the most important contributions of computational modeling to predict the structure and properties of lattice imperfections in graphene, namely, the point defects, topological defects, and edges, and draws connections to the subsequent experimental works confirming these predictions.

1   Introduction

The first isolation of *graphene*, a single layer of carbon atoms obtained by micromechanical cleavage of graphite (Novoselov et al. 2004, 2005; Zhang et al.

O. V. Yazyev (✉)

Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
e-mail: oleg.yazyev@epfl.ch

2005), started an entire new direction of research into two-dimensional (2D) materials extending from condensed matter physics to numerous other disciplines and promising novel technological applications. One particularly striking quality of graphene is its simplicity. The crystal structure of this material represents the bipartite honeycomb lattice with two equivalent carbon atoms per unit cell in sp^2 hybridization (Fig. 1a). The low-energy electronic properties of this material are dominated by the two π symmetry bands formed by the p_z atomic orbitals of carbon atoms (Fig. 1c). The bands form two conical intersection features in the corners of the Brillouin zone (points K and K′; Fig. 1d). These peculiar band dispersions called Dirac cones are responsible for many extraordinary electronic properties discovered in graphene (Castro Neto et al. 2009). Even the simplest theoretical descriptions, such as the one-orbital tight-binding model with only nearest-neighbor hopping integrals, are able to describe this feature accurately. Even though numerous other 2D materials have been discovered by now, graphene remains the most intriguing member of the family.
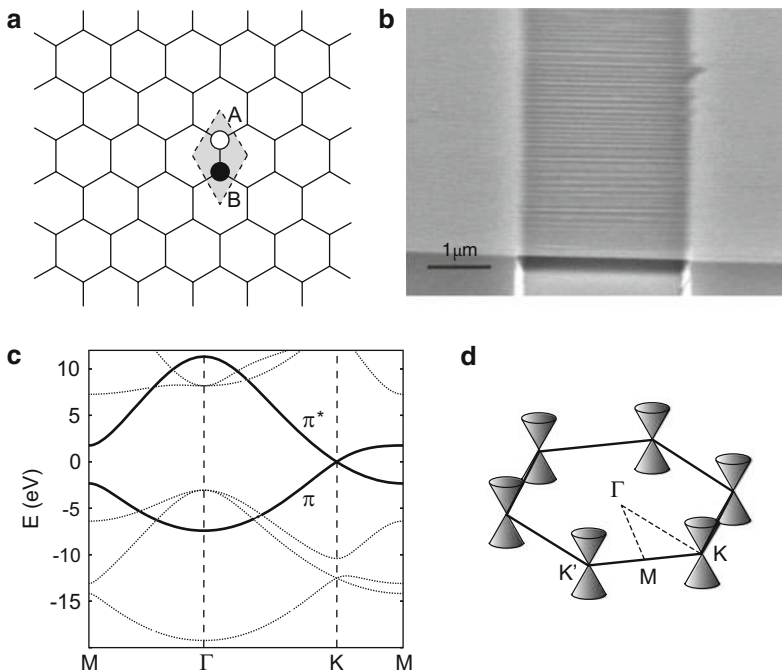
Fig. 1 (**a**) Two-dimensional honeycomb lattice of graphene showing the unit cell with two carbon atoms belonging to sublattices A and B. (**b**) Scanning electron microscopy (SEM) image of graphene bilayer freely suspended over a 2-micrometer-wide trench (Bao et al. 2009). (**c**) Band structure of graphene plotted along a high-symmetry k point path with the π symmetry bands highlighted. (**d**) Schematic representation of the band structure showing only the Dirac cone features in the corners of the Brillouin zone (points K and K′)

Despite the basic statistical physics arguments forbidding long-range crystalline order in two dimensions (Mermin 1968), relatively large samples of graphene can nevertheless be handled using basic laboratory equipment (Fig. 1b). Yet, reduced dimensionality of this material implies that structural imperfections, as well boundaries, have particularly strong effects on its physical properties. Furthermore, modern experimental techniques, in particular transmission electron microscopy (TEM) and scanning tunneling microscopy (STM), not only enabled the direct observation of structure and dynamics of defects in graphene at atomic resolution but also allowed their versatile creation and manipulation. Surprisingly, being a material with very simple crystal structure, graphene revealed an astonishing diversity of defects.

The purpose of this chapter is to give a brief overview of the most important contributions of computational modeling to predict the structure and properties of defects and boundaries in graphene that were consequently confirmed by the experiments. Without going into details of computational methods as such, we will consider the models of point defects, topological defects, and edges in graphene and draw a connection with the experiments.

2   Point Defects

Point defects in graphite, the bulk parent material of graphene, have been the subject of considerable research interest well before the first experimental isolation of this 2D material. Numerous modeling works have been performed in the context of understanding point defects that are created upon irradiation of graphite that is used as a moderator material in nuclear reactors. Even though graphite is an elemental material with a relative simple crystal structure, a very diverse family of point defects was proposed. These point defects include interstitials and vacancies as primary defects created upon radiation damage, their aggregates, metastable interstitial-vacancy pairs (Frenkel pairs) (Telling et al. 2003; Ewels et al. 2003; Akhmerov and Beenakker 2008), as well as defects produced by the rearrangement of carbon-carbon bonds within individual graphene layers (for an overview see Li et al. 2005). Further research interest was stimulated by the experimental observations of room-temperature ferromagnetism in graphite irradiated by high-energy protons (Esquinazi et al. 2003). It was suggested that defects play a crucial role in the induced magnetic ordering. This observation promised the possibility of creating novel magnetic materials based on light elements belonging to the *sp* block of the periodic table.

In atomically thin graphene, many predicted defects were directly observed using transmission electron microscopy (TEM). Not only this experimental tool allowed studying the structure of these defects at atomic resolution but it also allowed creating these defects and studying their transformations and migration in out-of-equilibrium conditions. Figure 2 reproduces TEM images of representative point defects in monolayer graphene (Kotakoski et al. 2011). The two primary defects induced by high-energy electrons in the TEM conditions are the Stone-Wales
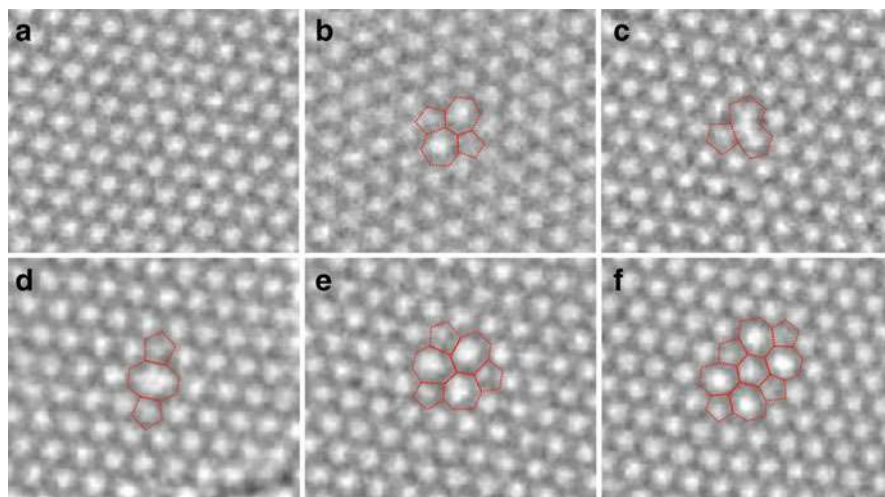
Fig. 2 Experimental observations of point defects in graphene. (**a**) Transmission electron microscopy (TEM) image of pristine graphene lattice. (**b**, **c**) TEM images of simplest point defects in graphene – the Stone-Wales defect (**c**) and the single-atom vacancy (**c**). (**d**) Double vacancy in graphene and (**e**, **f**) more complex defects derived from it by means of in-plane bond rotations. The red dashed lines outline the defect structures (Kotakoski et al. 2011)

defects (Fig. 2b) and the single-atom vacancies (Fig. 2c). The Stone-Wales defect is produced by means of a 90° in-plane rotation of a pair of carbon atoms followed by the reconnection of covalent bonds (Stone and Wales 1986). This results in a structure composed of two pentagons and two heptagons, while the coordination environment of each carbon atom remains unaffected. The calculations show that the formation energy of the Stone-Wales defect is 5 eV, while the energy barrier of bond rotation is approximately 10 eV (Li et al. 2005). Sputtering a single carbon atom creates a vacancy defect which reconstructs by forming a new covalent bond, while one carbon atom remains undercoordinated in agreement with theoretical predictions (El-Barbary et al. 2003). The formation energy of single-atom vacancy defect was predicted to be 7.5 eV. The double vacancy defect produced by removing two neighboring carbon atoms allows a reconstruction involving two new covalent bonds (Fig. 2d). Such a defect structure composed of two pentagons and an octagon has no undercoordinated atoms. The calculated formation energy of this defect is 8 eV, comparable to the one of the single-atom vacancy (Krasheninnikov et al. 2006). However, the formation energy per removed carbon atom is significantly lower for the double vacancy defect, which reflects the tendency of vacancy aggregation in graphene. The energy of this defect can be further lowered by a series of in-plane bond rotations leading to more complex defect structures composed of pentagons and heptagons. Calculations show that the observed point defects in Fig. 2e, f both have lower formation energies than the initial double vacancy defect (Lee et al. 2005; Kotakoski et al. 2011).

The single-atom vacancy defect in graphene deserves to be discussed in more detail. It was shown by means of first-principles calculations that the dangling bond of the undercoordinated carbon atom at the vacancy defect carries a local magnetic moment due to an unpaired σ electron (Lehtinen et al. 2004). This could account for the observed ferromagnetic ordering in irradiated graphite if sufficiently strong magnetic exchange coupling between such localized magnetic moments exists. Even more interesting are the so-called quasilocalized electronic states induced by single-atom vacancies. These resonant non-bonding states result from the local breaking of the bipartite symmetry of graphene lattice (Pereira et al. 2006). The vacancy defect removes a carbon atom from one of the sublattices, and this gives rise to a localized π-electron state in the complementary sublattice. Such quasilocalized states have been observed in STM images as very distinct triangular features around point defect on the surface of graphite (Mizes and Foster 1989), while the corresponding dI/dV measurements exhibit at sharp peak at the Fermi level (Ugeda et al. 2010). First-principles calculations show that these states indeed form a narrow band at the Fermi level that undergoes spin polarization in order to relieve the electronic instability associated with high density of states at the Fermi level (Yazyev and Helm 2007). This suggests a novel mechanism of itinerant-like magnetism induced by disorder in graphene-based materials. Interestingly, the coupling between quasilocalized in the same sublattice is ferromagnetic, while it is antiferromagnetic between the opposite sublattices (Yazyev 2008). This points at a possibility of tailoring the type of magnetic ordering in disordered graphene by controlling the distribution of defects in the two sublattices.

A closely related imperfection in graphene is the hydrogen chemisorption defect. A hydrogen atom binds covalently to one carbon atom of graphene lattice causing its transformation from the sp^2 to sp^3 hybridization. The latter is no longer able to provide a p_z atomic orbital to the π-electron system of graphene and hence is equivalent to a single-atom vacancy. For the reasons described above, hydrogen chemisorption defects also induce quasilocalized states. However, these simple defects are free of symmetry-breaking reconstructions and can be easily introduced to or removed from graphene. These properties established the hydrogen chemisorption defect as the reference disorder in graphene. A number of recent studies modeled the effect of such defects on the electronic transport properties of graphene (Wehling et al. 2010; Gargiulo et al. 2014). Similar to the vacancies, hydrogen chemisorption defects induce magnetic ordering with a clear physical picture (Yazyev and Helm 2007). Figure 3a shows the characteristic triangle-shaped distribution of spin density around a hydrogen chemisorption defect with a nominal magnetic moment of 1 Bohr magneton. Figure 3b reproduces the density of states for spin-up and spin-down channels that clearly illustrates the fully spin-split narrow band at the Fermi level. Experimentally, the quasilocalized states due to hydrogen chemisorption defects in graphene have recently been investigated using STM revealing the presence of sharp resonances (Fig. 3c) and their splitting consistent with the predicted spin polarization (Fig. 3d; González-Herrero et al. 2016).
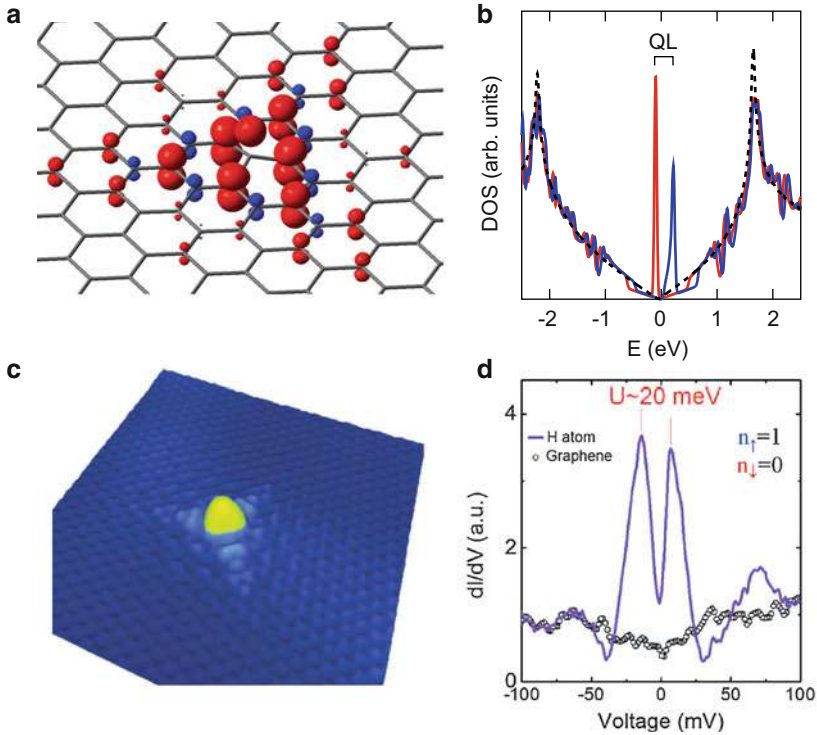
Fig. 3 Hydrogen chemisorption defect in graphene. (**a**) Spin density distribution and (**b**) the spin-resolved density of states obtained from first-principles calculations for the hydrogen chemisorption defect (Yazyev and Helm 2007). Red and blue correspond to majority spin and minority spin channels. (**c**) Scanning tunneling microscopy (STM) image of the quasilocalized state at the hydrogen chemisorption defect on graphene and (**d**) the corresponding experimental *dI/dV* curve showing the predicted spin splitting (González-Herrero et al. 2016)

3 Topological Defects

Topological defects are distinct from point defects as these types of disorder are characterized by certain quantities defined nonlocally and hence do not change upon a local modification of the atomic structure of such defects. These quantities are further referred to as the structural topological invariants. An example of a topological defect is the *dislocation*, and the corresponding topological invariant is its Burgers vector. The nonlocality of a dislocation stems from the fact that the Burgers vector can be defined on an arbitrarily large contour encircling the dislocation core; therefore any local structural modification to it would not change the Burgers vector. Another relevant type of topological defects is the *grain boundary*, the interface between two single-crystalline domains of same material.

Orientation of the crystalline lattice changes across the grain boundary giving rise to the notion of misorientation angle, the corresponding topological invariant. Grain boundaries are the defects inherent to polycrystalline materials. Both dislocations and grain boundaries in bulk materials are extensively covered in practically all materials science textbooks.

As already mentioned above, 2D materials in general and graphene in particular are very susceptible to wide range of structural disorder. Hence, one can expect that obtaining large single crystals of 2D materials would be a challenging task, while polycrystalline morphologies of 2D materials should be abundant. In addition, the space of topological invariants of the corresponding defects in 2D materials is "narrower" than in the case of bulk, three-dimensional materials. For instance, the Burgers vector of a dislocation is constrained to lie in the plane of 2D crystalline lattice; hence only edge dislocations are possible in 2D materials. Similarly, only tilt grain boundaries exist in 2D materials. Dislocations and grain boundaries in 2D materials can be regarded as point-like and one-dimensional objects, respectively.

In order to proceed with the construction of dislocations and grain boundaries in graphene, we will borrow from soft matter community a third type of topological defect called the disclination (Seung and Nelson 1988). A disclination is the result of either removing from or adding to the ideal lattice a semi-infinite wedge of material. The corresponding topological invariant is the disclination strength s, which measures the wedge angle. Wedge removal results in positive ($s > 0$) disclinations while wedge introduction – in negative ($s < 0$). The particular cases of $s = 60°$ and $s = -60°$ disclinations in graphene appear as pentagons and heptagons embedded in the array of hexagons forming the honeycomb lattice as shown in Fig. 4.

The result of combining two complementary disclinations is the dislocation (Seung and Nelson 1988), and its Burgers vector depends on the distance between the pentagon and the heptagon. Two examples of dislocations in graphene are shown in Fig. 4. The pentagon-heptagon pair sharing an edge is the dislocation with the smallest Burgers vector $b = (1, 0)$ equal to the lattice constant of graphene (2.46 Å). Above, we defined the Burgers vectors as a pair of integers in terms of lattice vectors, similar to the chirality indices of carbon nanotubes. The second example shown in Fig. 4 corresponds to the second smallest Burgers vector $b = (1, 1)$ (or 4.26 Å) (Yazyev and Louie 2010a).

Finally, grain boundaries can be viewed as arrays of dislocations following the model introduced by Read and Shockley (1950). The misorientation angle θ is related to the Burgers vector of constituent dislocations and the distance between them. Two examples of symmetric grain boundaries in graphene are shown in Fig. 4. These two grain boundaries are formed by particularly dense arrangements of $b = (1, 0)$ dislocations and correspond to misorientation angles $\theta = 21.8°$ and $\theta = 32.3°$, respectively (Yazyev and Louie 2010a).

Interestingly, dislocations characterized by all allowed Burgers vectors and grain boundaries spanning the entire range of misorientation angles $0° < \theta < 60°$ can be constructed starting from pentagons and heptagons only. All carbon atoms in defect structures built following the described approach maintain their original threefold
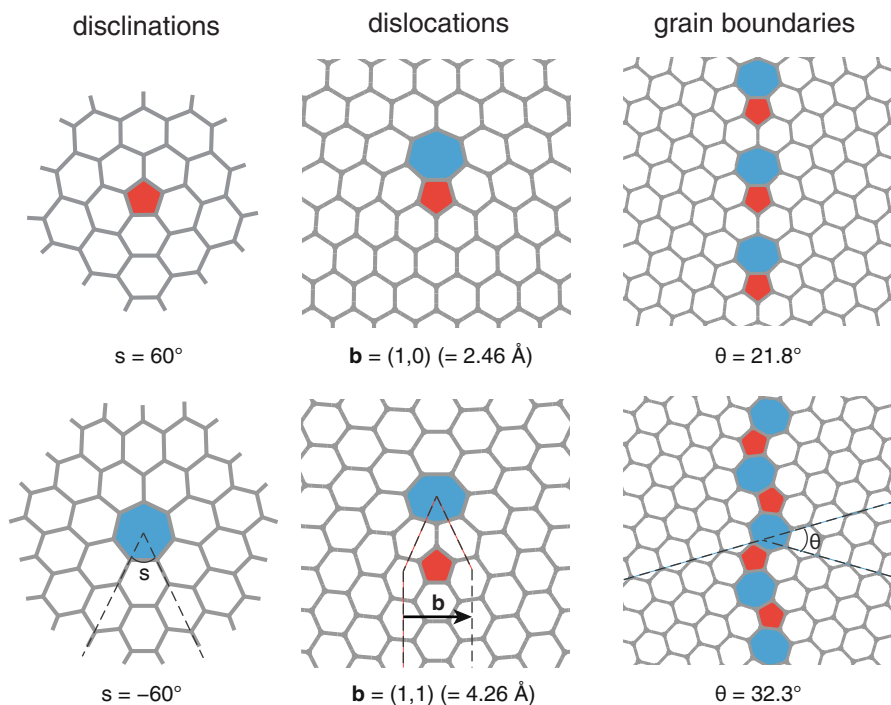
Fig. 4 Three hierarchically related types of topological defects in graphene: disclinations, disloca-
tions, and grain boundaries. The definitions of the corresponding structural topological invariants
(disclination strength *s*, Burgers vector **b**, and misorientation angle *θ*) and their values are given.
The structures of all three types of topological defects are composed of pentagons (red) and
hexagons (blue) (Yazyev and Louie 2010a; Yazyev and Chen 2014)

coordination, no dangling bonds or overcoordinated atoms have been introduced,
and hence the defects are expected to have low formation energies by construction.
This hierarchical construction introduced by us in 2010 (Yazyev and Louie 2010a),
almost simultaneously with the similar work from the group of Boris Yakobson (Liu
and Yakobson 2010), opened a possibility to investigate computationally realistic
models of topological defects in graphene ahead of their experimental observation.

First experimental studies of topological defects in graphene followed less than
a year after our theoretical predictions (Huang et al. 2011; Kim et al. 2011; An
et al. 2011; followed by many other). Figure 5a shows the dark-field TEM image of
epitaxial graphene grown using chemical vapor deposition (CVD) on Cu substrate
(Huang et al. 2011). The image reveals an intricate polycrystalline morphology with
micrometer-size crystallites. High-resolution TEM further exposes the atomic-scale
details of grain boundaries in these epitaxial graphene samples. The segment of
large-angle grain boundary shown in Fig. 5b shares many similarities with the mod-
els in Fig. 4. Specifically, it is composed of an alternating chain of pentagons and
heptagons with no under- or overcoordinated carbon atoms whatsoever. However,
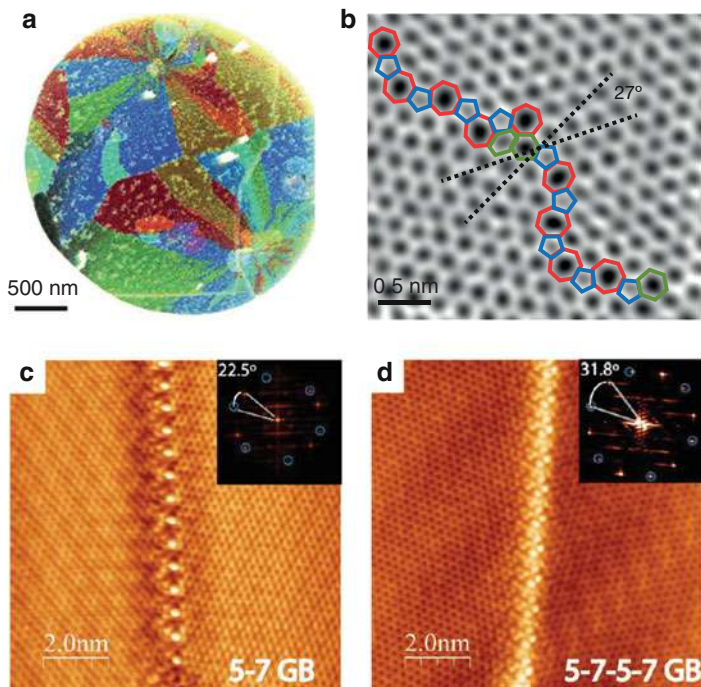
Fig. 5 Experimental observations of grain boundaries in graphene. (**a**) Dark-field TEM image color-coded according to the orientation of the crystalline lattice reveals the polycrystalline morphology of CVD-grown epitaxial graphene with micrometer-size single-crystalline domains. (**b**) High-resolution TEM image shows atomic structure details of the grain boundary as an alternating chain of pentagons and heptagons (Huang et al. 2011). (**c**, **d**) STM images of grain boundaries in thermally annealed epitaxial graphene showing highly periodic patterns that can be attributed to the grain boundary models shown in Fig. 4 (Yang et al. 2014)

unlike the models in Fig. 4, the observed structure is not linear and lacks clear periodicity. This discrepancy can be explained by assuming that the performed CVD growth process does not allow achieving thermodynamic equilibrium in full. In fact, thermal annealing of such epitaxial graphene samples leads to highly periodic structures as exemplified in Fig. 5c, d (Yang et al. 2014). Moreover, the two grain boundaries shown in Fig. 5c, d correspond exactly to the models in Fig. 4. It is striking that until this moment all predictions confirmed by experiments were done by means of modeling, but without doing any explicit computations. Further considerations, for example, rationalizing the stability of these two grain boundary structures would necessitate computer simulations.

Figure 6a reproduces the formation energies per unit length γ as a function of misorientation angle θ calculated using density functional theory for a number of periodic models of grain boundaries in graphene (Yazyev and Louie 2010a). Filled symbols correspond to models that were artificially constrained to be flat. One can
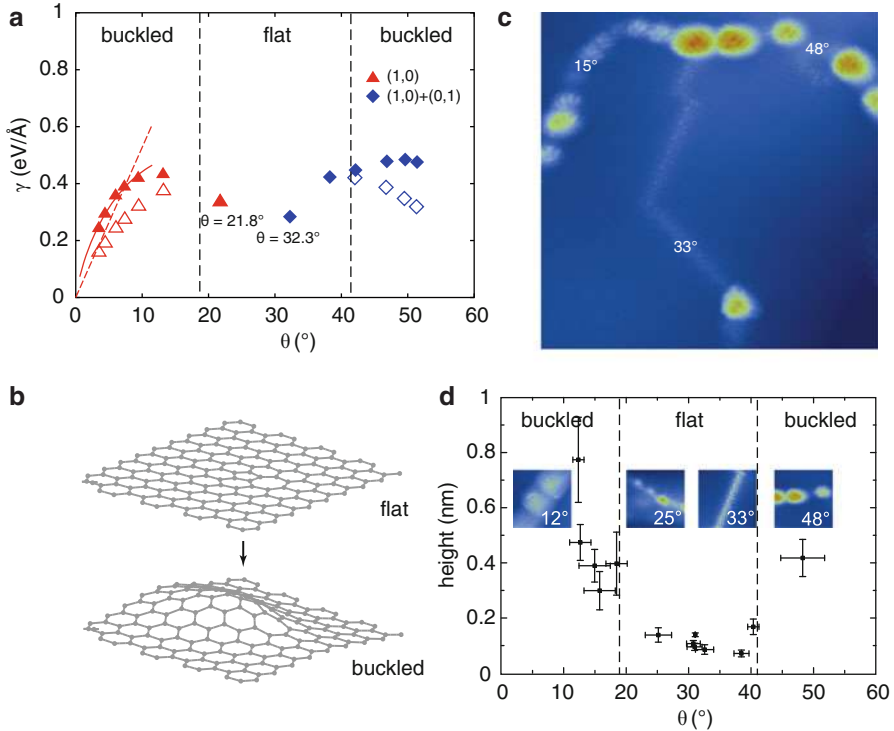
Fig. 6 Formation energies and buckling of grain boundaries in graphene. (**a**) Calculated grain boundary formation energy per unit length γ as a function misorientation angle θ. Formation energies of flat and buckled models are shown as filled and empty symbols. The vertical dashed lines delimit the buckled and flat regimes (Yazyev and Louie 2010a). (**b**) Illustration of the buckling deformation produced by a single $b = (1, 0)$ dislocation. (**c**) STM image of epitaxial graphene grown on the C-face of SiC showing three grain boundaries with their corresponding misorientation angles indicated (Tison et al. 2014). (**d**) Apparent height of the observed grain boundaries as a function misorientation angle θ. The flat and buckled regimes with transition angle $\theta_c = 19 \pm 2°$ are clearly distinguished (Tison et al. 2014)

recognize a clear trend with a minimum at ca. 30° and the two large-angle grain boundary models introduced in Fig. 4 ($\theta = 21.8°$ and $\theta = 32.3°$) having particularly low formation energies, which agrees with their frequent experimental observation (see, e.g., Fig. 5c, d). This picture is similar to what is usually observed for bulk materials. The existence of a formation energy minimum at large values of θ can be rationalized by the fact that in large-angle grain boundaries the distance between the constituent dislocations is small; hence the in-plane elastic fields produced by the dislocations effectively cancel each other reducing the contribution of elastic energy to the defect formation energy.

By allowing the atoms to displace out of plane, we enable a physical mechanism which is not present in bulk materials. The fact that the dimensionality of graphene,

or any other 2D material, is one less compared to the dimensionality of host space gives an additional degree of freedom. Displacement in this extra dimension, below referred to as buckling, effectively relieves the in-plane compressive strain. Figure 6b illustrates the buckling pattern produced by a single $b = (1, 0)$ dislocation in graphene, and the experimentally observed pictures appear to be similar (Lehtinen et al. 2013). Such a screening of the in-plane elastic fields gives rise to a mechanism that competes with the abovementioned mutual cancellation of elastic fields produced by closely spaces dislocations. Indeed, calculations show that apart from several large-angle models buckling significantly lowers the grain boundary formation energies (empty symbols in Fig. 6a) (Yazyev and Louie 2010a). The transition from flat to buckled configurations, called the buckling transition, takes place at a well-defined critical misorientation angle (Carraro and Nelson 1993). The theoretically predicted buckling transition in graphene was experimentally observed in epitaxial graphene samples grown on SiC(000$\bar{1}$) substrate (Tison et al. 2014). Figure 6c reproduces a typical STM image that shows three single-crystalline domains separated by a junction of three grain boundaries. While the ordered large-angle grain boundary ($\theta = 33°$) shows little topographic contrast, the two small-angle grain boundaries ($\theta = 15°$ and $\theta = 48°$) display pronounced buckling, and the location of individual dislocations can be clearly tracked. All observed line defects can be distinguished as either flat or buckled grain boundaries (Fig. 6d). The measured buckling transition angle $\theta_c = 19 \pm 2°$ (and the symmetrically reflected value $\theta_c = 41 \pm 2°$) is in quantitative agreement with the results of DFT calculations (dashed vertical lines in Fig. 6a; Yazyev and Louie 2010a).

The above demonstrates spectacular success of modeling in predicting novel defect structures and achieving quantitative agreement with the experiments. Many physical phenomena related to topological defects in graphene have been studied computationally beyond the ones described above. For instance, electronic transport is of special importance because graphene is considered as a possible key material in the future information and communication technologies. Novel transport phenomena resulting from the periodic structure of grain boundaries such as the controllable transport gaps (Yazyev and Louie 2010b) and valley filtering (Gunlycke and White 2011; Chen et al. 2014) have been discussed. Thermal transport and mechanical properties of polycrystalline graphene are two more examples of properties relevant to practical applications (for review see Yazyev and Chen 2014). Other types of one-dimensional topological defects resulting from stacking order in multilayer graphene systems are possible. The so-called stacking domain boundaries have been observed, modeled, and extensively discussed in the context of topologically nontrivial electronic states (Alden et al. 2013; Butz et al. 2014; San-Jose et al. 2014; Ju et al. 2015; Gargiulo and Yazyev 2018). Finally, dislocations, grain boundaries, and polycrystalline morphologies are equally relevant to other 2D materials, such as monolayer transition metal dichalcogenides (TMDs). Due to the binary nature of these materials, however, dislocations and grain boundaries show a broader diversity in terms of atomic structure and local chemical composition (van der Zande et al. 2013; Najmaei et al. 2013; Zhang et al. 2013b).

4   Edges and Nanoribbons

Since the very first isolation of graphene, the absence of band gap in this material was recognized as the most important challenge on the way of realizing its potential in electronics. In particular, sizable band gaps are required for manufacturing field-effect transistors with on/off current ratios sufficient for applications in digital electronics. Exploiting quantum confinement in nanostructures based on graphene is probably the most evident route to controlling band gaps. Early works on top-down lithographically patterned nanostructures showed that the transport properties and deduced bands gaps indeed shows strong dependence on the nanostructure dimensions (Fig. 7a, b; Han et al. 2007). However, it also became clear that the local atomic structure details and the crystallographic orientation of the edges play even a more important role in defining the electronic properties of graphene
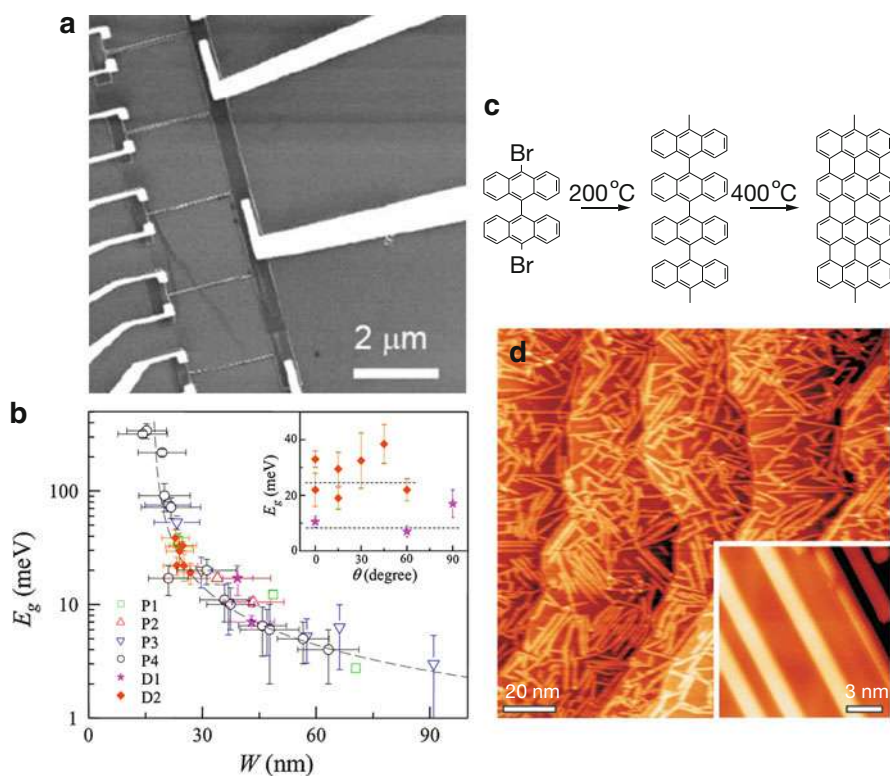
Fig. 7 Experimental approaches to manufacturing graphene nanostructures. (**a**) SEM image of a lithographically patterned device set containing multiple graphene nanoribbons of varying width (Han et al. 2007). (**b**) Dependence of the measured band gaps on the width and orientation of graphene nanoribbons (Han et al. 2007). (**c**) Scheme of the synthetic route leading to the $N = 7$ armchair nanoribbon (Cai et al. 2010). (**d**) STM image of the resulting $N = 7$ armchair graphene nanoribbons on Au(111) surface (Cai et al. 2010)

nanostructures. Common lithographic techniques are not able to provide the desired atomic precision; hence a broad range of alternative methods have been developed. It is worth highlighting the bottom-up self-assembly technique (Cai et al. 2010) covered in ▶ Chap. 31, "Electronic Structure of Atomically Precise Graphene Nanoribbons," by L. Talirz and C. Pignedoli. This approach relies on the surface-assisted self-assembly starting from precursor molecules, as, for instance, shown in Fig. 7c. It is possible to design molecular precursors that lead to uniquely defined nanostructures. An experimental STM image of highly ordered one-dimensional strips of graphene of 1 nm width produced by means of the synthetic route of Fig. 7c is shown in Fig. 7d. It is not our intention to cover in this chapter all proposed approaches for manufacturing graphene nanostructures. Instead, we will focus on reviewing the most important contributions from materials modeling to understanding the effects of reduced dimensions and *edges* on the electronic structure of nanostructured graphene.

The vast majority of computational studies on nanostructured graphene have been performed using one-dimensional periodic models commonly called *graphene nanoribbons*. Such models first introduced by Nakada et al. (1996) well before the first experimental isolation of graphene have become exceptionally popular. This is largely due to the fact that one-dimensional band structures of graphene nanoribbons proved to be a very insightful tool for analyzing the effects of reduced dimensions and edges. In addition, computationally tractable models of modest size could be constructed due to a small lattice constant of basic graphene nanoribbons. Such calculations have been performed using a broad range of techniques – from simple one-orbital nearest-neighbor tight-binding model to first-principles methods based on density functional theory or many-body perturbation theory. In the case of first-principles calculations, the outmost edge atoms are typically terminated by hydrogen atoms that provide a neutral from the point of view of electronic structure passivation of the dangling bonds. Among the short periodicity models, one distinguishes armchair and zigzag graphene nanoribbons that received their names from analogy with carbon nanotubes. The structure of an example of *armchair* graphene nanoribbon is shown in Fig. 8a, while in the case of *zigzag* graphene nanoribbon (Fig. 8b), the edge direction is rotated by 30° to be aligned with one of the lattice vectors of graphene.

The band structure of the armchair graphene nanoribbon in Fig. 7a calculated using the one-orbital tight-binding model with a constant nearest-neighbor transfer integral t (≈ 3 eV in graphene) shows the presence of a direct band gap. The density of states shows a number of van Hove singularities that associated with band extrema of a one-dimensional system. In the case of armchair nanoribbons, the band gaps Δ_a are due to the lateral confinement of the Dirac fermion of graphene. Importantly, the band gap magnitude shows an oscillating behavior as the nanoribbon width increases, changing between metallic and semiconducting band structures in the case of nearest-neighbor tight-binding model calculations. When edge effects are taken into account in first-principles calculations, a small band gap opens in metallic armchair graphene nanoribbons (Fig. 9a; Son et al. 2006a; Barone et al. 2006). Nevertheless, such a non-monotonic dependence of the
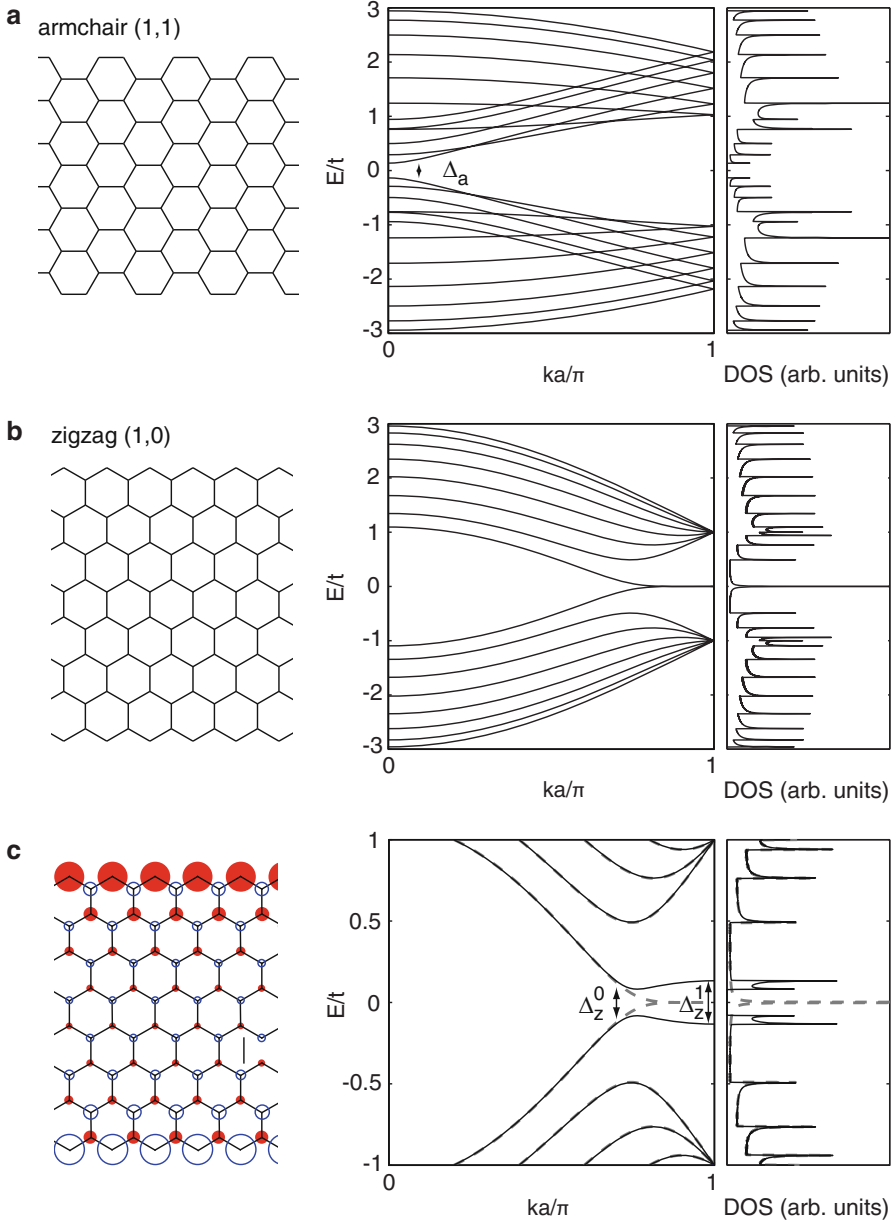
Fig. 8 Atomistic models of (**a**) an armchair graphene nanoribbon and (**b**) a zigzag graphene nanoribbon, their corresponding band structures and density of states plots calculated using the nearest-neighbor tight-binding model. (**c**) Local magnetic moments in the ground-state configuration of the zigzag graphene nanoribbon calculated using the mean-field Hubbard model ($U/t = 1$). The area of red (filled) and blue (empty) circles represent the magnitudes of spin-up and spin-down local magnetic moments, respectively. The corresponding band structure and density of states plots are compared to the tight-binding solutions (dashed lines)

band gap magnitude on the nanoribbon width demonstrates the need of achieving atomic-precision control of graphene nanostructures if electronic properties are to be controlled. Recent works reporting bottom-up synthesis of armchair graphene nanoribbons of different width showed band gaps in very good agreement with the results of first-principles calculations, provided these calculations are performed at the level of many-body perturbation theory with screening effects of the metallic substrate taken into account (Ruffieux et al. 2012).

In contrast, zigzag graphene nanoribbons show a dramatically different electronic structure with a pronounced flat-band segment at the Fermi level when the effect of electron-electron interactions is neglected (Fig. 8b). This flat band feature is due to electronic states localized at the edges, and hence it exhibits no pronounced dependence on the nanoribbon width. The diverging density of states at the Fermi level resulting from the edge states is intrinsically unstable. One evident mechanism for relieving this electronic instability via electron-electron interactions is by developing magnetic ordering (Fujita et al. 1996). Indeed, both the spin-polarized density functional theory and the simple mean-field Hubbard model calculations result in a gapped band structure (Fig. 8c). The ground state is a broken symmetry spin-polarized configuration with localized magnetic moments at the edges (Fig. 7c). The edge spins show ferromagnetic coupling along the edge and antiferromagnetic across the nanoribbon. The band gap Δ_z^0 is related to the coupling across the nanoribbon and therefore decreases with increasing the width. In contrast, the zone-boundary splitting Δ_z^1 is related to exchange coupling along the edge and hence shows only a weak dependence on the width (Fig. 9b; Son et al. 2006a). The theoretical prediction of edge magnetism in graphene has stimulated considerable research efforts resulting in a large number of interesting ideas. For instance, Son, Cohen, and Louie predicted that by applying electric field across such magnetic zigzag graphene nanoribbons it would be possible to close the gap selectively in
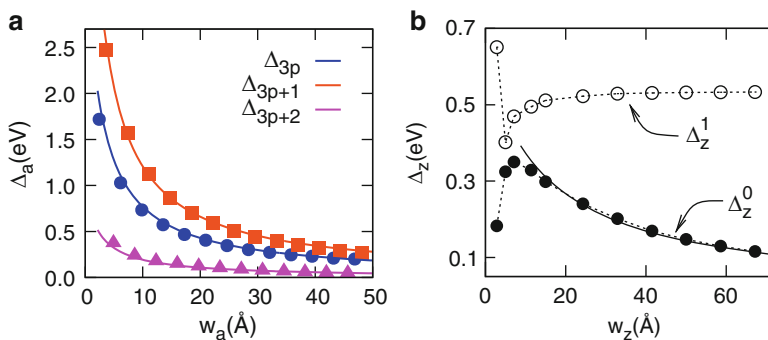
Fig. 9 (**a**) Band gaps of armchair graphene nanoribbons Δ_a as a function width w_a from local density approximation (LDA) calculations. The oscillating dependence of Δ_a is due to three families of armchair graphene nanoribbons. (**b**) Dependence of band gap Δ_z^0 and zone-boundary splitting Δ_z^1 on the width w_z of zigzag graphene nanoribbons obtained from spin-polarized LDA calculations (Son et al. 2006a)

one of the spin channels, in other words induce a half-metallic state (Son et al. 2006b). Such electric control of spin properties could find numerous applications in spintronic devices. On the other hand, one has to clearly understand the limitations of magnetism at graphene edges resulting from their reduced dimensionality and the intrinsic weak magnetic anisotropy of carbon (Yazyev and Katsnelson 2008). Producing well-ordered zigzag graphene nanoribbons experimentally has proven to be a very challenging task. Only very recently zigzag graphene nanoribbons were prepared using an STM-based etching method (Magda et al. 2014) and the already mentioned bottom-up self-assembly technique (Ruffieux et al. 2016). In both cases the observed electronic structure properties support the existence of spin-polarized edge states.

Armchair and zigzag graphene nanoribbons represent convenient models due to the orientation of the edges along the high-symmetry directions of the honeycomb lattice of graphene. One would immediately start wondering about the properties of graphene nanoribbons with edges oriented in an arbitrary way. It turns out that the electronic structure of such arbitrarily oriented nanoribbons, commonly referred to as the *chiral* graphene nanoribbons, resembles that of zigzag graphene nanoribbons. Both in zigzag and chiral nanoribbons, electronic states localized at the edges are a result of the local breaking of the sublattice symmetry of the bipartite lattice of graphene. In this respect the armchair edges represent a special case with no edge states. The electronic properties of chiral edges were addressed both analytically (Akhmerov and Beenakker 2008) and computationally (Nakada et al. 1996; Yazyev et al. 2011). In presence of electron-electron interactions, the localized states at chiral edges were predicted to undergo spin polarization with edge magnetic moment per unit length continuously spanning the cases of magnetic zigzag and nonmagnetic armchair edges as the edge orientation changes (Yazyev et al. 2011). Edge states were experimentally observed in chiral graphene nanoribbons produced by chemical unzipping of carbon nanotubes (Tao et al. 2011; Zhang et al. 2013a). The detailed STM images of such chiral edges show a clear contrast due to the localized edge states (Fig. 10a), while first-principles simulations (Fig. 10b) reproduce the experimentally observed picture allowing to identify the atomic structure of the edge in an atom-by-atom fashion (Zhang et al. 2013a).

5   Conclusions and Outlook

To summarize, this chapter has illustrated how first-principles modeling combined with experimental techniques capable of providing atomically resolved images (TEM and STM) allows to elucidate the structure of defects and edges in graphene in an atom-by-atom fashion. The same simulations complemented by theoretical models reveal the effect of these structural irregularities on the electronic and transport properties of graphene. Understanding the effects of disorder is crucial for designing novel functional materials and devices based on graphene. Further advance in predicting power is expected upon a continuing development of computational methods capable of treating large-scale realistic models, e.g.,
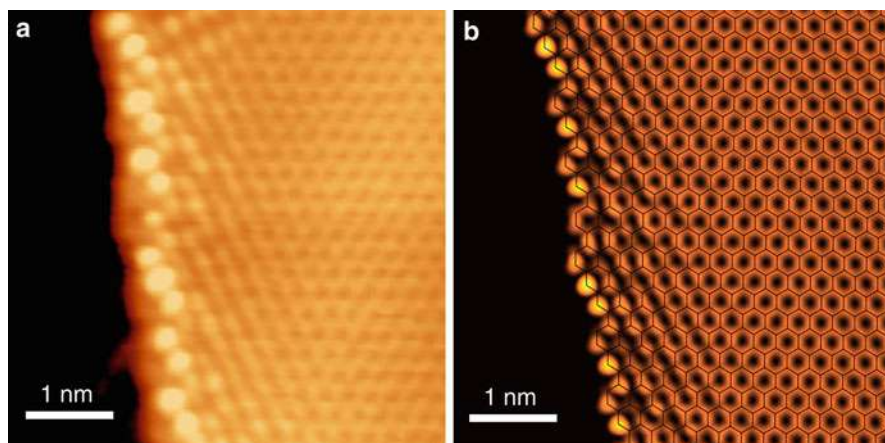
Fig. 10 (**a**) Experimentally observed and (**b**) simulated from first-principles STM images of an edge of a chiral graphene nanoribbon obtained by chemical unzipping of carbon nanotubes followed by hydrogen plasma etching (Zhang et al. 2013a). The contrast due to the localized edge states is clearly seen in both images

explicitly taking into account the substrate or performing device-scale simulations. The biggest challenge, however, lies in the ability to control experimentally the amount, structure, and possibly location, of the discussed structural irregularities. Overcoming this challenge can potentially result in a number of novel technological applications. Point defects in graphene are considered as a platform for magnetic materials and spintronic devices based on light elements. While the presence of topological defects such as grain boundaries is inevitable in large-area samples of 2D materials, it should be possible to minimize their undesirable effect on the electrical characteristics. At the same time, when introduced in a controlled way, such defects may serve as components of novel electronic devices. Finally, the abovementioned surface-assisted self-assembly approach to graphene nanoribbons and junctions can lead to a disruptive technology for manufacturing nanometer-scale all-graphene integrated circuits.

References

Akhmerov AR, Beenakker CWJ (2008) Boundary conditions for Dirac fermions on a terminated honeycomb lattice. Phys Rev B 77:085423

Alden JS et al (2013) Strain solitons and topological defects in bilayer graphene. Proc Natl Acad Sci 110:11256–11260

An J et al (2011) Domain (grain) boundaries and evidence of "Twinlike" structures in chemically vapor deposited grown graphene. ACS Nano 5:2433–2439

Bao W, Miao F, Chen Z, Zhang H, Jang W, Dames C, Lau CN (2009) Controlled ripple texturing of suspended graphene and ultrathin graphite membranes. Nat Nanotechnol 4:562–566

Barone V, Hod O, Scuseria GE (2006) Electronic structure and stability of semiconducting graphene nanoribbons. Nano Lett 6:2748

Butz B et al (2014) Dislocations in bilayer graphene. Nature 505:533–537

Cai J et al (2010) Atomically precise bottom-up fabrication of graphene nanoribbons. Nature 466:470–473

Carraro C, Nelson DR (1993) Grain-boundary buckling and spin-glass models of disorder in membranes. Phys Rev E 48:3082

Castro Neto AH, Guinea F, Peres NMR, Novoselov KS, Geim AK (2009) The electronic properties of graphene. Rev Mod Phys 81:109

Chen JH et al (2014) Controlled growth of a line defect in graphene and implications for gate-tunable valley filtering. Phys Rev B 89:121407

El-Barbary AA, Telling RH, Ewels CP, Heggie MI, Briddon PR (2003) Structure and energetics of the vacancy in graphite. Phys Rev B 68:144107

Esquinazi P, Spemann D, Höhne R, Setzer A, Han K-H, Butz T (2003) Induced magnetic ordering by proton irradiation in graphite. Phys Rev Lett 91:227201

Ewels CP, Telling RH, El-Barbary AA, Heggie MI, Briddon PR (2003) Metastable Frenkel pair defect in graphite: source of Wigner energy? Phys Rev Lett 91:025505

Fujita M, Wakabayashi K, Nakada K, Kusakabe K (1996) Peculiar localized state at zigzag graphite edge. J Phys Soc Jpn 65:1920–1923

Gargiulo F, Yazyev OV (2018) Structural and electronic transformation in low-angle twisted bilayer graphene. 2D Mater 5:015019

Gargiulo F et al (2014) Electronic transport in graphene with aggregated hydrogen adatoms. Phys Rev Lett 113:246601

González-Herrero H et al (2016) Atomic-scale control of graphene magnetism by using hydrogen atoms. Science 352:437–441

Gunlycke D, White CT (2011) Graphene valley filter using a line defect. Phys Rev Lett 106:136806

Han MY, Ozyilmaz B, Zhang Y, Kim P (2007) Energy band-gap engineering of graphene nanoribbons. Phys Rev Lett 98:206805

Huang PY et al (2011) Grains and grain boundaries in single-layer graphene atomic patchwork quilts. Nature 469:389–392

Ju L et al (2015) Topological valley transport at bilayer graphene domain walls. Nature 520:650–655

Kim K, Lee Z, Regan W, Kisielowski C, Crommie MF, Zettl A (2011) Grain boundary mapping in polycrystalline graphene. ACS Nano 5:2142–2146

Kotakoski J, Krasheninnikov AV, Kaiser U, Meyer JC (2011) From point defects in graphene to two-dimensional amorphous carbon. Phys Rev Lett 106:105505

Krasheninnikov AV, Lehtinen PO, Foster AS, Nieminen RM (2006) Bending the rules: contrasting vacancy energetics and migration in graphite and carbon nanotubes. Chem Phys Lett 418:132

Lee G-D, Wang CZ, Yoon E, Hwang N-M, Kim D-Y, Ho KM (2005) Diffusion, coalescence, and reconstruction of vacancy defects in graphene layers. Phys Rev Lett 95:205501

Lehtinen PO, Foster AS, Ma Y, Krasheninnikov AV, Nieminen RM (2004) Irradiation-induced magnetism in graphite: a density functional study. Phys Rev Lett 93:187202

Lehtinen O, Kurasch S, Krasheninnikov AV, Kaiser U (2013) Atomic scale study of the life cycle of a dislocation in graphene from birth to annihilation. Nat Commun 4:2098

Li L, Reich S, Robertson J (2005) Defect energies of graphite: density-functional calculations. Phys Rev B 72:184109

Liu Y, Yakobson BI (2010) Cones, pringles, and grain boundary landscapes in graphene topology. Nano Lett 10:2178–2183

Magda GZ et al (2014) Room-temperature magnetic order on zigzag edges of narrow graphene nanoribbons. Nature 514:608–611

Mermin ND (1968) Crystalline order in two dimensions. Phys Rev 176:250–254

Mizes HA, Foster JS (1989) Long-range electronic perturbations caused by defects using scanning tunneling microscopy. Science 244:559

Najmaei S et al (2013) Vapour phase growth and grain boundary structure of molybdenum disulphide atomic layers. Nat Mater 12:754–759

Nakada K, Fujita M, Dresselhaus G, Dresselhaus MS (1996) Edge state in graphene ribbons: nanometer size effect and edge shape dependence. Phys Rev B 54:17954

Novoselov KS et al (2004) Electric field effect in atomically thin carbon films. Science 306:666

Novoselov KS et al (2005) Two-dimensional gas of massless Dirac fermions in graphene. Nature 438:197–200

Pereira VM, Guinea F, Lopes dos Santos JMB, Peres NMR, Castro Neto AH (2006) Disorder induced localized states in graphene. Phys Rev Lett 96:036801

Read WT, Shockley W (1950) Dislocation models of crystal grain boundaries. Phys Rev 78:275

Ruffieux P et al (2012) Electronic structure of atomically precise graphene nanoribbons. ACS Nano 6:6930–6935

Ruffieux P et al (2016) On-surface synthesis of graphene nanoribbons with zigzag edge topology. Nature 531:489–492

San-Jose P, Gorbachev RV, Geim AK, Novoselov KS, Guinea F (2014) Stacking boundaries and transport in bilayer graphene. Nano Lett 14:2052–2057

Seung HS, Nelson DR (1988) Defects in flexible membranes with crystalline order. Phys Rev A 38:1005

Son Y-W, Cohen ML, Louie SG (2006a) Energy gaps in graphene nanoribbons. Phys Rev Lett 97:216803

Son Y-W, Cohen ML, Louie SG (2006b) Half-metallic graphene nanoribbons. Nature 444:347

Stone AJ, Wales DJ (1986) Theoretical studies of icosahedral C_{60} and some related species. Chem Phys Lett 128:501–503

Tao C et al (2011) Spatially resolving edge states of chiral graphene nanoribbons. Nat Phys 7:616–620

Telling RH, Ewels CP, El-Barbary AA, Heggie MI (2003) Wigner defects bridge the graphite gap. Nat Mater 2:333–337

Tison Y et al (2014) Grain boundaries in graphene on SiC(0001$\bar{1}$) substrate. Nano Lett 14:6382–6386

Ugeda MM, Brihuega I, Guinea F, Gómez-Rodríguez JM (2010) Missing atom as a source of carbon magnetism. Phys Rev Lett 104:096804

van der Zande AM et al (2013) Grains and grain boundaries in highly crystalline monolayer molybdenum disulphide. Nat Mater 12:554–561

Wehling TO, Yuan S, Lichtenstein AI, Geim AK, Katsnelson MI (2010) Resonant scattering by realistic impurities in graphene. Phys Rev Lett 105:056802

Yang B, Xu H, Lu J, Loh KP (2014) Periodic grain boundaries formed by thermal reconstruction of polycrystalline graphene film. J Am Chem Soc 136:12041–12046

Yazyev OV (2008) Magnetism in disordered graphene and irradiated graphite. Phys Rev Lett 101:037203

Yazyev OV, Chen YP (2014) Polycrystalline graphene and other two-dimensional materials. Nat Nanotechnol 9:755–767

Yazyev OV, Helm L (2007) Defect-induced magnetism in graphene. Phys Rev B 75:125408

Yazyev OV, Katsnelson MI (2008) Magnetic correlations at graphene edges: basis for novel spintronics devices. Phys Rev Lett 100:047209

Yazyev OV, Louie SG (2010a) Topological defects in graphene: dislocations and grain boundaries. Phys Rev B 81:195420

Yazyev OV, Louie SG (2010b) Electronic transport in polycrystalline graphene. Nat Mater 9:806–809

Yazyev OV, Tavernelli I, Rothlisberger U, Helm L (2007) Early stages of radiation damage in graphite and carbon nanostructures: a first-principles molecular dynamics study. Phys Rev B 75:115418

Yazyev OV, Capaz RB, Louie SG (2011) Theory of magnetic edge states in chiral graphene nanoribbons. Phys Rev B 84:115406

Zhang X et al (2013a) Experimentally engineering the edge termination of graphene nanoribbons. ACS Nano 7:198–202

Zhang Y, Tan Y-W, Stormer HL, Kim P (2005) Experimental observation of the quantum Hall effect and Berry's phase in graphene. Nature 438:201–204

Zhang Z, Zou X, Crespi VH, Yakobson BI (2013b) Intrinsic magnetism of grain boundaries in two-dimensional metal dichalcogenides. ACS Nano 7:10475–10481

Understanding Novel Superconductors with Ab Initio Calculations

6

Lilia Boeri

Contents

Abstract

This chapter gives an overview of the progress in the field of computational superconductivity. Following the MgB_2 discovery (2001), there has been an impressive acceleration in the development of methods based on density functional theory to compute the critical temperature and other physical properties of actual superconductors from first principles. State-of-the-art ab initio methods have reached predictive accuracy for conventional (phonon-mediated) superconductors, and substantial progress is being made also for unconventional superconductors. The aim of this chapter is to give an overview of the existing

L. Boeri (✉)
Dipartimento di Fisica, Sapienza Università di Roma, Roma, Italy
e-mail: lilia.boeri@uniroma1.it

computational methods for superconductivity and present selected examples of material discoveries that exemplify the main advancements.

1   Introduction

The aim of this chapter is to offer an up-to-date perspective on the field of ab initio superconductivity and of the related development of numerical methods to compute critical temperatures and other physical properties of superconductors. The material-specific aspect is what distinguishes ab initio (= from first principles) approaches, based on density functional theory (DFT) and its extensions, from other theoretical approaches to superconductivity, which mainly focus on the general description of the phenomenon. The typical questions addressed by computational superconductivity are as follows: (i) What makes a certain compound a good (or bad) superconductor? (ii) How are its properties modified by external parameters, such as doping, pressure, and strain? (iii) Is it possible to find new materials with improved superconducting properties compared to existing ones?

The most relevant parameter that defines the performance of a superconductor for large-scale applications is its critical temperature (T_c): this means that addressing the questions above requires the development of methods accurate enough to predict the T_c of a superconductor and its dependence on external parameters. The progress in this direction, in the last 20 years, has been impressive.

For a large class of superconductors, i.e., conventional, phonon-mediated ones, ab initio methods are now so accurate that the focus of the field is gradually shifting from the description of existing superconductors to the design of new materials. The first successful example was the prediction of high-T_c conventional superconductivity in SH_3 (2014) (Drozdov et al. 2015a; Duan et al. 2014).

For unconventional superconductors, which comprise two of the most studied classes of materials, the high-T_c cuprates (Bednorz and Mueller 1986) and Fe-based superconductors, (Kamihara et al. 2008) ab initio approaches are still far from being predictive, but it is becoming more and more widely accepted that the single-particle electronic structure determines crucial properties of these materials, such as the symmetry of the superconducting gap and the behavior of magnetic excitations.

The topics and the structure of this chapter have been specifically thought out to illustrate the parallel progress in ab initio methods and material research for superconductors. I have chosen three discoveries that I consider the fundamental milestones of this process: (a) the report of superconductivity in MgB_2 in 2001, which has disproved the Cohen-Anderson limit for conventional superconductors; (b) the discovery of Fe-based superconductors, which has led to a much deeper understanding of the interplay between electronic structure, magnetism, and superconductivity in unconventional superconductors; and (c) the discovery of high-temperature superconductivity at megabar pressures in SH_3, which has given a spectacular demonstration of the predictive power of ab initio calculations.

Although I will give a general introduction to the theory of superconductivity and briefly describe the most recent advancements in ab initio methods, methodological

developments are not the main topic of this chapter: I refer the interested reader to excellent reviews in literature for a detailed discussion (Giustino 2017; Sanna 2017; Sanna et al. 2018). For space reasons, I am also forced to leave out some currently very active directions of superconductivity research, such as topological superconductivity (Sato and Ando 2017) superconductivity in 2D transition metal dichalcogenides (Klemm 2015), and artificial superlattices (Boschker and Mannhart 2017), and other more traditional topics, such as cuprates and other oxides (Chu et al. 2015), fullerenes (Gunnarsson 1997), layered halonitrides (Kasahara et al. 2015), etc.

On the other hand, I have included at the end a short perspective describing possible routes to high-T_c superconductivity which exploit novel developments in experimental and ab initio techniques, since I believe that in the next years, the combination of the two may lead to the discovery of many new superconductors.

The structure of the chapter is the following: I will start by giving a concise historical review of the most important discoveries in Sect. 2. In Sect. 3, I will then introduce the basic concepts of superconductivity theory and describe the most recent developments in ab initio methods. The main body of the paper is contained in Sect. 4, where, using selected material examples, I will try to give an impression of the rapid progress of the field in the understanding of both conventional and unconventional superconductors. Finally, in Sect. 5, I will propose possible practical routes to high-T_c superconductivity.

2 A Brief History of Research in Superconductivity

Superconductivity was discovered more than 100 years ago when H.K. Onnes observed that, when cooled below 4 K, mercury exhibits a vanishing resistivity (Onnes 1913). Perfect diamagnetism, which is the second fingerprint of a superconductor, was discovered by Meissner and Ochsenfeld around 20 years later (Meissner and Ochsenfeld 1933). While it was immediately clear that superconductors could have an enormous potential for applications, the low critical temperatures represented an insurmountable obstacle to large-scale applications.

In addition to presenting practical problems, superconductivity proved to be a major challenge also for theorists: fully microscopic theories of superconductivity – the Bardeen-Cooper-Schrieffer (BCS) and Migdal-Eliashberg (ME) theories – were developed only after almost 50 years after the original discovery (Bardeen et al. 1957; Migdal 1958; Eliashberg 1960; Allen and Mitrovic 1982; Scalapino 1969; Carbotte 1990). They describe superconductivity as due to the condensation of pairs of electrons of opposite spin and momentum (Cooper pairs), held together by an attractive *glue*. In conventional superconductors, the glue is provided by phonons (lattice vibrations), but other excitations such as plasmons, spin fluctuations, etc. can also mediate the superconducting pairing. We will treat here only the case of the so-called *boson-exchange* superconductors, and not other mechanisms, such as resonant valence bond, hole superconductivity, etc.

The understanding of the microscopic mechanism of superconductivity did not lead to any immediate, appreciable progress in the search for new superconductors; this translated into a general skepticism toward theory, which is well exemplified by one of Matthias' rules for superconductivity (*stay away from theorists!*). Indeed, rather than a predictive theory, ME theory was long considered a sophisticated phenomenological framework to describe existing superconductors, while the search for new materials was (unsuccessfully) guided by semiempirical rules. Even worse, two leading theorists used ME theory to demonstrate the existence of an intrinsic limit of around 25 K to the T_c of *conventional* superconductors. Although conceptually wrong, the *Cohen-Anderson limit* is still cited today as an argument against high-T_c superconductivity (Cohen and Anderson 1972).

The notion of an upper limit to T_c was first challenged by the discovery of the first unconventional superconductors, the cuprates, in 1986 (Bednorz and Mueller 1986). In contrast to conventional superconductors, which above T_c behave as ordinary metals, cuprates exhibit a complex phase diagram, with many coexisting phases and physical phenomena (charge and spin density waves, metal-insulator transitions, transport anomalies, etc.). In 1987 a cuprate, YBCO, broke the liquid N_2 barrier, with a T_c of 92 K (Wu et al. 1987), causing a general excitement in the media about a possible *superconducting revolution*; the highest T_c ever attained in this class is 156 K (Schilling et al. 1993). Despite almost 30 years of research and many different proposals, a quantitative theory of superconductivity in the cuprates is still lacking; furthermore, their large-scale applicability is also limited due to their high brittleness and manufacturing costs (Gurevich 2011).

The search for new materials took a different turn at the beginning of this century, when a T_c of 39 *K* was reported in a simple $s-p$ binary compound, magnesium diboride (MgB_2) (Nagamatsu et al. 2001). In contrast to the cuprates, MgB_2 is a conventional superconductor. In less than 2 years, ab initio calculations provided a key quantitative understanding of very specific aspects of superconductivity in this material, such as two-gap superconductivity, anharmonicity, role of magnetic and nonmagnetic impurities, doping, etc. (An and Pickett 2001; Kong et al. 2001; Kortus et al. 2001; Liu et al. 2001; Choi et al. 2002; Mazin et al. 2002). This stimulated a renewed enthusiasm in the search for new superconductors and, in parallel, the development of accurate ab initio methods to model them.

Indeed, in the last 17 years, superconductivity has been discovered in B-doped semiconductors (Ekimov et al. 2004; Bustarret et al. 2006), intercalated graphites (Weller et al. 2005), unconventional Fe-based superconductors (Kamihara et al. 2008), and, finally, high-pressure hydrides (Drozdov et al. 2015a,b). The T_cs as a function of the year of their discovery are shown as colored symbols in Fig. 1. The most important methodological developments are reported on the top axis: superconducting density functional theory (Lüders et al. 2005; Marques et al. 2005), electron-phonon interaction with Wannier functions (Giustino et al. 2007a; Calandra et al. 2010; Giustino 2017), ab initio anisotropic ME theory (Sanna et al. 2018; Margine and Giustino 2013), (stochastic) self-consistent harmonic approximation (Errea et al. 2014), and ab initio spin fluctuations (Essenberger et al. 2014, 2016).
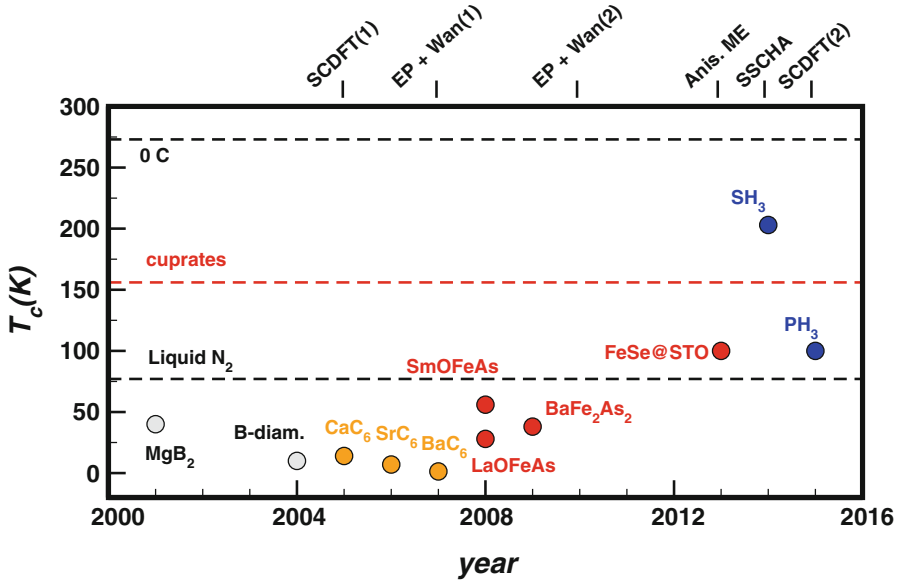
Fig. 1 Main developments in the field of superconductivity in the twenty-first century: the T_cs of the most important experimental discoveries as a function of year are shown as colored symbols. The top axis reports the most important methodological developments in ab initio superconductivity: superconducting density functional theory (*SCDFT(1)*) (Lüders et al. 2005; Marques et al. 2005), electron-phonon interaction with Wannier functions (*EP+WAN(1) and EP+WAN(2)*) (Giustino 2017; Giustino et al. 2007a; Calandra et al. 2010), ab initio anisotropic ME theory (*Anis. ME*) (Margine and Giustino 2013; Sanna et al. 2018), (stochastic) self-consistent harmonic approximation (*SSCHA*) (Errea et al. 2014), ab initio spin fluctuations (*SCDFT(2)*) (Essenberger et al. 2014, 2016)

Thanks to these advancements, ab initio calculations for conventional superconductors have now reached an accuracy which gives them fully predictive power. Methods are being developed also to treat other types of interactions, such as plasmons and spin fluctuations (Essenberger et al. 2014, 2016; Akashi and Arita 2013), and parameter ranges where the standard approximations of strong-coupling ME theory break down (Grimaldi et al. 1995). Combined with the development of efficient methods for ab initio crystal structure prediction and the progress in synthesis and characterization techniques (Woodley and Catlow 2008; Zhang et al. 2017), this opens unprecedented possibilities for material discovery.

3   Methods

In this section, I describe the methodological background of computational superconductivity. The first part introduces the main concepts behind the microscopic theories of superconductivity, i.e., the early Bardeen-Cooper-Schrieffer (BCS)

theory and the strong-coupling Migdal-Eliashberg (ME) theory. Although extremely accurate and elegant, ME theory was for a long time employed only as a semi-phenomenological theory, relying on electronic and vibrational spectra extracted from experiments. Early attempts of obtaining these quantities from density functional theory date back to the early 1980s, but with limited success, due to inadequate computational resources and insufficient accuracy of methods to treat phonons and electron-phonon (*ep*) interaction (Gaspari and Gyorffy 1972; Chang et al. 1985). The required accuracy was only achieved with density functional perturbation theory (DFPT) (Savrasov and Savrasov 1996; Baroni et al. 2001) and recently substantially increased with Wannier function interpolation methods (Giustino et al. 2007a; Calandra et al. 2010; Marzari et al. 2012).

In a very influential paper, already in 1996, Savrasov and Savrasov demonstrated that linear response calculations combined with Migdal-Eliashberg theory could reproduce the critical temperatures and other properties of elemental metals (Savrasov and Savrasov 1996). However, since at that time the T_cs of known conventional superconductors were much smaller than those of the cuprates, this result was erroneously perceived as of limited importance.

The report of superconductivity in MgB_2 in 2001 gave a strong impulse to the development of ab initio methods for superconductors, which resulted in two parallel lines of research: ab initio Migdal-Eliashberg theory and superconducting density functional theory, described in Sect. 3.2. In their fully anisotropic versions, including screened Coulomb interactions, they have a comparable accuracy of 5–10% on the critical temperature, gap, etc. (Sanna 2017; Sanna et al. 2018; Margine and Giustino 2013). This gives them fully predictive power, and, combined with methods for crystal structure prediction, offers the unprecedented possibility of designing superconductors ab initio, overcoming the practical limitations of experiments.

3.1   A Short Compendium of Superconductivity Theory

The first fully microscopic theory of superconductivity was formulated by Bardeen, Cooper, and Schrieffer in 1957 and is known as BCS theory (Bardeen et al. 1957).

BCS theory describes the transition of superconductors from an ordinary metallic state (*normal state*) to a new state, characterized by vanishing resistivity and perfect diamagnetism. In this *superconducting state*, the electronic spectrum develops a gap Δ around the Fermi level, which is maximum at zero temperature and vanishes at the critical temperature T_c. The critical temperature exhibits an *isotope effect*, i.e., T_c increases (or decreases) upon partial replacement of an element with a lighter (heavier) isotope. Isotope effects are also measured for other characteristic properties of superconductors (gap, specific heat, etc.).

BCS theory reconciles all the above experimental observations in a consistent framework, based on three key concepts:

(i) A Fermi sea of electrons in the presence of an attractive interaction is unstable toward the formation of a pair of electrons with opposite spin and momentum

(**Cooper pair**), which effectively behaves as a boson. In a superconductor, below T_c, a small but macroscopic fraction of electrons, of order $\Delta/E_F \simeq 10^{-3}$, forms Cooper pairs – this is sometimes referred to as the *condensate fraction* of a superconductor.

(ii) A variational many-body wavefunction for the electrons is constructed from a superposition of ordinary single-particle states and Cooper pairs (**BCS wavefunction**). The existence of a condensate fraction leads to the appearance of a gap in the electronic spectrum $\varepsilon_\mathbf{k}$, which satisfies the self-consistent equation:

$$\Delta_\mathbf{k} = -\frac{1}{2} \sum_{\mathbf{k,k'}} \frac{V_{\mathbf{k,k'}} \Delta_{\mathbf{k'}}}{\sqrt{\varepsilon_\mathbf{k}^2 + \Delta_\mathbf{k}^2}} \cdot \tanh\left(\frac{\sqrt{\varepsilon_\mathbf{k}^2 + \Delta_\mathbf{k}^2}}{2T}\right). \tag{1}$$

iii) Using a simple model for the electron-electron interaction $V_{\mathbf{k,k'}}$, the so-called **BCS potential**, which is attractive only if the two electrons with wavevector $\mathbf{k}$, $\mathbf{k'}$ both lie in a small energy shell ω_D around the Fermi energy E_F, Eq. 1 can be solved analytically, and the gap and T_c are given by:

$$\Delta(T = 0) \simeq 2\omega_D \exp\left(-\frac{1}{N(E_F)V}\right), \qquad k_b T_c = 1.13\omega_D \exp\left(-\frac{1}{N(E_F)V}\right), \tag{2}$$

The original idea of Bardeen, Cooper, and Schrieffer is that the attractive interaction V between electrons is mediated by lattice vibrations (phonons). In this case, ω_D is a representative phonon energy scale, such as the Debye frequency; $N(E_F)$ is the electronic density of states at the Fermi level.

One of the first successes of BCS theory has been the explanation of the isotope effect on T_c, $\alpha_{T_c} = -\frac{d\ln(T_c)}{d\ln(M)} = 0.5$, where M is the ionic mass, as well as the prediction of several *magic ratios*, satisfied in most elemental superconductors: the most famous is probably the ratio $\frac{2\Delta(0)}{k_B T_c} = 3.53$.

However, BCS theory is valid only at weak-coupling ($\lambda = N(E_F)V < 0.2 - 0.3$) and instantaneous interactions; these assumptions are not verified in many conventional superconductors, where the actual values of T_c, isotope effect, magic ratios, etc. are spectacularly different from the BCS predictions (Marsiglio and Carbotte 1986).

A quantitative description of the strong-coupling, retarded regime is given by the many-body ME theory, based on a set of self-consistent coupled diagrammatic equations for the electronic and bosonic propagators (Allen and Mitrovic 1982; Scalapino 1969; Carbotte 1990; Margine and Giustino 2013). The bosons that mediate the superconducting pairing can be phonons or other excitations of the crystal, such as plasmons or spin fluctuations (Berk and Schrieffer). Below the critical temperature, electrons are described by a normal and an *anomalous* electronic propagator, the latter accounting for Cooper pairs. ME equations are then obtained from Dyson's equations for the normal and anomalous propagators; in their most commonly used, T-dependent form, they can be written as:

$$Z(\mathbf{k}, i\omega_n) = 1 + \frac{\pi T}{\omega_n} \sum_{\mathbf{k}'n'} \frac{\delta(\varepsilon_{\mathbf{k}'})}{N(E_F)} \frac{\omega_{n'}}{\sqrt{\omega_{n'}^2 + \Delta^2(\mathbf{k}, i\omega_{n'})}} \lambda(\mathbf{k}, \mathbf{k}', n - n') \quad (3)$$

$$Z(\mathbf{k}, i\omega_n)\Delta(\mathbf{k}, i\omega_n) = \pi T \sum_{\mathbf{k}'n'} \frac{\delta(\varepsilon_{\mathbf{k}'})}{N(E_F)} \frac{\Delta(\mathbf{k}', i\omega_{n'})}{\sqrt{\omega_{n'}^2 + \Delta^2(\mathbf{k}, i\omega_{n'})^2}} \times$$

$$\times \left[\lambda(\mathbf{k}, \mathbf{k}', n - n') - \mu(\mathbf{k} - \mathbf{k}')) \right] \quad (4)$$

where $Z(\mathbf{k}, i\omega_n)$ and $Z(\mathbf{k}, i\omega_n)\Delta(\mathbf{k}, i\omega_n)$ are the self-energy of the normal and anomalous electronic propagators, respectively; $i\omega_n = i(2n\pi T + 1)$ are Matsubara frequencies, and $\mathbf{k}, \mathbf{k}'$ are the electronic momenta; and the δ function restricts the sum over $\mathbf{k}'$ only to electronic states at the Fermi level. The electrons interact through a retarded attractive interaction $\lambda(\mathbf{k}, \mathbf{k}', n - n')$ and an instantaneous Coulomb repulsion $\mu(\mathbf{k} - \mathbf{k}')$. The interaction $\lambda(\mathbf{k}, \mathbf{k}', n - n')$ is usually expressed in terms of an electron-boson spectral function $\alpha^2 F(\mathbf{k}, \mathbf{k}', \omega)$ as :

$$\lambda(\mathbf{k}, \mathbf{k}', n - n') = \int_0^\infty d\omega \frac{2\omega}{(\omega_n - \omega_{n'})^2 + \omega^2} \alpha^2 F(\mathbf{k}, \mathbf{k}', \omega) \quad (5)$$

Equations (3) and (4) can be solved numerically to obtain the gap and other thermodynamic quantities. It is very common, and in most cases sufficiently accurate, to approximate the more general form with an isotropic version, replacing the sums on the electronic momenta $(\mathbf{k}, \mathbf{k}')$ with Fermi surface averages. If one is only interested in the T_c, there are excellent approximation formulas; a very popular choice for phonon-mediated superconductors is the McMillan-Allen-Dynes expression (Allen and Dynes 1975):

$$T_c = \frac{\omega_{\log}}{1.2 k_B} \exp\left[-\frac{1.04(1 + \lambda)}{\lambda - \mu^*(1 + 0.62\lambda)} \right], \quad (6)$$

where $\lambda = 2 \int d\omega \frac{\alpha^2 F(\omega)}{\omega}$ and $\omega_{\log} = \exp\left[\frac{2}{\lambda} \int \frac{d\omega}{\omega} \alpha^2 F(\omega) \ln(\omega) \right]$ are the *ep* coupling constant and logarithmic averaged phonon frequency, respectively; μ^* is the so-called Morel-Anderson pseudopotential, obtained by screening the full Coulomb potential up to a characteristic cutoff energy (Morel and Anderson 1962).

3.2   Ab Initio Methods

The two methods described in this section, ab initio anisotropic Migdal-Eliashberg theory (DFT-ME in the following) (Margine and Giustino 2013; Sanna et al. 2018) and superconducting density functional theory (SCDFT) (Oliveira et al. 1988; Lüders et al. 2005; Marques et al. 2005), represent a fundamental step forward in the study of actual superconductors, because they permit to obtain a full characterization of the normal and superconducting state of a system from the sole knowledge of the

chemical composition and crystal structure. Although there are fundamental and practical differences between the two, both methods rely crucially on the ability of DFT of providing accurate electronic and bosonic spectra for most materials at an affordable computational cost (Jones and Gunnarsson 1989; Baroni et al. 2001).

The basic assumptions are the following:

1. *Electronic quasiparticles* appearing in Eqs. (3) and (4) are replaced by the Kohn-Sham quasiparticles.
2. *Bosonic* excitation energies and electron-boson spectral functions are obtained from density functional perturbation theory (DFPT) (Baroni et al. 2001).

   For *phonons*, the spectral function is:

   $$\alpha^2 F_{ph}(\mathbf{k}, \mathbf{k}', \omega) = N(E_F) \sum_{\nu} g_{\mathbf{k}\mathbf{k}',\nu} \delta(\omega - \omega_{\mathbf{k}-\mathbf{k}',\nu}), \qquad (7)$$

   where $g_{\mathbf{k},\mathbf{k}'} = \left\langle \mathbf{k}' | \delta V_{scf}^{\nu,\mathbf{k}'-\mathbf{k}} | \mathbf{k} \right\rangle$ is the *ep* matrix element for the mode ν and $\delta V_{scf}^{\nu,\mathbf{q}}$ is the variation of the Kohn-Sham self-consistent potential due to an infinitesimal displacement along the eigenvector of the phonon mode ν with wavevector $\mathbf{q} = \mathbf{k}' - \mathbf{k}$.

   For *spin fluctuations*, the spectral function is proportional to the imaginary part of the longitudinal interacting spin susceptibility $\chi_{zz}(\mathbf{q} = \mathbf{k}-\mathbf{k}', \omega)$(Vignale and Singwi 1985), which, using linear response within the time-dependent density functional theory (TDDFT) framework (Runge and Gross 1984), can be written as:

   $$\chi_{zz}(\mathbf{q}, \omega) = \frac{\chi^{KS}(\mathbf{q}, \omega)}{1 - f_{xc}(\mathbf{q}, \omega)\chi^{ks}(\mathbf{q}, \omega)}, \qquad (8)$$

   where $\chi^{KS}(\mathbf{q}, \omega)$ is the Kohn-Sham susceptibility and $f_{xc}(\mathbf{q}, \omega)$ is the exchange and correlation kernel (Essenberger et al. 2012).
3. The *Coulomb potential* $\mu(\mathbf{k} - \mathbf{k}')$ which, in most empirical approaches, is treated as an adjustable parameter within the Morel-Anderson approximation (Morel and Anderson 1962), is treated fully ab initio by screening the bare Coulomb potential within RPA. Substantial deviations from typical values of $\mu^* = 0.1 - -0.15$ are found in strongly anisotropic systems such as MgB_2 and layered superconductors; in some cases, such as alkali metals at high pressure, the effect of Coulomb interactions is even stronger, giving rise to plasmonic effects (Akashi and Arita 2013).

Once the spectra of the quasiparticles and the interactions between them are known from first principles, DFT-ME or SCDFT can be applied to describe the superconducting state. DFT-ME theory amounts to solving the fully anisotropic ME equations, for electronic and bosonic spectra computed in DFT; the current implementations solve the equations in Matsubara frequencies and use Padé approximants

to continue them to real space. The obvious advantage of this method is that all quantities have an immediate physical interpretation through many-body theory.

SCDFT is a fundamentally different approach, which generalizes the original Hohenberg-Kohn idea of one-to-one correspondence between ground-state density and potential (Hohenberg and Kohn 1964), introducing two additional densities (and potentials) for the ionic system $\Gamma(\mathbf{R}_i)(V_{ext}(\mathbf{R}_i))$ and the superconducting electrons $\chi(\mathbf{r}, \mathbf{r}')(\Delta(\mathbf{r}, \mathbf{r}'))$ and finding the values that minimize a suitable energy functional. This permits to derive a gap equation which is analogous to the BCS one but instead of an empirical potential contains a *kernel* with all the relevant physical information on the system. I refer the reader to the original references for a full derivation (Lüders et al. 2005; Marques et al. 2005; Oliveira et al. 1988) and to Sanna (2017) for an excellent pedagogical introduction.

SCDFT equations are more easily solvable on a computer than fully anisotropic ME equations because they do not require expensive sums over Matsubara frequencies; however, the interpretation of many physical quantities, including the frequency dependence of the gap, is not equally transparent and straightforward. Another intrinsic limitation is that, like in all DFT-like methods, the quality of the results depends strongly on the quality of the functional.

The latest-developed functionals yield results with an *accuracy* comparable to that of the best DFT-ME calculations, which for most conventional superconductors is between 5 and 10% of the critical temperature. The most severe source of inaccuracy in DFT calculations for superconductors is usually an underconverged integration in reciprocal space in Eqs. (3) and (4), an issue that has considerably improved thanks to the use of Wannier interpolation techniques.

Achieving quantitative accuracy for conventional pairing also encouraged to address ab initio effects, which are often disregarded even in model approaches, such as anharmonicity, vertex corrections, and zero-point effects (Pietronero et al. 1995; Boeri et al. 2005; Lazzeri et al. 2003).

These turned out to be relevant for a wide variety of compounds, particularly for the newly discovered superconducting hydrides, where the energy scales of phonons and electrons are comparable (Errea et al. 2014, 2016; Sano et al. 2016). I refer the reader to the relevant references for an in-depth discussion.

For unconventional superconductors, on the other hand, the most severe source of inaccuracy is intrinsic and is the possible divergence in the spin-fluctuation propagator, which completely destroys the predictive power of DFT approaches for currently available functionals. Moreover, most unconventional superconductors suffer from the lack of accuracy of DFT in strongly correlated systems (Yin et al. 2013).

3.3   Developments in Related Fields: Ab Initio Material Design

The term *ab initio material design* indicates the combination of methods for ab initio crystal structure prediction and thermodynamics to predict the behavior of materials at arbitrary conditions of pressure and temperature, knowing only the initial chemical composition of the system. The development of these methods

represents a substantial step forward in computational condensed matter research, as it overcomes one of its biggest limitations, exemplified by the *Maddox Paradox* (1988): *"One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition."*

The basic working principle of **ab initio crystal structure prediction** is quite simple. Predicting the crystal structure of a material for a given regime of chemical composition and pressure amounts to finding the global minimum of a complicated landscape, generated by the ab initio total energies (or enthalpies) of all possible structures. The number of possible configurations for a typical problem is so large that a purely enumerative approach is unfeasible; in the last years, several methods have been devised to make the problem computationally manageable, such as ab initio random structure search, minima hopping, metadynamics, genetic algorithms, particle swarm algorithms, etc. (Woodley and Catlow 2008).

Once the most favorable crystal structure for a given composition and pressure is known, the ab initio CALPHAD (CALculation of PHAse Diagrams) approach permits to predict accurate **phase diagrams** (Lukas et al. 2007), as illustrated in Fig. 2. The binary Li-S phase diagram in panel (*a*) shows the stability ranges of different Li-S compositions and has been constructed repeating several *convex hull* calculations at different pressures.

The *convex hull* construction is shown in panel (b). The points represent the lowest-energy structure predicted by an evolutionary search for a given composition; for a binary phase with composition $A_x B_y$, the formation enthalpy ΔH is defined as: $\Delta H = H(A_x B_y) - [x H(A) + y H(B)]$. If this quantity is negative, the phase $A_x B_y$ is stable with respect to elemental decomposition; if it is positive, the phase is highly (or weakly) metastable, i.e., if formed, it will decompose into its elemental constituents in finite time. However, the decomposition into the two elements is often not the relevant one, as a compound could decompose into other phases, preserving the correct stoichiometry. To estimate all possible decompositions for the binary system, one constructs the most convex curve connecting the formation enthalpy of all known phases for a given stoichiometry. Points on this convex hull represent stable compositions into which phases which lie above the convex hull will decompose into given a sufficient interval of time. The figure also shows that, as pressure increases, the diversity of the phase diagram increases, i.e., off-stoichiometry compositions become possible (*forbidden chemistry*). The convex hull construction can be easily extended to multinary systems (Gibbs diagrams) and finite temperatures including entropic effects.

4   Materials

The aim of this section is to illustrate how, within a little bit more than a decade, an increased understanding of material-specific aspects of superconductivity gained from ab initio calculations has permitted to replace empirical rules to search for new superconductors with quantitative strategies.
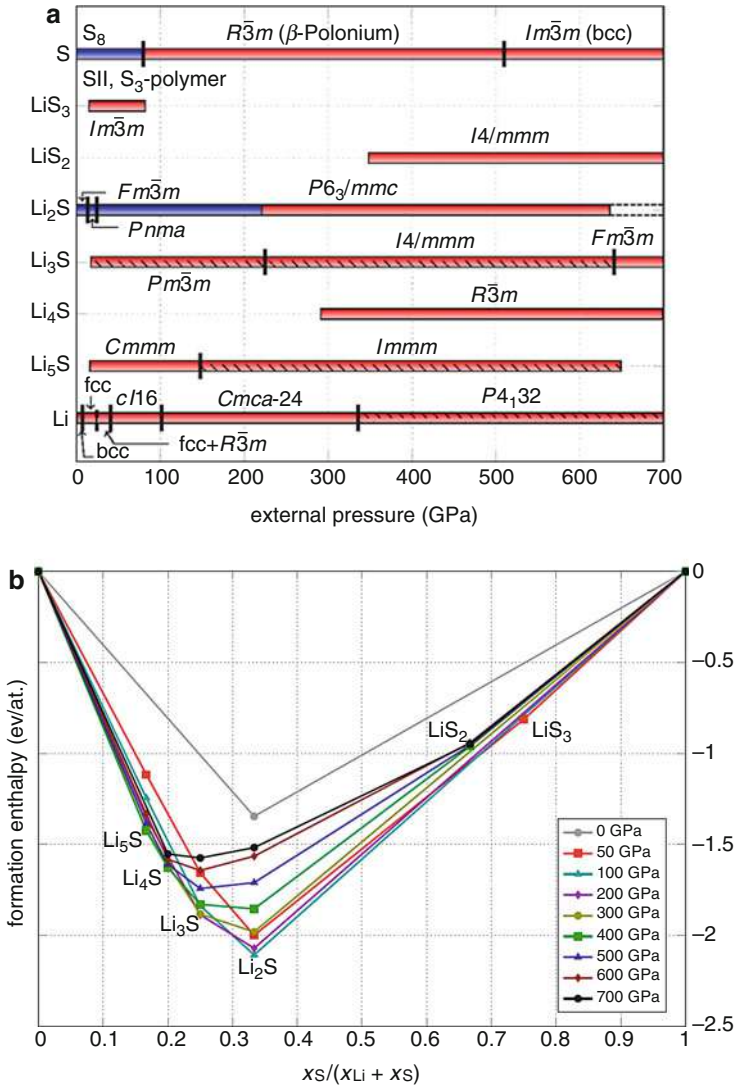
Fig. 2 Basic steps of ab initio materials prediction: the complete phase diagram of a binary alloy as a function of pressure (**a**) can be constructed combining several convex hull constructions at different pressures (**b**) (see text). (Readapted from Kokail et al., Phys. Rev. B 94, 060502(R) (2016) – Copyright (2016) by the American Physical Society)

I will start with a general discussion that determine the T_c of conventional superconductors (Sect. 4.1), introducing the concepts of *dormant ep* interactions and lattice instabilities and showing how these can be used to interpret both the old empirical knowledge (Matthias' rules and Cohen-Anderson limit) and the latest experimental discoveries (Savrasov and Savrasov 1996; Boeri et al. 2002, 2007, 2008; Profeta et al. 2012).

I will then use a toy numerical model (simple graphite), to see how these concepts are realized in an actual physical system, and a very simple approximation to doping (rigid-band) to simulate the effect of physical doping and detect the sources of *ep* interaction in graphite-like materials. Both models are useful for a first exploration but inadequate to make accurate predictions for actual superconductors, where the doping is usually obtained via chemical substitution, which causes a major rearrangement of phononic and electronic states and hence sizable changes in the values of the *ep* interaction and T_c.

For simple graphite, nature provides two simple realizations of two of its sources of dormant *ep* interactions: MgB_2, a prototype of *covalent* superconductors, and graphite intercalated compounds, where superconductivity is correlated with the filling of *interlayer* states. In Sects. 4.3 and 4.4, I will discuss these two examples in detail and also indicate the main theoretical predictions and experimental discoveries which were inspired by them.

In particular, the line of research on covalent superconductors culminated in the discovery of high-T_c conventional superconductivity at extreme pressures in SH_3, discussed in Sect. 4.5, which also represents a fundamental step forward in the direction of the search of new superconductors using first-principles methods.

After describing the incredible evolution of the state of research in conventional superconductors, in Sect. 4.6, I will present a representative example of unconventional superconductors, iron pnictides, and chalcogenides (Fe-based superconductors), discovered in 2008, which shares many similarities with the high-T_c cuprates. This example will allow me to give an idea of the many challenges that theory faces in the description of unconventional superconductors, already in the normal state, and currently represent a fundamental obstacle to the derivation of numerical methods to compute T_cs.

4.1   Conventional Superconductors: Search Strategies

Matthias' rules The so-called Matthias' rules are a set of empirical rules that summarize the understanding of superconductors in the 1970s. The rules were allegedly formulated, and revised several times, by Bernd Matthias, one of the leading material scientists in superconductivity: they are usually cited in this form:

1. High symmetry is good; cubic symmetry is the best.
2. High density of electronic states is good.
3. Stay away from oxygen.
4. Stay away from magnetism.
5. Stay away from insulators.
6. Stay away from theorists.

Matthias' rules were inspired by A15 superconductors (cubic transition metal alloys, which can be easily doped, exhibit sharp peaks in the electronic DOS, and are prone to lattice and magnetic instabilities) and clearly disproved by subsequent experimental discoveries: cuprates and pnictides (first, third, fourth, and fifth rule) (Bednorz and Mueller 1986; Kamihara et al. 2008) but also conventional superconductors such as MgB_2 (Nagamatsu et al. 2001) (first and second rule) (Drozdov et al. 2015a). However, their impact on superconductivity research has been so important that they are still sometimes cited as arguments against conventional superconductivity or the possibility of theoretically predicting new superconductors, together with another old quasi-empirical rule, the Cohen-Anderson limit.

In order to derive more general, nonempirical strategies to search for new superconductors, I will begin with a simple analytical model. Instead of taking into account complexity of the electronic and vibrational properties of real superconductors, for the moment I consider an ideal case, in which a single phonon branch with frequency ω and a single electronic band with density of states at the Fermi level $N(E_F)$ is coupled through an average matrix element I. In this case, the T_c is well described by the McMillan-Allen-Dynes formula (Eq. 6) with all constant factors set to one, $\omega_{\ln} = \omega$, and the coupling constant is given by the Hopfield expression $\lambda = (N(E_F)I^2)/(M\omega^2)$:

$$T_c = \frac{\omega}{k_B} \exp\left[-\frac{(1+\lambda)}{\lambda - \mu^*(1+\lambda)}\right], \qquad (9)$$

These formulas indicate that there are three main strategies to optimize the critical temperature of a conventional superconductor: (i) maximize the value of the electronic DOS at the Fermi level $N(E_F)$ (first two Matthias' rules); (ii) select compounds which contain light elements (and stiff bonds), to maximize the characteristic lattice energy scales (ω); and (iii) increase the *ep* matrix elements (I).

While the first two strategies were already understood in the early 1970s, it became apparent only with the MgB_2 discovery that it is possible to find compounds where the intrinsic *ep* matrix elements I are much larger than in transition metals and their alloys, where the maximum T_c does not exceed 25 K. In a seminal paper, An and Pickett pointed out that the (relatively) high T_c of MgB_2 occurs because of *covalent bonds driven metallic* (An and Pickett 2001).

The reason why covalent metals have larger intrinsic *ep* coupling than ordinary metals can be intuitively understood looking at Fig. 3, which shows isosurfaces of the electronic localization function (ELF) for three different superconductors: Nb ($T_c = 9$ K), MgB_2 ($T_c = 39$ K), and SH_3 ($T_c = 203$ K). The ELF indicates regions where electrons are concentrated. It is clear that in MgB_2 and SH_3, the electrons localize along the bonds, while in Nb they are delocalized over the whole volume. When atoms undergo phonon vibrations, electrons localized along a bond will feel a much stronger perturbation than those spread out over the whole crystal. However, the arguments above are oversimplified, as the existence of strong directional bonds is a necessary prerequisite for large *ep* coupling, but it is not sufficient: due to the small energy scales involved in the superconducting pairing, it is also essential that
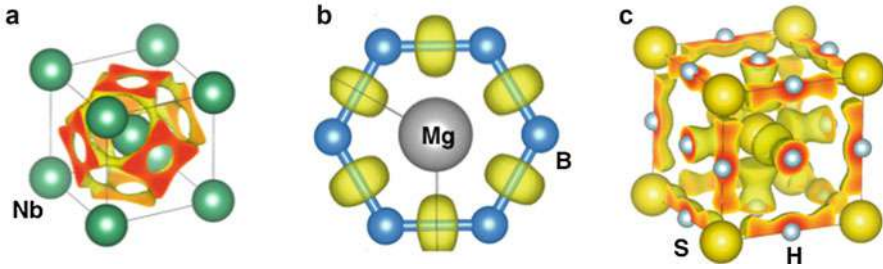
Fig. 3 Electronic localization function in conventional superconductors: bcc Nb ($T_c = 9$ K); MgB_2 ($T_c = 39$ K); SH_3 ($T_c = 203$ K)

the electronic states which contribute to this bond lie at the Fermi level; otherwise, they remain *dormant* and do not contribute to the superconducting pairing. More precisely, this means that shifting the position of the Fermi level (E_F) selects different matrix elements g in Eq. (7) when performing averages over the Fermi surfaces in in Eqs. (3) and (4), and the *ep* interaction λ, and hence T_c, is appreciable only for some positions of E_F.

A possible strategy to search for new conventional superconductors thus amounts to identifying, first, possible **dormant** *ep* **interactions** within a given material class and, second, physical mechanisms to activate them, such as doping, pressure, and alloying. Ab initio approaches permit to explore both steps of this process, at different levels of approximation. In the following, I will illustrate the basic working principle, using an example (simple graphite) and two practical realizations (MgB_2 and intercalated graphites); I will also refer to the same principles when discussing possible perspective of future research in Sect. 5.

Before moving on with the discussion, I need to introduce a second concept which is crucial in the search for new conventional superconductors: **lattice instabilities**. This argument is important because it is at the heart of the Cohen-Anderson limit (Cohen and Anderson 1972).

According to Eq. (9), T_c can apparently be increased indefinitely, increasing λ, which contradicts what is observed in practice, since the critical temperatures of actual superconductors are limited. However, in my discussion, I have so far disregarded the *feedback* effect between phonon frequencies and *ep* interaction, which is one of the main limiting factors to high T_c in actual materials. Indeed, the same *ep* interaction that pairs electrons leading to superconductivity also causes a decrease (softening) of the phonon frequencies. This means that the frequency ω appearing in McMillan's formula for T_c (Eq. 9) should be more correctly rewritten as: $\omega^2 = (\Omega_0)^2 (1 - 2\lambda_0)$, where Ω_0 is the bare frequency of the lattice, in the absence of *ep* interaction, and λ_0 is the corresponding coupling constant. It is now easy to see that the T_c for a fixed Ω_0 has a maximum for a given value of λ_0 and then decreases approaching a lattice instability ($\omega \to 0$); this means that within a given material class, T_c can only be increased up to a threshold value determined by λ_0, before incurring in a lattice instability. Using typical parameters for the

A15 compounds, which were the best known superconductors when the Cohen-Anderson limit was formulated, gives a maximum value of T_c of ~ 25 K. However, covalent superconductors such as MgB_2, doped diamond, and SH_3 have much larger characteristic phonon scales (Ω_0, λ_0) and can sustain much larger T_c.

4.2   Dormant *ep* Interaction in Graphite

To put general arguments on more physical grounds, we now consider a toy model based on an actual physical system, simple graphite, which is realized stacking several layers of graphene on top of each other (the stacking is thus AAA, in contrast to the ABA and ABC stacking of Bernal and rhombohedral graphite); for this experiment, we keep the interlayer distance equal to that of Bernal graphite.

The unit cell contains two inequivalent carbon atoms, which form a hexagonal lattice. This means that the four orbitals of carbon will split into three sp^2 hybrids and a p_z orbital, forming six σ and two π bands, of bonding and antibonding character. The blue arrow points to a fifth band, which has no carbon character: this is the so-called interlayer (IL) band, which is essentially a free-electron state confined between the carbon layers. The relative Wannier function, shown in the inset of the figure, is indeed centered in the middle of the interstitial region between the graphitic layers and has spherical symmetry (Boeri et al. 2007; Csaniy et al. 2005).

The phonon spectrum of pure graphite (not shown) is even simpler than its electronic structure: in-plane bond-stretching optical phonons form a rather narrow band between 150 and 200 meV, while optical out-of-plane and acoustical modes are more dispersive and extend up to 100 meV.

For the doping of pure graphite, corresponding to four electrons/carbon, the Fermi level cuts the band structure near the crossing point between π and π^* bands; the *ep* coupling is extremely low and gives rise to a negligible T_c. However, different *dormant ep* interactions can be activated if, with a *gedankenexperiment*, the Fermi level is shifted to higher or lower energies. This rigid-band shift is the simplest approximation to physical doping within an ab initio calculation.

The right panel of Fig. 4 shows a rigid-band calculation of the T_c of simple graphite, in an energy range of ± 8 eV around the Fermi level, corresponding to a doping of ± 1.5 electrons/carbon; in this calculation, the phonons and *ep* matrix elements are computed only once, for the physical doping of four electrons/carbon, but the averages and sums on the Fermi surface in Eqs. (3) and (4) are recomputed for different positions of the Fermi level (E_F).

The figure clearly shows that T_c is still negligible for small variations of energy around the original Fermi level but has a marked increase as soon as holes or electrons are doped into the σ (bonding or antibonding) or IL bands, reaching a maximum of ~ 100 K, when the Fermi level reaches a large van Hove singularity at ~ 8 eV, corresponding to the bottom of the σ^* band. Apart from causing substantial deviations in T_cs, the coupling of phonons to σ, IL, and π electrons is also qualitatively different. σ and π electrons couple mostly to high-energy optical phonons, which modulate the interatomic distances in the hexagonal layers, and hence the
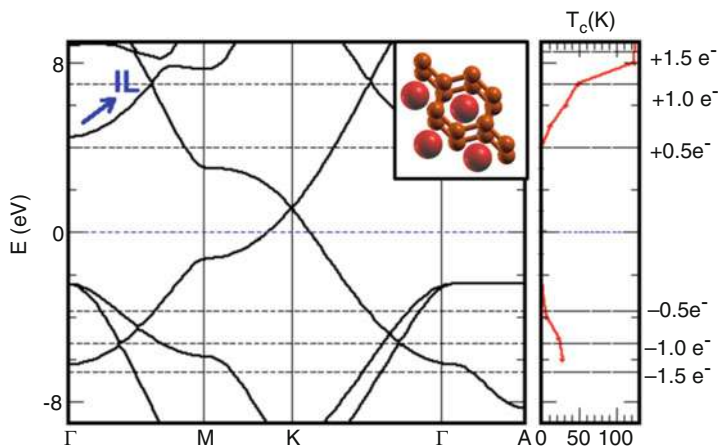
Fig. 4 Rigid-band study of superconductivity in simple graphite; dashed lines indicate the position of the Fermi level for a semi-integer doping of holes or electrons; the blue arrow marks the position of the interlayer band, whose Wannier function is shown in the inset

hopping between neighboring carbon sites; IL electrons, instead, respond to out-of-plane phonons, which modulate their overlap with the π wavefunctions that stick out-of-plane. This effect causes a substantial variation in the shape of the Eliashberg functions (not shown) and ω_{log}. The right scale on Fig. 4 indicates the position of the Fermi level for semi-integer values of hole($-$) or electron($+$) doping; in both cases, a finite T_c is obtained for a minimum doping of half an electron/carbon if needed, but obtaining a high T_c requires a much higher doping (~ 1 e^-/C atom). Note that the highest doping levels that can be obtained with field effect using liquid electrolytes are much lower, i.e., of the order of a few tenths of e^-/carbon, so that the only viable alternative to realize superconductivity in doped graphite is via chemical doping.

The rigid-band approximation is instructive to identify dormant *ep* interactions, but the calculated T_cs are often severely overestimated compared to other experiments or more sophisticated approximations for doping. In fact, the rigid-band approach neglects important effects due to the self-consistent rearrangement of electrons produced by doping, such as band shifts, renormalization of the phonon frequencies, and screening of the *ep* matrix elements (Subedi and Boeri 2011; Casula et al. 2012).

For simple graphite, nature provided two extremely ingenious realizations of our predictions, discussed in the next two sections.

4.3   Magnesium Diboride and Other Covalent Superconductors

With a T_c of 39K, magnesium diboride (MgB_2) currently holds the record for conventional superconductivity at ambient pressure (Nagamatsu et al. 2001; Bhaumik et al. 2017). Its crystal structure is layered: boron forms graphite-like hexagonal planes; magnesium is placed in between, at the center of the hexagons. Mg is
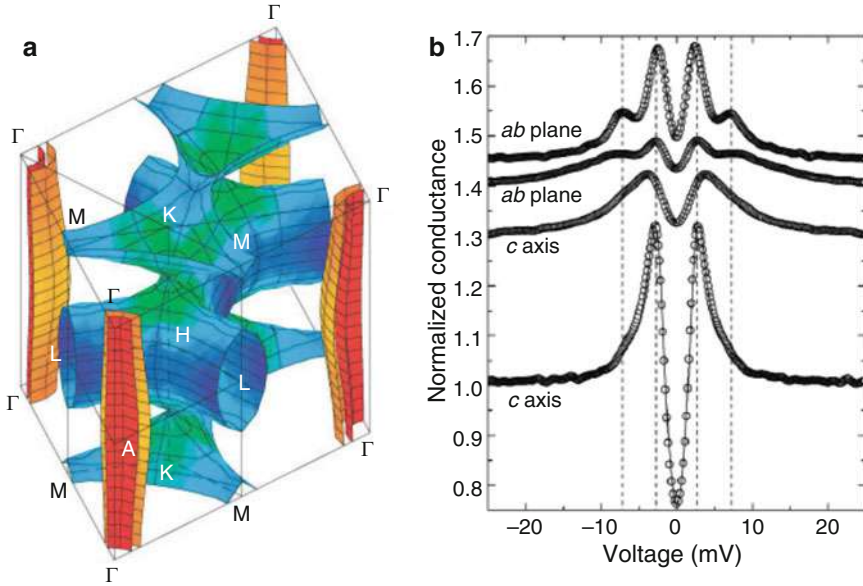
Fig. 5 Two-gap superconductivity in MgB$_2$: anisotropic distribution of the gap on the Fermi surface, predicted by DFT (**a**) (Figure reprinted with permission from Nature publishing group, H.J. Choi et al., The origin of the anomalous superconducting properties of MgB2, Nature, 418, 758 (2002); Copyright (2002); evidence of two-gap behavior from tunneling experiments (**b**) (Figure reprinted with permission from R.S. Gonnelli et al., Phys. Rev. Lett. 89, 247004 (2002); Copyright (2002) by the American Physical Society)

completely ionized (Mg^{2+}) and thus MgB$_2$ is not only isostructural but also isoelectronic to graphite. However, due to the attractive potential of the Mg ions, the center of mass of the π bands is shifted up with respect to that of the σ bands compared to simple graphite, so that MgB$_2$ behaves effectively as a compensated (semi)-metal. The σ holes and π electrons form two cylinders around the center and a 3D tubular network around the corners of the Brillouin zone, respectively, as shown in Fig. 5a.

The two groups of electronic states experience a rather different coupling to phonons: over 70% of the total *ep* coupling, in fact, comes from σ holes and bond-stretching phonons; the rest is distributed over the remaining phonon modes and electronic states (Kong et al. 2001). The simplest approximation to account for the strong anisotropy of the *ep* coupling over the Fermi surface is to replace the *ep* coupling constant λ in Eq. (3) and (4), with 2×2 matrices of the form (Mazin et al. 2002):

$$\lambda = \begin{pmatrix} \lambda_{\sigma\sigma} & \lambda_{\sigma\pi} \\ \lambda_{\pi\sigma} & \lambda_{\pi\pi} \end{pmatrix} = \begin{pmatrix} 1.02 & 0.30 \\ 0.15 & 0.45 \end{pmatrix} \tag{10}$$

When the interband coupling is finite but appreciably smaller than the intraband one, $|\lambda_{\sigma\pi} + \lambda_{\pi\sigma}| < |\lambda_{\sigma\sigma} + \lambda_{\pi\pi}|$, the theory predicts that experiments should

observe two distinct gaps, closing at the same T_c (Suhl et al. 1959). Two-gap superconductivity was indeed observed by different experimental techniques: specific heat, tunneling, and angle-resolved photoemission spectroscopy (ARPES) for the first time in MgB_2 (Mazin et al. 2002).

Indeed, in most superconductors, multiband and anisotropic effects are extremely difficult to detect because they are suppressed by the interband scattering induced by sample impurities; in MgB_2 the real-space orthogonality of the σ and π electronic wavefunctions prevents interband scattering, and two-gap superconductivity can be detected also by techniques with a limited resolution (Mazin et al. 2002).

Figure 5b shows in- and out-of-plane tunneling spectra of MgB_2, which permit to unambiguously identify a σ (large) and π (small) gap (Gonnelli et al. 2002). The experimental spectra compare extremely well with the theoretical prediction of two different gaps on the σ and π sheets of the Fermi surface; the image in the left panel of the figure shows an anisotropic DFT-ME calculation of the gap; in the figure, the color of the Fermi surface is proportional to the size of the gap on that sheet (Choi et al. 2002).

Besides providing the first clear case of two-gap superconductivity, MgB_2 is the first example of superconductivity from *doping covalent bonds*. The most spectacular realization of this possibility came in 2004, with the report of a T_c of 4 K in heavily boron-doped diamond, raised to 11 K in thin films (Ekimov et al. 2004; Takano et al. 2004).

Diamond is a wideband insulator ($\Delta \sim 5.5$ eV); when boron is doped at small concentrations, an acceptor band forms within the gap. At the much higher doping levels ($\sim 1-10\%$) realized in superconducting samples, the impurity band overlaps so strongly with the valence band of diamond, that the net effect of B doping is to create holes at the top of the valence band (Yokoya et al. 2005). This band is formed by the bonding combination of the four carbon sp^3 hybrids. Boron-doped diamond can thus be seen as a 3D analogue of MgB_2, where, similarly to MgB_2, a small fraction of σ holes created by B doping exhibits an extremely strong coupling to bond-stretching phonons.

Since the *ep* coupling is concentrated in a single type of phonon modes and electronic states, the simplified formulas introduced in Sect. 4.1 give a reasonable approximation to T_c for both MgB_2 and diamond (Boeri et al. 2004; Lee and Pickett 2004; Blase et al. 2004; Giustino et al. 2007b; Hoesch et al. 2007). Furthermore, they permit to understand why, even though the C-C bonds in diamond are actually stiffer than the B-B bonds in MgB_2, the measured T_cs are much lower. In fact, near the bottom (or top) of a band, i.e., at low dopings, the DOS increases as $\sqrt{E}$ in 3D, while it is virtually constant in 2D. This implies that, as holes or electrons are doped into the system, $N(E_F)$, and hence λ, increases much more slowly in 3D systems than in 2D ones. Thus, for physical ranges of dopings, the maximum T_c in doped semiconductors is typically very low (Boeri et al. 2004; Bustarret et al. 2006).

A very ingenious strategy that attains high-T_c conventional superconductivity in a doped sp^3 system avoiding negative dimensionality effects was proposed a few years later by Savini et al. (2010), i.e., realizing superconductivity in p-doped graphane (fully hydrogenated graphene). Graphane can be considered a 2D version
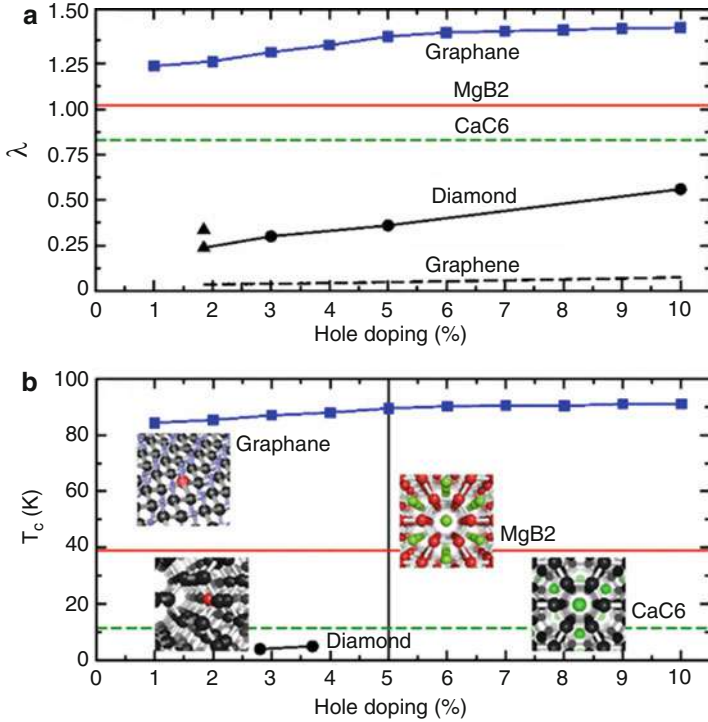
Fig. 6 Electron-phonon coupling (top) and T_c (bottom) as a function of doping in different carbon-based superconductors. (Figure reprinted with permission from G. Savini et al., Phys. Rev. Lett. 105, 037002 (2010), Copyright by the American Physical Society 2010)

of diamond, in which the bonding is still sp^3; hence, the matrix elements are as large as those of diamond, but the DOS is 2D; hence, λ is sizable already at low dopings. For this compound, the authors of Savini et al. (2010) estimate a T_c of $\sim$90 K already for 1% doping. Figure 6, from the original reference, compares the behavior of T_c with doping in the difference class of carbon-based superconductors discussed so far.

Although not as spectacular as graphane, in general, many carbon-based compounds doped with boron are good candidates for superconductivity with relatively high temperatures. For example, recent experiments indicate that Q-carbon, an amorphous form of carbon, in between diamond and graphite, can achieve T_cs as high as 56 K upon boron doping (Bhaumik et al. 2017).

A complementary idea is that of doping boron-rich phases with carbon; boron, being an electron-deficient material, forms a variety of structures with two- and three-center bonds. One of the most common motifs is the icosahedron (B_{12}), found, for example, in the α and β phases of elemental boron, as well as in superconducting dodecaborides (Matthias et al. 1968). Boron icosahedra doped with carbon are predicted to superconduct with T_cs as high as 37 K (Calandra et al. 2004).

4.4   Intercalated Graphites

While MgB_2 can be considered a natural realization of hole-doped graphite σ bonds, superconducting graphite intercalation compounds (GICs) are the practical realization of superconductivity by doping into the interlayer states. Although the T_cs are substantially lower than in covalent metals, this type of superconductivity is quite interesting, because it can be more easily manipulated by external means (doping, pressure) and also realized in "artificial" systems, such as mono- and bilayers of graphene decorated with alkali or noble metals.

Intercalated graphites had been extensively studied in the 1970s, because of their high mobilities, but the only known superconducting member of the family, KC_8, had an extremely small T_c ($<1 K$). Only in 2005, superconductivity with relatively high T_cs was reported in several AC_6 compounds: CaC_6 ($T_c = 11.5$ K), YbC_6 ($T_c = 6.5$ K), and later SrC_6 ($T_c = 1.6$ K) (Weller et al. 2005; Kim et al. 2007). At the same time, it was observed that for all newly discovered members of the family, the occurrence of superconductivity clearly correlates with the filling of the interlayer band, which is empty in non-superconductors (Csaniy et al. 2005).

A strong correlation is also found between the T_c and the distance between two subsequent graphitic layers in different AC_6 compounds – Fig. 7a. This can be easily
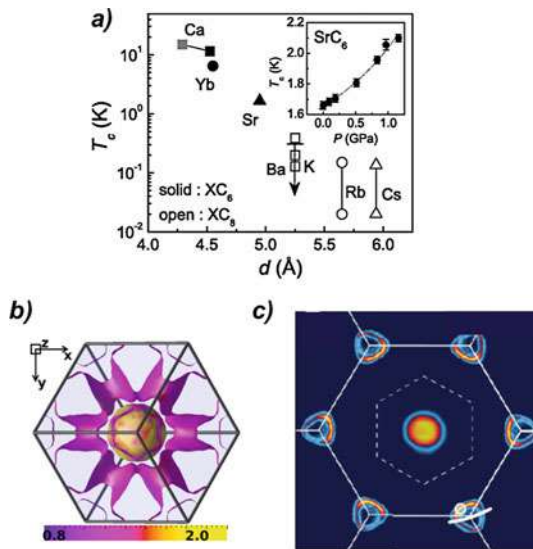
Fig. 7 Physical properties of superconducting graphite intercalation compounds: (**a**) dependence of T_c on the interlayer separation in different GICs (Figure reprinted with permission from Kim et al., Phys. Rev. Lett. 99, 027001 (2007), Copyright by the American Physical Society (2007)); (**b**) anisotropic superconducting gap predicted from SCDFT calculations in CaC_6 (Figure reprinted with permission from A. Sanna et al., Phys. Rev. B 75, 020511(R) (2007), copyright by the American Physical Society (2007)); (**c**) superconducting gap measured by ARPES in Li-decorated graphene (Figure reprinted with permission from B.M. Ludbrook et al., PNAS 112, 11795 (2015))

explained within the conventional *ep* scenario (Calandra and Mauri 2005; Kim et al. 2006, 2007; Boeri et al. 2007), since, as demonstrated in Sect. 4.1, when localized due to doping and confinement effects, interlayer electrons can couple to π electrons through out-of-plane phonons of the carbon layers (in GICs, additional coupling is also provided by intercalant modes).

Like in MgB_2, the distribution of the *ep* interaction, and hence of the superconducting gap, is very anisotropic on the Fermi surface, being much larger for the interlayer electrons – central sphere in Fig. 7b – than for π ones, which form the outer tubular manifold.

The idea of superconductivity due to the localization of interlayer states was exploited in Profeta et al. (2012), proposing to achieve superconductivity in Li-decorated graphene. A T_c of 5.9 K, close to the theoretical prediction of 8.6 K, was reported in 2015 by Ludbrook et al. (2015). Figure 7c shows the superconducting gap, measured by ARPES. Note that also in this case there is a visible variation of the gap between π and IL sheets of the Fermi surface.

4.5 High-T_c Conventional Superconductivity in High-Pressure Hydrides

The most spectacular realization of high-T_c conventional superconductivity from doped covalent bonds can be found in SH_3, which, so far, holds the T_c the record among all (conventional and unconventional) superconductors.

Indeed, superconductivity at high pressures is a rather ubiquitous phenomenon, because high pressures tend to increase the hopping between neighboring sites and hence metallicity. Almost all the elements of the periodic table can be made superconducting; the typical T_cs are rather low, reaching a maximum of $\sim 20\,K$ in Li, Ca, Sc, Y, V, B, P, and S (Hamlin 2015), well reproduced by ab initio calculations (Profeta et al. 2006; Yao et al. 2009; Monni et al. 2017; Flores-Livas et al. 2017).

Hydrogen and its compounds represent a notable exception, having provided, in 2014, the first example of high-T_c conventional superconductivity. This discovery was the coronation of a 50-year-long search, inspired by two insightful papers by Neil Ashcroft, predicting high-T_c superconductivity in metallic hydrogen (1968) and covalent hydrides (2004) (Ashcroft 1968, 2004). Both predictions rely on the general arguments for high-T_c conventional superconductivity introduced in Sect. 4.1, i.e., large phonon frequencies due to the light hydrogen mass (ii) combined with large *ep* matrix elements due to the lack of screening from core electrons (iii) can yield remarkable superconducting transition temperatures, even if the density of states at the Fermi level is moderate (i).

In 1968, the first of Ashcroft's predictions, superconductivity in metallic hydrogen, seemed merely an academic speculation. The pressure required to metallize hydrogen, which is a large-gap insulator at ambient pressure ($\Delta \sim 10\,eV$), was clearly beyond reach for the experimental techniques of the time. However, 50 years

later, high-pressure research has advanced to such a point that at least three groups have reported evidences of hydrogen metallization, at pressures ranging from 360 to 500 GPa (3.6 to 5 Mbar) (Dalladay-Simpson and Howie 2016; Eremets et al. 2017; Dias and Silvera 2017).

These reports are still controversial, since the pressure ranges are close to the limit of current high-pressure techniques (Eremets and Drozdov 2017; Goncharov and Struzhkin 2017). The insulator-to-metal transition has two possible origins: band overlap, in the molecular $CmCa$ phase, or a structural transition to an atomic $\beta - Sn$ phase. The two phases are almost impossible to discern experimentally because hydrogen is a poor X-ray scatterer, and theoretical studies also predict an unusual spread of values (300–500) GPa for the transition boundary between the two phases, due to the different approximations used to account for quantum lattice effects (McMahon et al. 2012). However, according to DFT calculations, both the $CmCa$ and the $\beta - Sn$ structures should become superconducting at or above ambient temperature, suggesting that the first of Ashcroft's predictions may soon be realized (Cudazzo et al. 2008; McMahon and Ceperley 2011; Borinaga et al. 2016).

The second prediction, superconductivity in covalent hydrides, was experimentally verified at the end of 2014. The underlying idea is that in covalent hydrides, metallization of the hydrogen sublattice should occur at lower pressure than in pure hydrogen, because the other atoms exert an additional chemical pressure.

Indeed, in 2008, superconductivity was measured in compressed silane (SiH_4) at 120 GPa, but the measured T_c (17 K) was disappointingly low compared to theoretical prediction of 100 K (Eremets et al. 2008; Li et al. 2010). However, this finding proved that the experimental knowledge to make covalent hydrides was available, and it was only a matter of time before high-T_c conventional superconductivity would actually be observed. Ab initio calculations revealed two essential missing pieces of the puzzle: (a) megabar pressures can stabilize *superhydrides*, i.e., phases with much larger hydrogen content than the hydrides stable at ambient pressure, and (b) some of these *superhydrides* are metals with unusually strong bonds, which can lead to high-T_c superconductivity (Zurek et al. 2009).

The first high-T_c superconductor discovered experimentally is a sulfur superhydride (SH_3), which is stabilized under pressure, by compressing gaseous sulfur dihydride (SH_2) in a hydrogen-rich atmosphere. The compound metallizes at $\sim$100 GPa, where it exhibits a T_c of 40 K. The T_c increases reaching a maximum of over 203 K at 200 GPa; isotope effect measurements confirmed that superconductivity is of conventional origin. I refer the reader to the original experimental (Drozdov et al. 2015a; Goncharov et al. 2016; Einaga et al. 2016; Capitani et al. 2017) and theoretical (Duan et al. 2014; Heil and Boeri 2015; José et al. 2016; Errea et al. 2015, 2016; Bernstein et al. 2015; Quan and Pickett 2016; Akashi et al. 2016; Sano et al. 2016) references for a more detailed discussion of specific aspects of the SH_3 discovery, such as the nature and thermodynamics of the SH_2 to SH_3 transition and anharmonic and nonadiabatic effects (Gor'kov and Kresin 2018).

In the history of superconducting materials, SH_3 stands out for one main reason: it is the first example of a high-T_c superconductor whose chemical composition, crystal structure, and superconducting transition temperature were predicted from

first principles before the actual experimental discovery. In fact, a few months before the experimental report, Duan et al. (2014) predicted that SH_2 and H_2 would react under pressure and give rise to a highly symmetric bcc structure, later confirmed by X-ray experiments (Einaga et al. 2016), which could reach a T_c as high as 200 K at 200 GPa.

From the point of view of electronic structure, SH_3 is a paradigmatic example of high-T_c conventional superconductivity. Indeed, in this case, all three conditions reported in Sect. 4.1 are verified: SH_3 is a hydride, and hence its characteristic vibrational frequencies are high (i); the Fermi level falls in the vicinity of a van Hove singularity of the DOS of the bcc lattice (ii); and the *ep* matrix elements are high due to the unusual H-S covalent bonds stabilized by pressure (iii).

Replicating a similar combination is not simple, even in other high-pressure hydrides, where one can hope that high pressure may help stabilize unusual bonding environments and hydrogen-rich stoichiometries. Despite an intense theoretical exploration of the high-pressure superconducting phase diagrams of binary hydrides (Zurek et al. 2009), only a few candidates match or surpass the T_c of SH_3: Ca, Y, and La (Wang et al. 2012; Liu et al. 2017).

In fact, the formation of covalent directional bonds between hydrogen and other atoms appears to be very sensitive to their electronegativity difference (Bernstein et al. 2015; Fu et al. 2016), and in binary hydrides, the possibilities to optimize this parameter are obviously limited. A possible strategy to overcome this limitation is to explore ternary hydrides, where the electronegativity, atomic size, etc. can be tuned continuously combining different elements. However, since ternary Gibb's diagrams are computationally much more expensive than binary convex hulls, fully ab initio studies of ternaries are rare. Two recent studies explore two different strategies toward high T_c in ternary hydrides: doping low-pressure molecular phases of covalent binary hydrides, like water (Flores-Livas et al. 2017), and off-stoichiometry phases of alkali-metal alanates and borates, which permit to independently tune the degree of metallicity and covalency (Kokail et al. 2017).

Due to the intrinsic difficulty of reaching megabar pressures and the limited information that can be extracted from X-ray spectra, the experimental information on high-pressure hydrides is much scarcer than theoretical predictions. Nevertheless, there has been a substantial progress in the last 5 years: metallic superhydrides predicted by theory have been reported in hydrides of alkali metals (Li,Na), (Pépin et al. 2015; Struzhkin et al. 2016) transition metals (Fe), (Pépin et al. 2017) group IV elements, etc. For some of these systems, first-principles calculations predict substantial superconducting temperatures (Zurek et al. 2009), while other cases are controversial (Majumdar et al. 2017; Kvashnin et al. 2018; Heil et al. 2018).

Resistivity and susceptibility measurements required to detect superconductivity under pressure are even more challenging, and therefore the available information on superconductivity in high-pressure hydrides is still very scarce: besides SH_3, only one other hydride, PH_3, has been shown to superconduct at high pressures, albeit with a lower T_c ($\sim$100 K) (Drozdov et al. 2015b). At variance with SH_3, PH_3 is highly metastable, and samples rapidly degrade over time, consistently with ab initio predictions of metastability (Fu et al. 2016; Flores-Livas et al. 2016; Shamp et al.

2016). SeH_3, which should exhibit T_cs comparable to SH_3, has been successfully synthesized at the end of last year (Pace et al. 2017), but superconductivity has not been measured yet.

The same arguments that motivated the search for high-T_c conventional superconductivity in high-pressure hydrides can be applied to other compounds that contain light elements, such as Li, B, C, etc. Some of these elements form covalent structures with strong directional bonds already at ambient pressure, and high pressures could be used to optimize doping, stoichiometry, etc. However, element-specific factors can unpredictably affect properties relevant in high-pressure superconductivity. For example, elemental phases of alkali metals at high pressure exhibit a characteristic *interstitial charge localization*, due to avoided core overlap. The same effect, which is extremely detrimental for conventional superconductivity (charge localized in the interstitial regions has an intrinsically low coupling to phonons), also occurs in many of their compounds. This was shown, for example, for the Li-S system, whose behavior at high pressure is remarkably different from H-S (Kokail et al. 2016).

4.6   Unconventional Superconductivity in Fe Pnictides and Chalcogenides

After illustrating the remarkable progress in the research on conventional superconductors, I have chosen to discuss the single largest class of unconventional superconductors, formed by Fe pnictides and chalcogenides (Fe-based superconductors, FeSC). The T_cs of these compounds, discovered in 2008, go up to 56 K in the bulk, and allegedly up to 100 K in monolayers of FeSe grown on $SrTiO_3$ (Kamihara et al. 2008; Ge et al. 2014; Paglione and Greene 2010; Johnston 2010; Huang and Hoffman 2017). Their rich phase diagram, the quasi-2D crystal structure and the unconventional behavior of the superconducting gaps, is strongly reminiscent of the high-T_c cuprates. Thus, I will use FeSC as a representative example to illustrate the challenges faced by ab initio methods in unconventional superconductors, both in the normal and in the superconducting state (Basov and Chubukov 2011).

Figure 8 shows the main features, common to most compounds:

- (a)–(b) A common structural motif, consisting of square planes of iron atoms, and X_4 tetrahedra; X is either a pnictogen (As,P) or a chalchogen (Se,Te).
- (c) A quasi-two-dimensional Fermi surface, comprising two hole and two electron sheets, strongly nested with one other.
- (d) A phase diagram, in which superconductivity sets in after suppressing a spin density wave (SDW) ordered state, by means of doping or pressure. The most common SDW pattern is a stripe one, in which the Fe spins are aligned ferromagnetically along one of the edges of the Fe squares and antiferromagnetically along the other. The SDW transition is usually preceded by a structural-nematic transition (Fernandes et al. 2014).
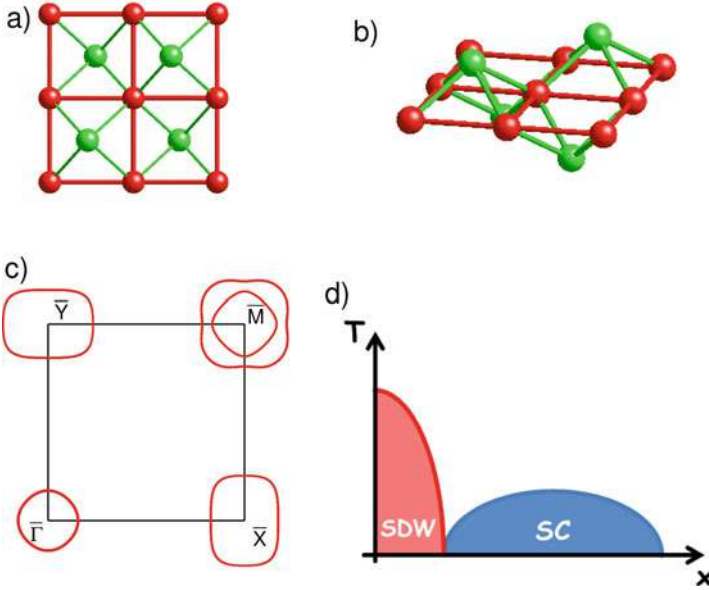
Fig. 8 Common features of Fe-based superconductors (FeSC): FeX layers, seen from the top (**a**) and side (**b**); Fe and X atoms are red and green, respectively. (**c**) Two-dimensional Fermi surface of LaOFeAs, from Andersen and Boeri (2011), with the hole and electron pockets centered at $\bar{X}$, $\bar{Y}$, and $\bar{M}$, respectively. Here a third hole pocket at $\bar{\Gamma}$ is also present. (**d**) Phase diagram, showing the transition from SDW to superconducting (SC) state; x is an external tuning parameter (doping, pressure)

FeSC can be grouped in different *families*, depending on the nature of the FeX layers and of the intercalating blocks; the most common are 11 Fe chalcogenides (FeS, FeSe, FeTe), 111 alkali-metal pnictides (LiFeAs, LiFeP, NaFeAs, NaFeP, etc.), 122 pnictides (Ba/KFe$_2$As$_2$, BaFe$_2$P$_2$, CaFeAs$_2$, EuFe$_2$As$_2$, etc.), 1111 pnictides (LaOFeAs, LaOFeP, etc.), and 122 chalcogenides (KFe$_2$Se$_2$, etc.). In most of these cases, superconductivity appears around a Fe d^6 configuration, but it survives up to high hole or electron dopings in 122 K pnictides and chalcogenides.

In most FeSC, the superconducting gap exhibits a feature which is distinctive of unconventional superconductors, i.e., changes sign over the Fermi surface. According to Eq. 1, this is only possible if, unlike *ep* interaction, the pairing interaction $V_{\mathbf{k},\mathbf{k}'}$ is repulsive over some regions of reciprocal space.

In contrast to the cuprates, where the Fermi surface topology favors d-wave superconductivity, the gap of most FeSC exhibits a characteristic $s^{\pm}$ symmetry, with opposite signs on the hole and electron sheets of the Fermi surface. However, substantial variations in the symmetry and magnitude of the superconducting gap are observed among and within different FeSC families, which can be related to changes in the general shape and orbital distribution of Fermi surface sheets (Hirschfeld et al. 2011).

The general topology of the Fermi surface is well described by DFT calculations; however, quasiparticle bands measured by ARPES exhibit renormalizations and shifts, which hint to strong local correlation effects beyond DFT, which require the use of specialized methods, such as DFT+DMFT (Kotliar et al. 2006).

Figure 9a, from Yin et al. (2011), shows the Fermi surfaces calculated within DFT (upper panel) and DFT+DMFT (lower panel) for representative FeSC: LaFePO, $BaFe_2As_2$, LiFeAs, and KFe_2As_2. The first three are d^6 pnictides, with the typical hole-electron topology shown in Fig. 8, eventually deformed due to interlayer hopping (Andersen and Boeri 2011). In KFe_2As_2, where the electron count is $d^{5.5}$, the hole pocket is expanded, and the electron pocket has lost its typical circular shape.

The figure shows that, in most cases, the inclusion of electronic correlation modifies the Fermi surface quantitatively, but not qualitatively, with respect to the DFT prediction. The renormalization of the quasiparticle band dispersion depends on their orbital character, being stronger for Fe t_2 (xz, yz, xy) than for e ($x^2 - y^2$, $3z^2 - 1$) orbitals, and on the nature of the X atom; correlation effects are in fact weakest for $X=P$ and increase moving to As,S,Se and Te (de' Medici et al. 2014; Yin et al. 2011). Orbital-selective mass renormalizations and poor Fermi liquid behavior are characteristic of a special correlation regime, known as *Hund's metal* regime, which has been extensively studied by several authors (Werner et al. 2008; Haule and Kotliar 2009; de' Medici et al. 2011; Yin et al. 2011); excellent reviews can be found in Georges et al. (2013) and de' Medici (2017).

A very important consequence of the interplay between metallicity, strong correlations, and multiband character in FeSC is their anomalous magnetic behavior. Early DFT studies pointed out that magnetism cannot be consistently described within neither the itinerant (Slater) nor the localized (Heisenberg) scenario (Mazin et al. 2008b; Mazin and Johannes 2008). A regime which is intermediate between the two cannot be treated in DFT, which is a mean-field theory, but requires methods able to capture the *dynamics* of the magnetic moments at different frequency scales (Yin et al. 2011; Aichhorn et al. 2011; Hansmann et al. 2010; Toschi et al. 2012; Skornyakov et al. 2011; Schickling et al. 2012).

Panel (c) shows a DFT+DMFT calculation of the magnetically ordered state for a variety of FeSC (Yin et al. 2011). Circles (and stars) indicate the theoretical (experimental) value of the ordered SDW magnetic moment, which shows large variations among different compounds. Chalcogenides exhibit large ordered moments, while 111 and 1111 pnictides and phosphides exhibit smaller values. (Gray) Squares indicate the value of the local magnetic moment, which oscillates over a much smaller range of values around 2.5 μ_B. In the SDW phase, the two moments coexist; when the itinerant moment is suppressed with doping or pressure, only the local moment survives, meaning that in the nonmagnetic regions of the phase diagram, FeSC is in a paramagnetic state.

The superconducting state that emerges from this complicated normal state is clearly unconventional. According to linear response calculations, *ep* coupling plays a marginal role in the superconducting pairing, since the values of the calculated coupling constant λ in all FeSC are extremely small, even including
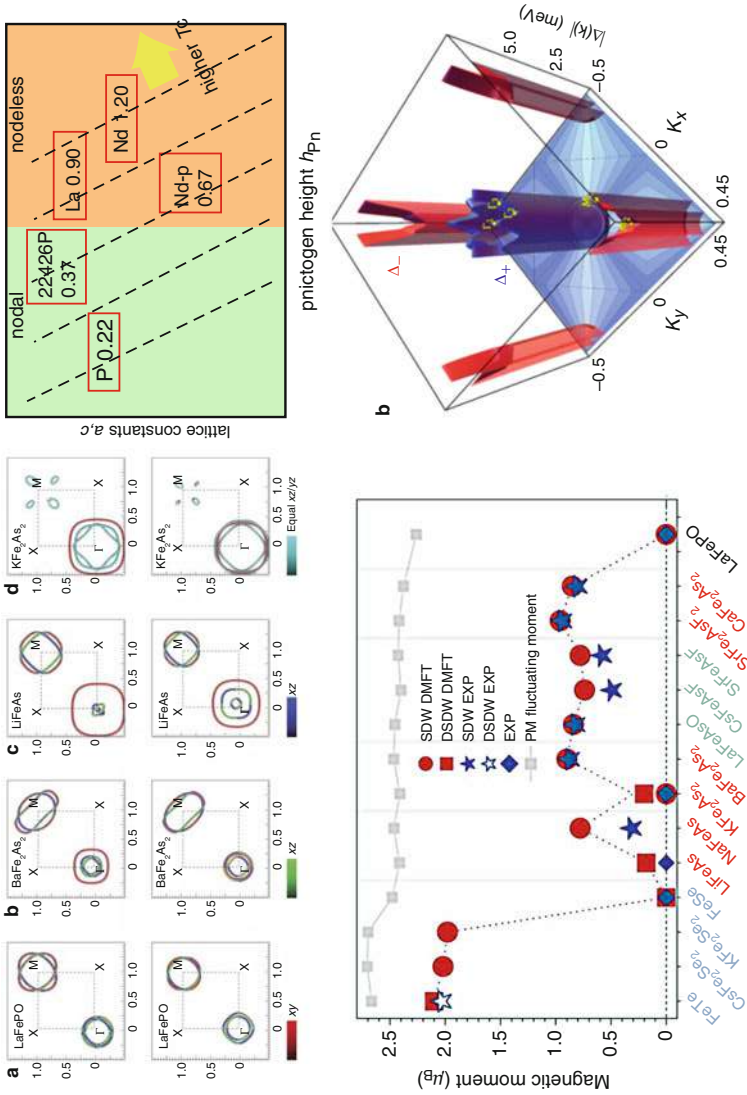
Fig. 9 Normal and superconducting-state properties of FeSC, predicted by DFT+many-body methods. (**a**) Renormalization of the Fermi surfaces and (**b**) variation of the ordered and local magnetic moment across different families of FeSC, predicted by in DFT+DMFT (Figures reprinted with permission from Nature publishing group, Y. Yin et al., Nat. Phys. 7, 294 (2011), Copyright Nature (2011)); (**c**) superconducting trends in 1111 pnictides predicted by DFT+RPA (Figures reprinted with permission from K. Kuroki et al., Phys. Rev. B (2009), Copyright American Physical Society (2009)); (**d**) superconducting gap of FeSe predicted by SCDFT (Figure reprinted with permission from F. Essenberger et al., Phys. Rev. B 79, 224511 (2009), Copyright American Physical Society (2009)).

magnetoelastic effects[1] (Boeri et al. 2008, 2010; Subedi et al. 2008). The strongest candidate as superconducting mediator are spin fluctuations, as suggested by the unconventional symmetry of the superconducting gap and the proximity of SDW and superconductivity in the phase diagram (Chubukov et al. 2008; Kuroki et al. 2008; Mazin et al. 2008a). As discussed in Sect. 3.2, there is currently no first-principles theory of spin fluctuations with an accuracy comparable to that for *ep* interaction. Most studies of superconductivity in FeSC have therefore adopted a *hybrid* approach, in which the electronic structure in the vicinity of the Fermi level is downfolded or projected to an effective analytical tight-binding model (Graser et al. 2009; Kuroki et al. 2009; Miyake et al. 2010; Andersen and Boeri 2011), and electron-electron interactions leading to spin and charge fluctuations are included at a second stage with many-body techniques. The most common are the random phase approximation (RPA), fluctuation exchange (FLEX), and functional renormalization group (FRG) (Chubukov et al. 2008; Thomale et al. 2009; Hirschfeld et al. 2011; Platt et al. 2012). Although not quantitatively predictive, these *weak-coupling* approaches have provided a detailed understanding of the superconducting gap symmetry, competition of superconductivity with other ordered phases, and occurrence of nematic order and traced their common origin to multi-orbital physics (Yi et al. 2017). Figure 9c, from Kuroki et al. (2009), shows that the variations in T_c across different 1111 pnictides can be reproduced within RPA and derive from changes in orbital composition of the Fermi surface.

A first, very elegant, fully first-principles study of superconductivity in FeSC was carried out by the Gross group, who applied their recently derived ab initio theory for spin fluctuations to FeSe – see Sect. 3.2 for details. Figure 9d shows the superconducting gap calculated in SCDFT, which exhibits an $s^{\pm}$ symmetry, consistently with previous *hybrid* studies. It is important to remark, however, that the gap in Fig. 9d was calculated treating the contributions of all pairing channels (phonons, spin fluctuations, charge fluctuations) on equal footing and without any intermediate mapping on an effective many-body model. However, even this calculation cannot be considered fully ab initio as, in order to obtain a physically meaningful value for the superconducting gap and T_c, the authors had to introduce an artificial *scaling* factor in the spin-fluctuation propagator, which would otherwise diverge (Essenberger et al. 2014; Ortenzi et al. 2012).

The divergence is associated to the specific choice of TDDFT exchange functional (adiabatic local density approximation, ALDA) made by the authors which, being the time-dependent equivalent to the standard local density approximation (LDA), overestimates the tendency to magnetism in FeSC (Mazin et al. 2008b). Proposals to cure this critical divergence in a nonempirical fashion are underway (Sharma et al. 2018) and could lead to a very important step forward in the

---

[1](This result is generally accepted, although DFT+DMFT studies evidenced a strong renormalization of some phonon modes, due to strong electronic correlations (Mandal et al. 2014) *ep* coupling has also been suggested to play a primary role in the enhancement of the superconducting T_c in FeSe monolayers grown on $SrTiO_3$ (Huang and Hoffman 2017), although in this case, the modes involved in the pairing belong to the substrate.)

understanding of unconventional superconductors. However, it is also important to stress that, even if divergence issues are solved, T_c computed within DFT may still be inaccurate, because it neglects renormalization effects of quasiparticle energies and interaction vertices due to strong local electronic correlations, which, in some FeSC, may be sizable (Yin et al. 2013; de' Medici et al. 2017).

5   Outlook and Perspectives

The results described in this chapter show that in the last 20 years, there has been a substantial advancement in the understanding of superconductors, driven by the progress of ab initio electronic structure methods. The SH_3 discovery has demonstrated that room-temperature superconductivity can be attained, at least at extreme pressures (Drozdov et al. 2015a), and that ab initio methods can be very effectively employed to predict new superconductors. The next challenge in the field is clearly to devise practical strategies to replicate the same result at ambient conditions, exploiting novel synthesis and doping techniques. Obviously, the development of ab initio methods that can treat these regimes is a crucial step in this direction.

   The aim of this final section is to give a short overview of promising strategies to high-T_c superconductivity, exemplified in Fig. 10a–e.

(a) **Design by Analogy:** This is one of the most common routes to search for new superconductors, i.e., achieve high-T_c designing materials with a similar geometry, chemistry, electronic structure, etc. as the best existing superconductors. This route has lead to several predictions of high-T_c superconductivity in layered borides and carbides after the discovery of MgB_2, as well as hole-doped LiBC, and, more recently, doped graphane (Satta et al. 2001; Rosner et al. 2002; Savini et al. 2010). The main drawback of many of these early works is that many of these hypothetical compounds are thermodynamically unstable. Nowadays, studies of the thermodynamic stability of compounds are well established but were very rare at the time of the MgB_2 discovery. One of the first studies to take this aspect into account is the prediction of superconductivity in LiB, (Kolmogorov and Curtarolo 2006) shown in Fig. 10a.

(b) **Chemical Doping of Molecular Crystals:** Similar to covalent solids, molecular ones exhibit stiff bonds, and hence their electronic states can couple strongly to lattice vibrations, which is a prerequisite for high-T_c conventional superconductivity. However, similar to covalent solids, molecular ones are usually insulating at ambient pressures. Making them superconducting requires either very high pressures or chemical doping, which is often hard to control experimentally and to model ab initio.

   One of the most complete studies of the effect of doping on superconductivity in molecular crystals is the study of superconductivity in doped ice, by Flores-Livas et al. (2017) Here, analyzing the effect of different dopants using supercells, the authors found that nitrogen acts as an effective dopant
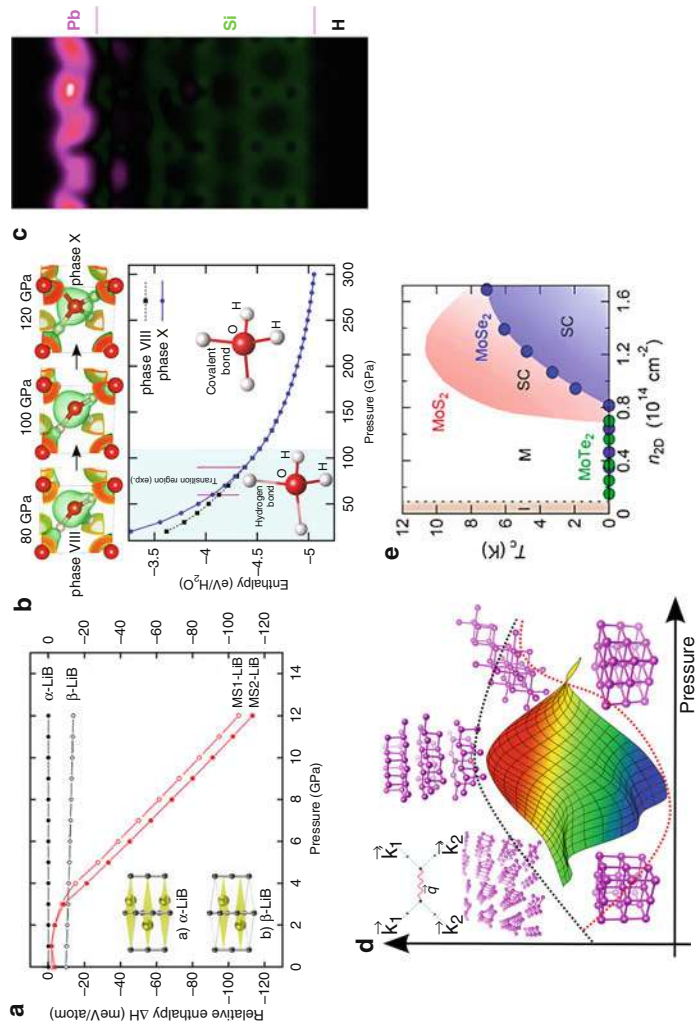
Fig. 10 Selected literature examples of possible strategies to room-temperature superconductivity at ambient pressures: (**a**) first-principles prediction of α and β-LiB (Figure reprinted with permission from A. Kolmogorov and S. Curtarolo, Phys. Rev. B 74, 224507 (2006), Copyright by the American Physical Society (2006)) (*Design by Analogy*); (**b**) superconductivity in doped ices (Flores-Livas et al. 2017) (*Chemical Doping of molecular crystals*); (**c**) charge-density profile of superconducting Pb@Si(111) surface (Linscheid et al. 2015a) (*Atomic-scale design and Dimensionality*); (**d**) high-pressure superconducting phase diagram of phosphorus (Figure reprinted with permission from Flores Livas et al., Phys. Rev. Mat. 1 024802 (2017), Copyright by the American Physical Society (2017)); (*Quenching of High-pressure metastable phases*); (**e**) T$_c$ vs field effect and chemical doping in monolayer MoS$_2$ (W. Shi et al. Sci. Rep. 5, 12534 (2015)) (*Doping by field effect*)

at the oxygen site, leading to a T_c of 60 K. The study nicely evidences the crucial difference between actual doping, modeled with supercells, and more approximate approaches, as rigid-band or jellium doping.

(c) **Atomic-Scale Design and Dimensionality:** Another recent trend in condensed matter research opened by the development of novel techniques, such as molecular beam epitaxy (MBE), or exfoliation (Novoselov et al. 2004), is the design of materials at the atomic scale, through heterostructuring, controlled deposition, strain engineering, etc. This permits to tune superconductivity in existing materials; induce it in semiconductors and semimetals through doping, like Li-decorated graphene and phosphorene; or create completely artificial superconductors, depositing superconducting elements on semiconductor surfaces, as in Fig. 10c, which shows a SCDFT calculation for Pb on the 111 surface (Zhang et al. 2010; Linschied et al. 2015a). For this type of problems, real-space approaches may be more appropriate than reciprocal-space ones (Linscheid et al. 2015b).

(d) **Quenching of High-Pressure Metastable Phases** down to ambient pressure is another very attractive route to stabilize unusual bonding environments, enabled by recent developments in high-pressure techniques. In fact, controlled heating and cooling cycles permit to selectively stabilize different metastable phases, realizing pressure hysteresis cycles. Figure 10d, from Flores-Livas et al. (2017), shows the crystal structures of the different phases that form the complicated superconducting phase diagram of elemental phosphorus. In this element, a high-T_c branch, clearly distinct from the low-T_c ground-state one, can be accessed by laser heating and is associated with metastable black phosphorus; similar branchings between high-T_c and low-T_c phases have been predicted at higher pressures.

(e) Doping via **Field Effect** is an attractive route to tune the properties of materials continuously, without impurities and distortions associated to chemical doping. Modern techniques based on liquid electrolytes permit to achieve doping levels as high as $10^{15} carriers/cm^3$, corresponding to tenths of electrons, around three orders of magnitude larger than standard solid-state techniques. The first successful applications have been to cuprates, ZrNiCl, and 2D materials – Fig. 10e. Rigorous modeling of field-effect devices has been derived in Sohier et al. (2015), Sohier et al. (2017).

Although not an experimental technique, methods for machine learning, which can be used to prescreen ab initio proposals, will most likely play a bigger and bigger role in the design and discovery of new materials (Curtarolo et al. 2013).

Note that all of the specific examples discussed above are based on the assumption of *conventional* (phonon-mediated) pairing. The same techniques have been, and can be, applied also to *unconventional* superconductors, which remain the likeliest candidates for high-T_c superconductivity. However, without a quantitative theory of the superconducting pairing, it is at the moment impossible to formulate reliable predictions of T_c and other superconducting properties. A progress in this direction is therefore an essential prerequisite to any meaningful computational search.

References

Aichhorn M, Pourovskii L, Georges A (2011) Importance of electronic correlations for structural and magnetic properties of the iron pnictide superconductor LaFeAsO. Phys Rev B 84:054529

Akashi R, Arita R (2013) Development of density-functional theory for a plasmon-assisted superconducting state: application to lithium under high pressures. Phys Rev Lett 111:057006

Akashi R, Sano W, Arita R, Tsuneyuki S (2016) Possible "magnéli" phases and self-alloying in the superconducting sulfur hydride. Phys Rev Lett 117:075503

Allen P, Dynes R (1975) Transition temperature of strong-coupled superconductors reanalyzed. Phys Rev B 12:905

Allen P, Mitrovic B (1982) Theory of superconducting T_c. In: Solid State Physics, vol 37. Academic, New York, pp 1–92

An JM, Pickett WE (2001) Superconductivity of MgB_2: covalent bonds driven metallic. Phys Rev Lett 86(19):4366–4369

Andersen O, Boeri L (2011) On the multi-orbital band structure and itinerant magnetism of iron-based superconductors. Annalen der Physik 523(1–2):8–50

Ashcroft N (1968) Metallic hydrogen: a high-temperature superconductor? Phys Rev Lett 21:1748–1749

Ashcroft N (2004) Hydrogen dominant metallic alloys: high temperature superconductors? Phys Rev Lett 92:187002

Bardeen J, Cooper LN, Schrieffer JR (1957) Theory of superconductivity. Phys Rev 108:1175

Baroni S, de Gironcoli S, Corso AD, Giannozzi P (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73:515

Basov DN, Chubukov AV (2011) Manifesto for a higher Tc. Nat Phys 7:272 EP Perspective

Bednorz J, Mueller K (1986) Possible high T_c superconductivity in the Ba-La-Cu-O system. Zeit Phys B 64:189

Berk NF, Schrieffer JR (1966) Effect of ferromagnetic spin correlations on superconductivity. Phys Rev Lett 17:433–435

Bernstein N, Hellberg CS, Johannes MD, Mazin II, Mehl MJ (2015) What superconducts in sulfur hydrides under pressure and why. Phys Rev B 91:060511

Bhaumik A, Sachan R, Narayan J (2017) High-temperature superconductivity in boron-doped q-carbon. ACS Nano 11(6):5351–5357. PMID:28448115

Blase X, Adessi C, Connétable D (2004) Role of the dopant in the superconductivity of diamond. Phys Rev Lett 93:237004

Boeri L, Bachelet G, Cappelluti E, Pietronero L (2002) Small fermi energy and phonon anharmonicity in MgB_2 and related compounds. Phys Rev B 65:214501

Boeri L, Kortus J, Andersen OK (2004) Three-dimensional MgB_2-type superconductivity in hole-doped diamond. Phys Rev Lett 93:237002

Boeri L, Cappelluti E, Pietronero L (2005) Small fermi energy, zero-point fluctuations, and nonadiabaticity in MgB$_2$. Phys Rev B 71:012501

Boeri L, Bachelet GB, Giantomassi M, Andersen OK (2007) Electron-phonon interaction in graphite intercalation compounds. Phys Rev B 76:064510

Boeri L, Dolgov OV, Golubov AA (2008) Is LaFeAsO$_{1-x}$F$_x$ an electron-phonon superconductor? Phys Rev Lett 101:026403

Boeri L, Calandra M, Mazin II, Dolgov OV, Mauri F (2010) Effects of magnetism and doping on the electron-phonon coupling in BaFe$_2$As$_2$. Phys Rev B 82:020506

Borinaga M, Errea I, Calandra M, Mauri F, Bergara A (2016) Anharmonic effects in atomic hydrogen: superconductivity and lattice dynamical stability. Phys Rev B 93:174308

Boschker H, Mannhart J (2017) Quantum-matter heterostructures. Ann Rev Condens Matter Phys 8(1):145–164

Bustarret E, Marcenat C, Achatz P, Kačmarčik J, Lévy F, Huxley A, Ortéga L, Bourgeois E, Blase X, Débarre D, Boulmer J (2006) Superconductivity in doped cubic silicon. Nature 444:465

Calandra M, Mauri F (2005) Theoretical explanation of superconductivity in C$_6$Ca. Phys Rev Lett 95:237002

Calandra M, Vast N, Mauri F (2004) Superconductivity from doping boron icosahedra. Phys Rev B 69:224505

Calandra M, Profeta G, Mauri F (2010) Adiabatic and nonadiabatic phonon dispersion in a Wannier function approach. Phys Rev B 82:165111

Capitani F, Langerome B, Brubach JB, Roy P, Drozdov A, Eremets MI, Nicol EJ, Carbotte JP, Timusk T (2017) Spectroscopic evidence of a new energy scale for superconductivity in H$_3$S. Nat Phys 13:859 EP Article

Carbotte JP (1990) Properties of boson-exchange superconductors. Rev Mod Phys 62:1027–1157

Casula M, Calandra M, Mauri F (2012) Local and nonlocal electron-phonon couplings in k$_3$ picene and the effect of metallic screening. Phys Rev B 86:075445

Chang KJ, Dacorogna MM, Cohen ML, Mignot JM, Chouteau G, Martinez G (1985) Superconductivity in high-pressure metallic phases of si. Phys Rev Lett 54:2375–2378

Choi JH, David R, Hong S, Cohen ML, Louie SG (2002) The origin of the anomalous superconducting properties of MgB$_2$. Nature 418:758

Chu CW, Deng LZ, Lv B (2015) Hole-doped cuprate high temperature superconductors. Phys C Supercond Appl 514:290–313. Superconducting materials: conventional, unconventional and undetermined (Eds JE Hirsch, MB Maple, F Marsiglio)

Chubukov AV, Efremov DV, Eremin I (2008) Magnetism, superconductivity, and pairing symmetry in iron-based superconductors. Phys Rev B 78:134512

Cohen M, Anderson P (1972) Superconductivity in $d-$ and $f-$ band metals. Amer Inst of Phys, New York

Csaniy G, Littlewood PB, Nevidomskyy AH, Pickard CJ, Simons BD (2005) The role of the interlayer state in the electronic structure of superconducting graphite intercalated compounds. Nat Phys 1:42

Cudazzo P, Profeta G, Sanna A, Floris A, Continenza A, Massidda S, Gross EKU (2008) Ab initio description of high-temperature superconductivity in dense molecular hydrogen. Phys Rev Lett 100:257001

Curtarolo S, Hart GLW, Buongiorno Nardelli M, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12:191–201

Dalladay-Simpson EG, Howie RT (2016) Evidence for a new phase of dense hydrogen above 325 gigapascals. Nature 529:63–67

de' Medici L (2017) Hund's metals, explained. arXiv/cond-mat/1707.03282

de' Medici L, Mravlje J, Georges A (2011) Janus-faced influence of Hund's rule coupling in strongly correlated materials. Phys Rev Lett 107:256401

de' Medici L, Giovannetti G, Capone M (2014) Selective Mott physics as a key to iron superconductors. Phys Rev Lett 112:177001

de' Medici L (2017) Hund's induced Fermi-liquid instabilities and enhanced quasiparticle interactions. Phys Rev Lett 118:167003

Dias RP, Silvera IF (2017) Observation of the Wigner-Huntington transition to metallic hydrogen. Science 355(6326):715–718

Drozdov AP, Eremets MI, Troyan IA, Ksenofontov V, Shylin SI (2015a) Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system. Nature 525: 73–76

Drozdov A, Eremets MI, Troyan IA (2015b) Superconductivity above 100 K in PH3 at high pressures. arXiv/cond-mat/1508.06224

Duan D, Liu Y, Tian F, Li D, Huang X, Zhao Z, Yu H, Liu B, Tian W, Cui T (2014) Pressure-induced metallization of dense $(H_2S)_2H_2$ with high-tc superconductivity. Sci Rep 4:6968

Einaga M, Sakata M, Ishikawa T, Shimizu K, Eremets MI, Drozdov AP, Troyan IA, Hirao N, Ohishi Y (2016) Crystal structure of the superconducting phase of sulfur hydride. Nat Phys 12:835 EP

Ekimov EA, Sidorov VA, Bauer ED, Mel'nik N, Curro NJ, Thompson J, Stishov S (2004) Superconductivity in diamond. Nature 428:542

Eliashberg GM (1960) Interactions between electrons and lattice vibrations in a superconductor Sov Phys JETP 11:696

Eremets MI, Trojan IA, Medvedev SA, Tse JS, Yao Y (2008) Superconductivity in hydrogen dominant materials: silane. Science 319(5869):1506–1509

Eremets M, Troyan I, Drozdov A (2017) Low temperature phase diagram of hydrogen at pressures up to 380 GPa. A possible metallic phase at 360 GPa and 200 K. arXiv/cond-mat 1601.04479

Eremets M, Drozdov AP (2017) Comments on the claimed observation of the Wigner-Huntington transition to metallic hydrogen. arxiv-condmat/1702.05125

Errea I, Calandra M, Mauri F (2014) Anharmonic free energies and phonon dispersions from the stochastic self-consistent harmonic approximation: application to platinum and palladium hydrides. Phys Rev B 89:064302

Errea I, Calandra M, Pickard CJ, Nelson J, Needs RJ, Li Y, Liu H, Zhang Y, Ma Y, Mauri F (2015) High-pressure hydrogen sulfide from first principles: a strongly anharmonic phonon-mediated superconductor. Phys Rev Lett 114:157004

Errea I, Calandra M, Pickard CJ, Nelson JR, Needs RJ, Li Y, Liu H, Zhang Y, Ma Y, Mauri F (2016) Quantum hydrogen-bond symmetrization in the superconducting hydrogen sulfide system. Nature 532:81–84

Essenberger F, Buczek P, Ernst A, Sandratskii L, Gross EKU (2012) Paramagnons in FeSe close to a magnetic quantum phase transition: ab initio study. Phys Rev B 86:060412

Essenberger F, Sanna A, Linscheid A, Tandetzky F, Profeta G, Cudazzo P, Gross EKU (2014) Superconducting pairing mediated by spin fluctuations from first principles. Phys Rev B 90:214504

Essenberger F, Sanna A, Buczek P, Ernst A, Sandratskii L, Gross EKU (2016) Ab initio theory of iron-based superconductors. Phys Rev B 94:014503

Fernandes RM, Chubukov AV, Schmalian J (2014) What drives nematic order in iron-based superconductors? Nat Phys 10:97 EP Review Article

Flores-Livas JA, Sanna A, Graužinytė M, Davydov A, Goedecker S, Marques MAL (2017) Emergence of superconductivity in doped H_2O ice at high pressure. Sci Rep, 7:6825 https://doi.org/10.1038/s41598-017-07145-4 arxiv-condmat/1610.04110

Flores-Livas JA, Amsler M, Heil C, Sanna A, Boeri L, Profeta G, Wolverton C, Goedecker S, Gross EKU (2016) Superconductivity in metastable phases of phosphorus-hydride compounds under high pressure. Phys Rev B 93:020508

Flores-Livas JA, Sanna A, Drozdov AP, Boeri L, Profeta G, Eremets M, Goedecker S (2017) Interplay between structure and superconductivity: metastable phases of phosphorus under pressure. Phys Rev Mater 1:024802

Fu Y, Du X, Zhang L, Peng F, Zhang M, Pickard CJ, Needs RJ, Singh DJ, Zheng W, Ma Y (2016) High-pressure phase stability and superconductivity of pnictogen hydrides and chemical trends for compressed hydrides. Chem Mater 28(6):1746–1755

Gaspari GD, Gyorffy BL (1972) Electron-phonon interactions, d resonances, and superconductivity in transition metals. Phys Rev Lett 28:801–805

Ge JF, Liu ZL, Liu C, Gao CL, Qian D, Xue QK, Liu Y, Jia JF (2014) Superconductivity above 100 k in single-layer FeSe films on doped $SrTiO_3$. Nat Mater 14:285

Georges A, de' Medici L, Mravlje J (2013) Strong correlations from Hund's coupling. Ann Rev Condens Matter Phys 4(1):137–178

Giustino F (2017) Electron-phonon interactions from first principles. Rev Mod Phys 89:015003

Giustino F, Cohen ML, Louie SG (2007a) Electron-phonon interaction using Wannier functions. Phys Rev B 76:165108

Giustino F, Yates JR, Souza I, Cohen ML, Louie SG (2007b) Electron-phonon interaction via electronic and lattice Wannier functions: Superconductivity in boron-doped diamond reexamined. Phys Rev Lett 98:047005

Goncharov AF, Struzhkin VV (2017) Comment on observation of the Wigner-Huntington transition to metallic hydrogen. preprint arXiv:1702.04246

Goncharov AF, Lobanov SS, Kruglov I, Zhao XM, Chen XJ, Oganov AR, Konôpková Z, Prakapenka VB (2016) Hydrogen sulfide at high pressure: change in stoichiometry. Phys Rev B 93:174105

Gonnelli RS, Daghero D, Ummarino GA, Stepanov VA, Jun J, Kazakov SM, Karpinski J (2002) Direct evidence for two-band superconductivity in MgB_2 single crystals from directional point-contact spectroscopy in magnetic fields. Phys Rev Lett 89:247004

Gor'kov LP, Kresin VZ (2018) Colloquium: high pressure and road to room temperature superconductivity. Rev Mod Phys 90:011001

Graser S, Maier TA, Hirschfeld PJ, Scalapino DJ (2009) Near-degeneracy of several pairing channels in multiorbital models for the Fe pnictides. New J Phys 11(2):025016

Grimaldi C, Pietronero L, Straessler S (1995) Nonadiabatic superconductivity: electron-phonon interaction beyond Migdal's theorem. Phys Rev Lett 75:1158

Gunnarsson O (1997) Superconductivity in fullerides. Rev Mod Phys 69:575–606

Gurevich A (2011) To use or not to use cool superconductors? Nature Mat 10:255

Hamlin J (2015) Superconductivity in the metallic elements at high pressures. Phys C Supercond Appl 514:59–76. Superconducting materials: conventional, unconventional and undetermined (Eds JE Hirsch, MB Maple, F Marsiglio)

Hansmann P, Arita R, Toschi A, Sakai S, Sangiovanni G, Held K (2010) Dichotomy between large local and small ordered magnetic moments in iron-based superconductors. Phys Rev Lett 104:197002

Haule K, Kotliar G (2009) Coherence-incoherence crossover in the normal state of iron oxypnictides and importance of Hund's rule coupling. New J Phys 11(2):025021

Heil C, Boeri L (2015) Influence of bonding on superconductivity in high-pressure hydrides. Phys Rev B 92:060508

Heil C, Bachelet GB, Boeri L (2018) Absence of superconductivity in iron polyhydrides at high pressures. Phys. Rev. B 97:214510 https://journals.aps.org/prb/abstract/10.1103/PhysRevB.97.214510

Hirschfeld PJ, Korshunov MM, Mazin II (2011) Gap symmetry and structure of Fe-based superconductors. Rep Prog Phys 74(12):124508

Hoesch M, Fukuda T, Mizuki J, Takenouchi T, Kawarada H, Sutter JP, Tsutsui S, Baron AQR, Nagao M, Takano Y (2007) Phonon softening in superconducting diamond. Phys Rev B 75:140508

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864

Huang D, Hoffman JE (2017) Monolayer FeSe on $SrTiO_3$. Ann Rev Condens Matter Phys 8(1):311–336

Johnston DC (2010) The puzzle of high temperature superconductivity in layered iron pnictides and chalcogenides. Adv Phys 59(6):803–1061

Jones RO, Gunnarsson O (1989) The density functional formalism, its applications and prospects. Rev Mod Phys 61:689–746

José FL, Sanna A, Gross EKU (2016) High temperature superconductivity in sulfur and selenium hydrides at high pressure. Eur Phys J B 89(3):63

Kamihara Y, Watanabe T, Hirano M, Hosono H (2008) Iron-based layered superconductor LaOFeAs. J Am Chem Soc 130(11):3296–3297

Kasahara Y, Kuroki K, Yamanaka S, Taguchi Y (2015) Unconventional superconductivity in electron-doped layered metal nitride halides MNX (M = Ti, Zr, Hf; X = Cl, Br, I). Phys C Supercond Appl 514:354–367. Superconducting materials: conventional, unconventional and undetermined (Eds JE Hirsch, MB Maple, F Marsiglio)

Kim JS, Boeri L, Kremer RK, Razavi FS (2006) Effect of pressure on superconducting ca-intercalated graphite CaC_6. Phys Rev B 74:214513

Kim JS, Boeri L, O'Brien JR, Razavi FS, Kremer RK (2007) Superconductivity in heavy alkaline-earth intercalated graphites. Phys Rev Lett 99:027001

Klemm RA (2015) Pristine and intercalated transition metal dichalcogenide superconductors. Phys C Supercond Appl 514:86–94. Superconducting materials: conventional, unconventional and undetermined (Eds JE Hirsch, MB Maple, F Marsiglio)

Kolmogorov AN, Curtarolo S (2006) Prediction of different crystal structure phases in metal borides: a lithium monoboride analog to MgB_2. Phys Rev B 73:180501

Kuroki K, Onari S, Arita R, Usui H, Tanaka Y, Kontani H, Aoki H (2008) Unconventional pairing originating from the disconnected fermi surfaces of superconducting LaFeAsO$_{1-x}$F$_x$. Phys Rev Lett 101:087004

Kuroki K, Usui H, Onari S, Arita R, Aoki H (2009) Pnictogen height as a possible switch between high-T_c nodeless and low-T_c nodal pairings in the iron-based superconductors. Phys Rev B 79:224511

Kokail C, Heil C, Boeri L (2016) Search for high-T_c conventional superconductivity at megabar pressures in the lithium-sulfur system. Phys Rev B 94:060502

Kokail C, von der Linden W, Boeri L (2017) Prediction of high-T_c conventional superconductivity in the ternary lithium borohydride system. Phys Rev Mater 1:074803

Kong Y, Dolgov OV, Jepsen O, Andersen OK (2001) Electron-phonon interaction in the normal and superconducting states of MgB_2. Phys Rev B 64(2):020501

Kortus J, Mazin II, Belashchenko KD, VP Antropov, Boyer LL (2001) Superconductivity of metallic boron in mgb$_2$. Phys Rev Lett 86:4656–4659

Kotliar G, Savrasov SY, Haule K, Oudovenko VS, Parcollet O, Marianetti CA (2006) Electronic structure calculations with dynamical mean-field theory. Rev Mod Phys 78:865–951

Kvashnin AG, Kruglov IA, Semenok DV, Oganov AR (2018) Iron superhydrides FeH_5 and FeH_6: stability, electronic properties, and superconductivity. J Phys Chem C 122(8):4731–4736

Lazzeri M, Calandra M, Mauri F (2003) Anharmonic phonon frequency shift in MgB_2. Phys Rev B 68:220509

Lee KW, Pickett WE (2004) Superconductivity in boron-doped diamond. Phys Rev Lett 93: 237003

Li Y, Gao G, Xie Y, Ma Y, Cui T, Zou G (2010) Superconductivity at approximately 100 k in dense $SiH_4(H_2)_2$ predicted by first principles. Proc Nat Acad Sci 107(36):15708–15711

Linschied A, Sanna A, Gross EKU (2015a) Ab-initio calculation of a Pb single layer on a Si substrate: two-dimensionality and superconductivity. airXiv cond-mat/1503.00977

Linscheid A, Sanna A, Floris A, Gross EKU (2015b) First-principles calculation of the real-space order parameter and condensation energy density in phonon-mediated superconductors. Phys Rev Lett 115:097002

Liu AY, Mazin II, Kortus J (2001) Beyond Eliashberg superconductivity in MgB_2: anharmonicity, two-phonon scattering, and multiple gaps. Phys Rev Lett 87:087005

Liu H, Naumov II, Hoffmann R, Ashcroft NW, Hemley RJ (2017) Potential high-tc superconducting lanthanum and yttrium hydrides at high pressure. Proc Natl Acad Sci USA 114(27):6990–6995

Ludbrook BM, Levy G, Nigge P, Zonno M, Schneider M, Dvorak DJ, Veenstra CN, Zhdanovich S, Wong D, Dosanjh P, Straßer C, Stöhr A, Forti S, Ast CR, Starke U, Damascelli A (2015) Evidence for superconductivity in li-decorated monolayer graphene. Proc Nat Acad Sci 112(38):11795–11799

Lüders M, Marques MAL, Lathiotakis NN, Floris A, Profeta G, Fast L, Continenza A, Massidda S, Gross EKU (2005) Ab initio. Phys Rev B 72:024545

Lukas H, Fries SG, Sundman B (2007) Computational thermodynamics: the Calphad method. Cambridge University Press, Cambridge

Majumdar A, Tse JS, Wu M, Yao Y (2017) Superconductivity in FeH_5. Phys Rev B 96:201107

Mandal S, Cohen RE, Haule K (2014) Strong pressure-dependent electron-phonon coupling in FeSe. Phys Rev B 89:220502

Margine ER, Giustino F (2013) Anisotropic Migdal-Eliashberg theory using Wannier functions. Phys Rev B 87:024505

Marsiglio F, Carbotte JP (1986) Strong-coupling corrections to Bardeen-Cooper-Schrieffer ratios. Phys Rev B 33:6141–6146

Marques MAL, Lüders M, Lathiotakis NN, Profeta G, Floris A, Fast L, Continenza A, Gross EKU, Massidda S (2005) Ab initio. Phys Rev B 72:024546

Marzari N, Mostofi AA, Yates JR, Souza I, Vanderbilt D (2012) Maximally localized Wannier functions: theory and applications. Rev Mod Phys 84:1419–1475

Matthias BT, Geballe TH, Andres K, Corenzwit E, Hull GW, Maita JP (1968) Superconductivity and antiferromagnetism in boron-rich lattices. Science 159(3814):530–530

Mazin II, Johannes MD (2008) A key role for unusual spin dynamics in ferropnictides. Nat Phys 5:141 EP Article

Mazin II, Andersen OK, Jepsen O, Dolgov OV, Kortus J, Golubov AA, Kuz'menko AB, van der Marel D (2002) Superconductivity in MgB_2: clean or dirty? Phys Rev Lett 89:107002

Mazin II, Singh DJ, Johannes MD, Du MH (2008a) Unconventional superconductivity with a sign reversal in the order parameter of $LaFeAsO_{1-x}F_x$. Phys Rev Lett 101:057003

Mazin II, Johannes MD, Boeri L, Koepernik K, Singh DJ (2008b) Problems with reconciling density functional theory calculations with experiment in ferropnictides. Phys Rev B 78:085104

Meissner W, Ochsenfeld R (1933) Ein neuer effekt bei eintritt dernsupraleitfŁaehigkeit. Naturwissenschaften 21:787

McMahon JM, Ceperley DM (2011) High-temperature superconductivity in atomic metallic hydrogen. Phys Rev B 84:144515

McMahon JM, Morales MA, Pierleoni C, Ceperley DM (2012) The properties of hydrogen and helium under extreme conditions. Rev Mod Phys 84:1607–1653

Migdal A (1958) Migdal's theorem. Sov Phys JETP 34:996

Miyake T, Nakamura K, Arita R, Imada M (2010) Comparison of ab initio low-energy models for LaFePO, LaFeAsO, $BaFe_2As_2$, LiFeAs, FeSe, and FeTe: electron correlation and covalency. J Phys Soc Japan 79(4):044705

Monni M, Bernardini F, Sanna A, Profeta G, Massidda S (2017) Origin of the critical temperature discontinuity in superconducting sulfur under high pressure. Phys Rev B 95:064516

Morel P, Anderson PW (1962) Calculation of the superconducting state parameters with retarded electron-phonon interaction. Phys Rev 125:1263–1271

Nagamatsu J, Nakagawa N, Muranaka T, Zenitani Y, Akimitsu J (2001) Superconductivity at 39 K in magnesium diboride. Nature (London) 410:63

Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA (2004) Electric field effect in atomically thin carbon films. Science 306(5696):666–669

Oliveira LN, Gross EKU, Kohn W (1988) Density-functional theory for superconductors. Phys Rev Lett 60:2430–2433

Onnes HK (1913) The resistance of pure mercury at helium temperatures. Commun Phys Lab Univ Leiden 133:37

Ortenzi L, Mazin II, Blaha P, Boeri L (2012) Accounting for spin fluctuations beyond local spin density approximation in the density functional theory. Phys Rev B 86:064437

Pace EJ, Binns J, Alvarez MP, Dalladay-Simpson P, Gregoryanz E, Howie RT (2017) Synthesis and stability of hydrogen selenide compounds at high pressure. J Chem Phys 147(18):184303

Paglione J, Greene RL (2010) High-temperature superconductivity in iron-based materials. Nat Phys 6:645 EP Review article

Pépin C, Loubeyre P, Occelli F, Dumas P (2015) Synthesis of lithium polyhydrides above 130 GPA at 300 K. Proc Nat Acad Sci 112(25):7673–7676

Pépin CM, Geneste G, Dewaele A, Mezouar M, Loubeyre P (2017) Synthesis of FeH$_5$: a layered structure with atomic hydrogen slabs. Science 357(6349):382–385

Pietronero L, Straessler S, Grimaldi C (1995) Nonadiabatic superconductivity I. Vertex corrections for the electron-phonon interactions. Phys Rev B 52:10516

Platt C, Thomale R, Honerkamp C, Zhang SC, Hanke W (2012) Mechanism for a pairing state with time-reversal symmetry breaking in iron-based superconductors. Phys Rev B 85:180502

Profeta G, Franchini C, Lathiotakis NN, Floris A, Sanna A, Marques MAL, Lüders M, Massidda S, Gross EKU, Continenza A (2006) Superconductivity in lithium, potassium, and aluminum under extreme pressure: a first-principles study. Phys Rev Lett 96:047003

Profeta G, Calandra M, Mauri F (2012) Phonon-mediated superconductivity in graphene by lithium deposition. Nature Phys 8:131–134

Quan Y, Pickett WE (2016) Van hove singularities and spectral smearing in high-temperature superconducting H$_3$S. Phys Rev B 93:104526

Rosner H, Kitaigorodsky A, Pickett WE (2002) Prediction of high T_c superconductivity in hole-doped LiBC. Phys Rev Lett 88:127001

Runge E, Gross EKU (1984) Density-functional theory for time-dependent systems. Phys Rev Lett 52:997–1000

Sanna A (2017) Introduction to superconducting density functional theory. https://www.cond-mat.de/events/correl17/manuscripts/sanna.pdf

Sanna A, Profeta G, Floris A, Marini A, Gross EKU, Massidda S (2007) Anisotropic gap of superconducting CaC$_6$: a first-principles density functional calculation. Phys Rev B 75:020511

Sanna A, Flores-Livas JA, Davydov A, Profeta G, Dewhurst K, Sharma S, Gross EKU (2018) Ab initio Eliashberg theory: making genuine predictions of superconducting features. J Phys Soc Japan 87(4):041012

Sano W, Koretsune T, Tadano T, Akashi R, Arita R (2016) Effect of van Hove singularities on high-T_c superconductivity in H$_3$S. Phys Rev B 93:094525

Sato M, Ando Y (2017) Topological superconductors: a review. Rep Prog Phys 80(7):076501

Satta G, Profeta G, Bernardini F, Continenza A, Massidda S (2001) Electronic and structural properties of superconducting MgB$_2$, CaSi$_2$, and related compounds. Phys Rev B 64:104507

Savini G, Ferrari AC, Giustino F (2010) First-principles prediction of doped graphane as a high-temperature electron-phonon superconductor. Phys Rev Lett 105:037002

Savrasov SY, Savrasov DY (1996) Electron-phonon interactions and related physical properties of metals from linear-response theory. Phys Rev B 54:16487

Scalapino DJ (1969) The Electron-Phonon Interaction and Strong-Coupling Superconductors. In: Superconductivity (Ed. R.D. Parks) Marcel Dekker, New York. pp. 449–560.

Schilling A, Cantoni M, Guo JD, Ott HR (1993) Superconductivity above 130 K in the Hg-Ba-Ca-Cu-O system. Nature 363:56

Schickling T, Gebhard F, Bünemann J, Boeri L, Andersen OK, Weber W (2012) Gutzwiller theory of band magnetism in LaOFeAs. Phys Rev Lett 108:036406

Shamp A, Terpstra T, Bi T, Falls Z, Avery P, Zurek E (2016) Decomposition products of phosphine under pressure: Ph$_2$ stable and superconducting? J Am Chem Soc 138(6):1884–1892. PMID:26777416

Sharma S, Gross EKU, Sanna A, Dewhurst JK (2018) Source-free exchange-correlation magnetic fields in density functional theory. J Chem Theory Comput 14(3):1247–1253. PMID: 29420031

Skornyakov SL, Katanin AA, Anisimov VI (2011) Linear-temperature dependence of static magnetic susceptibility in LaFeAsO from dynamical mean-field theory. Phys Rev Lett 106: 047007

Sohier T, Calandra M, Mauri F (2015) Density-functional calculation of static screening in two-dimensional materials: the long-wavelength dielectric function of graphene. Phys Rev B 91:165428

Sohier T, Calandra M, Mauri F (2017) Density functional perturbation theory for gated two-dimensional heterostructures: theoretical developments and application to flexural phonons in graphene. Phys Rev B 96:075448

Struzhkin VV, Kim DY, Stavrou E, Muramatsu T, Mao Hk, Pickard CJ, Needs RJ, Prakapenka VB, Goncharov AF (2016) Synthesis of sodium polyhydrides at high pressures. Nat Commun 7:12267

Subedi A, Boeri L (2011) Vibrational spectrum and electron-phonon coupling of doped solid picene from first principles. Phys Rev B 84:020508

Subedi A, Zhang L, Singh DJ, Du MH (2008) Density functional study of FeS, FeSe, and FeTe: electronic structure, magnetism, phonons, and superconductivity. Phys Rev B 78:134514

Suhl H, Matthias BT, Walker LR (1959) Bardeen-Cooper-Schrieffer theory of superconductivity in the case of overlapping bands. Phys Rev Lett 3:552–554

Takano Y, Nagao M, Sakaguchi I, Tachiki M, Hatano T, Kobayashi K, Umezawa H, Kawarada H (2004) Superconductivity in diamond thin films well above liquid helium temperature. Appl Phys Lett 85(14):2851–2853

Thomale R, Platt C, Hu J, Honerkamp C, Bernevig BA (2009) Functional renormalization-group study of the doping dependence of pairing symmetry in the iron pnictide superconductors. Phys Rev B 80:180505

Toschi A, Arita R, Hansmann P, Sangiovanni G, Held K (2012) Quantum dynamical screening of the local magnetic moment in Fe-based superconductors. Phys Rev B 86:064411

Vignale G, Singwi KS (1985) Effective two-body interaction in Coulomb fermi liquids. Phys Rev B 32:2156–2166

Wang H, Tse JS, Tanaka K, Iitaka T, Ma Y (2012) Superconductive sodalite-like clathrate calcium hydride at high pressures. Proc Natl Acad Sci USA 109(17):6463–6466

Weller TE, Ellerby M, Saxena SS, Smith RP, Skipper NT (2005) Superconductivity in the intercalated graphite compounds C_6Yb and C_6Ca. Nature Phys 1:39

Werner P, Gull E, Troyer M, Millis AJ (2008) Spin freezing transition and non-fermi-liquid self-energy in a three-orbital model. Phys Rev Lett 101:166405

Woodley SM, Catlow R (2008) Crystal structure prediction from first principles. Nature Mat 7:937–946

Wu MK, Ashburn JR, Torng CJ, Hor PH, Meng RL, Gao L, Huang ZJ, Wang YQ, Chu CW (1987) Superconductivity at 93 K in a new mixed-phase Y-Ba-Cu-O compound system at ambient pressure. Phys Rev Lett 58:908–910

Yao Y, Klug DD, Sun J, Martoňák R (2009) Structural prediction and phase transformation mechanisms in calcium at high pressure. Phys Rev Lett 103:055503

Ye JT, Zhang YJ, Akashi R, Bahramy MS, Arita R, Iwasa Y (2012) Superconducting dome in a gate-tuned band insulator. Science 338(6111):1193–1196

Yi M, Zhang Y, Shen ZX, Lu D (2017) Role of the orbital degree of freedom in iron-based superconductors. npj Quantum Mater 2(1):57

Yin ZP, Haule K, Kotliar G (2011) Kinetic frustration and the nature of the magnetic and paramagnetic states in iron pnictides and iron chalcogenides. Nat Mater 10:932

Yin ZP, Kutepov A, Kotliar G (2013) Correlation-enhanced electron-phonon coupling: applications of GW and screened hybrid functional to bismuthates, chloronitrides, and other high-T_c superconductors. Phys Rev X 3:021011

Yokoya T, Nakamura T, Matsushita T, Muro T, Takano Y, Nagao M, Takenouchi T, Kawarada H, Oguchi T (2005) Origin of the metallic properties of heavily boron-doped superconducting diamond. Nature 438:647

Zhang L, Wang Y, Lv J, Ma Y (2017) Materials discovery at high pressures. Nat Rev Mat 2:17005 EP. Review article

Zhang T, Cheng P, Li WJ, Sun YJ, Wang G, Zhu XG, He K, Wang L, Ma X, Chen X, Wang Y, Liu Y, Lin HQ, Jia JF, Xue QK (2010) Superconductivity in one-atomic-layer metal films grown on Si(111). Nat Phys 6:104

Zurek E, Hoffmann R, Ashcroft N, Oganov AR, Lyakhov AO (2009) A little bit of lithium does a lot for hydrogen. Proc Natl Acad Sci USA 106(42):17640–17643

Titanium Alloys: From Properties Prediction to Performance Optimization

7

Dongsheng S. Xu, Hao Wang, Jinhu H. Zhang, Chunguang G. Bai, and Rui Yang

Contents

D. S. Xu (✉) · H. Wang · J. H. Zhang · C. G. Bai · R. Yang
Institute of Metal Research, Chinese Academy of Sciences, Shenyang, China
e-mail: dsxu@imr.ac.cn; haowang@imr.ac.cn; jinhuzhang@imr.ac.cn; cgbai@imr.ac.cn; ryang@imr.ac.cn

Abstract

Like many structural materials, titanium alloys have complex multiscale hierarchical structures that enabled a variety of mechanical properties. The elastic modulus, strength, ductility, fatigue, and creep behavior may depend strongly on the multiscale structures, from the bonding between atoms, to microstructures up to grain level, and their texture. However, their interdependences are usually not straightforward and sometimes not easy to reveal solely by experiments. Multiscale modeling and simulation in the last two decades have promoted the understanding of various aspects of titanium alloys. The progresses range from electronic structural calculation of alloying effects on modulus, strength, ductility, creep and oxidation, etc., to the ordering behavior in intermetallic phases and interaction between alloying elements and dislocation or twinning; atomic simulations of dislocation interaction and point defect formation, deformation and fracture mechanisms of polycrystalline titanium with grain boundaries of different nature under various stress; phase transformation and microstructure evolution under different conditions, and the effects of internal and external factors on the microstructure evolution and mechanical behavior of titanium alloys; finally FEM type simulations to optimize the thermo-mechanical processing, such as rolling and canned extrusion, etc. In spite of the increasing capabilities of multiscale materials modeling and simulation, many challenges still exist, and further seamless interdisciplinary collaborations among materials scientists, system designers, and industry production sections are needed to solve the grand challenge of the overall optimization of the titanium alloys for specific applications.

1   Introduction

Titanium alloys are widely used in aero-space and chemical industry mainly owing to their high strength, low density and good corrosion resistance, and applied recently as human implant, due to their good biological and mechanical compatibility with human bones. Pure titanium melts at 1670 °C, and has a solid-state phase transformation at 882 °C, above which it stabilizes as β phase with bcc structure, while below as α phase with hcp structure. Adding alloying elements can alter this transformation temperature, and cause other phases with different crystal structures to form. Very different microstructures can be obtained by changing alloy composition, heat treatment, and thermomechanical treatment schemes (Leyens and Peters 2003; Lütjering and Williams 2007).

Traditionally, titanium alloys are classified into three groups, namely, α, β and α + β alloys, depending on the main phases present in the alloys (Lütjering and Williams 2007); later, intermetallic alloys were introduced. Each type of alloys has their field of application according to their specific properties. For example, pure titanium and α alloys with minor alloying are usually used as corrosion resistant alloy, for chemical reaction chamber, pipe, valve and pump, etc., while β alloys, containing β stabilizing elements, are used in situation such as critical structural

components in aircraft where high strength and toughness are required or as bio-implant materials requiring low elastic modulus to improve the stress transfer to bone (Zhao et al. 2006), while some near α alloys can work continuously up to 600 °C and can be used in jet engines (Banerjee and Williams 2013).

Above 600 °C, conventional disordered alloys cannot work for long anymore, and alloys based on ordered intermetallic compound such as TiAl, Ti_3Al, or Ti_2AlNb may be used (Leyens and Peters 2003). By adding large amount of Al, the compound formed becomes even lighter than conventional titanium alloy, in addition to increasing their working temperature, and is very attractive for improving the efficiency of jet engine, especially the γ-TiAl-based alloy (density: ~4 g/cm^3), allowing for making components rotating at higher speed in jet engine, and even working up to 700 °C or higher (Appel et al. 2011). However, these ordered phases are usually brittle and difficult to fabricate. Various studies have been carried out to improve their different properties in the last 20 years (Kim 1998; Appel et al. 2011; Clemens and Mayer 2013; Yang 2015a).

Similar to steels and many other structural materials, the properties of titanium alloys depend strongly on their multiscale microstructures. These hierarchical structures mainly result from the rich phase transformations in this system with various stable and metastable phases (Fig. 1a) during different thermo-mechanical processing. Multielement additions may cause more phases or even amorphous phase to form, making available very large room for microstructure and properties adjustment, especially for components obtained through thermo-mechanical processing route (Banerjee and Williams 2013). At the same time, the fine details of the hierarchical microstructures formed by the multistep deformation/transformation
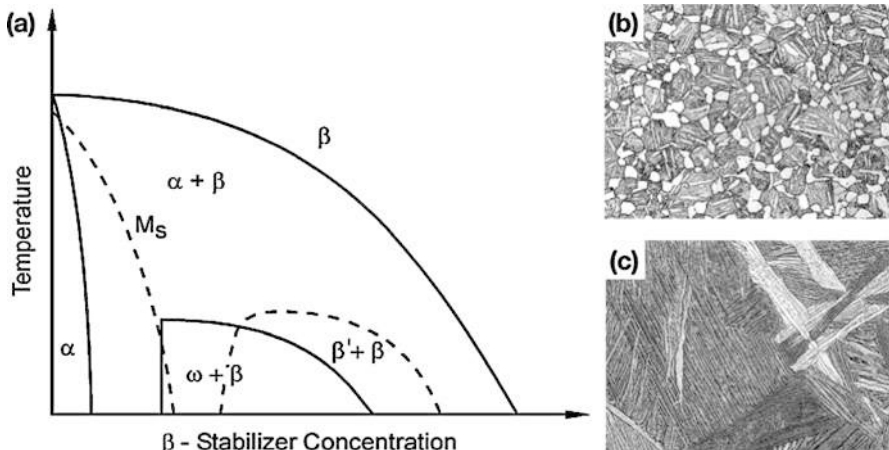
Fig. 1 (**a**) Schematic pseudo-binary phase diagram showing some metastable phases for alloys containing β stabilizers (Lütjering and Williams 2007). (**b**) and (**c**) Typical microstructure of titanium alloys: bimodal (TIMETAL 834) and lamellar (TIMETAL 1100), respectively, formed from different processing routes (Leyens and Peters 2003)

depend sensitively on the composition and hot working condition (Fig. 1b, c) (Leyens and Peters 2003) and are not easily characterized and related to the processing parameters in each step of high temperature deformation and heat treatment. Due to the lack of scientific understanding of each subprocess, a very good control over all the properties cannot be done rationally and can only be realized largely according to experience of engineers.

For ordered intermetallic compounds, it was shown in very early days that some high temperature properties are superior compared to conventional alloys, and the strength of certain compounds increases with temperature, hence suggesting higher service temperature. Some aluminides like Ni_3Al and TiAl have high strength and stiffness retention at high temperature and are lighter compared to their disordered counterpart, therefore ideal for high temperature rotating components in jet engines. For this reason, intensive investigations had been carried out; however, lack of room temperature ductility and low toughness, a general feature of intermetallics, made it difficult for system designers to accept them. After many years of work, the intermetallics community reached the conclusion in 1960 that this type of materials is not suitable for structural applications (Westbrook 1960). In 1980s, it was found that the poor room temperature ductility of Ni_3Al is not its intrinsic property, and grain boundary detachment due to hydrogen from water in moisture is the root cause of fracture. This discovery led to a period of concentrated research in intermetallic compounds. After several decades of effort, the TiAl intermetallic alloy has been used successfully in jet engine, e.g., on GEnx as low pressure turbine (LPT) blades (Kim and Kim 2018), each engine reducing ∼200 pounds. The successful insertion signified only the first step in overcoming many challenges facing the application of intermetallic alloys, and integrated study combining multiscale simulations and advanced characterization is needed for the rational design and optimization of new alloys. In the following, examples of the progress of computational studies will be given, and the prospective and challenges discussed.

2   Recent Approaches for Computational Design of Titanium-Based Alloys

Despite the great achievements in R&D of titanium alloys in the last 60 years, there is no straight forward way to design new alloys, and it takes 20 years or more to develop a new alloy for jet engine application. With the rapid progress of the computing technology in the last 30 years, multiscale modeling and simulation have been carried out extensively for many alloy systems. For alloy for specific application, identifying the bottleneck in each case and finding ways to overcome it utilizing multiscale simulation is an effective strategy to lower cost and accelerate development process.

In searching new alloys, first principles calculation can be a critical step in screening elements for property improvements, while for alloy optimization, analyzing limiting properties and process bottlenecks, identifying influencing factors, and utilizing multiscale simulations to reveal hidden mechanisms is crucial.

For conventional titanium alloys, how the multiscale microstructures arise during the long history of component fabrication and what their contribution is to each of the properties are not clear, while for intermetallic alloys such as TiAl, their poor deformability and long-term stability during high temperature service need further understanding. These problems need computational studies from electronic to macroscale, from fundamental defect properties, e.g., point defect, dislocation and twinning, to their contribution to comprehensive mechanical properties, e.g., strength, fatigue and creep. Phase stability, atomic ordering, point defect interaction, and various alloying effects are calculated on the electronic scale (Xu et al. 1997a; Song et al. 1999; Hu et al. 2001, 2003; Li et al. 2011), using simulation method mostly based on density functional theory, such as CASTEP (Clark et al. 2005), VASP (Hafner 2008) with plane wave basis, SIESTA (Soler et al. 2002), or DVM (Yang et al. 1993) using localized atomic basis. The recent high throughput computation following the emergence of ICME and MGI may greatly accelerate the materials simulation and analysis and promote alloying element selection (Curtarolo et al. 2012; Christodoulou 2013; Kubel 2013). The deformation mechanisms including dislocation and twinning properties have been studied with atomistic simulations (Wooding et al. 1995; Xu et al. 2008a; Girshick et al. 1998; Hu et al. 2001, 2004, 2014; Li et al. 2011; Ghazisaeidi and Trinkle 2012, 2014; Wang et al. 2014b; Clouet et al. 2015; He et al. 2018), e.g., molecular dynamics (MD). Phase transformation and microstructure evolution have been studied with mesoscale simulations, such as phase field (PF) and cellular automata (CA), often with help from atomistic simulation for input (Chen 2002, 2014; Trinkle et al. 2003; Hennig et al. 2005, 2008; Teng et al. 2012; Shi et al. 2014, 2015). Thermo-mechanical processing involving hot deformation and transformation has been studied with macroscale simulations (Manonukul and Dunne 2004; Dunne et al. 2007a, b, c, 2012; Dunne and Rugg 2008; Anahid and Ghosh 2013; Ghosh and Anahid 2013b; Ghosh and Chakraborty 2013a), e.g., finite element method (FEM) and crystal-plasticity finite element method (CPFEM). Thermo-dynamic phase diagram and kinetic calculation also play important roles in different stages of development, using ThermoCalc, Pandat, etc. It is worth noting that owing to the complex microstructure of titanium-based alloys, their behaviors usually depend on subprocesses on several different scales; hence, these results should always be incorporated with results on other scales for correct understanding. Nevertheless, it has been observed that multiscale modeling plays increasingly important roles in the development of titanium-based alloys, and these will be exemplified in the following (Hennig et al. 2005; Raabe et al. 2007; Ghosh and Anahid 2013b; Clouet et al. 2015; Yu et al. 2015; Appel et al. 2016).

3   First Principles Prediction of Alloying Effects for Composition Design

Most properties of materials have their electronic origin, some with direct relevance, e.g., stability and elastic modulus of phases are related to energy and its second

derivative, respectively, so that they are predictable by electronic structure calculation, while others may depend on structures at multiple scales, such as strength, toughness, ductility, and fatigue resistance, with indirect link via some elementary process, such as dislocation, twinning, or grain boundary sliding. In such case, multiscale simulations are needed for thorough understanding.

3.1   Electronic Structure Calculation for Alloying Selection

Before large-scale calculation of electronic structure was available, predicting the formation of phases was difficult for multielement alloys. Certain experience-based model was developed, e.g., Mo- or Al-equivalent was calculated according to the concentration of alloying additions and the resulting change rate of the phase transformation temperature, and criteria were established for phase stability estimation. However, sometimes the summation rule does not work well for multicomponent alloys. One early example considering electronic factor was to control strictly the precipitation of Ti_3X phase in titanium alloys (to achieve balance between long-term microstructure stability and high temperature strength) using the electron concentration criterion but with experimentally determined valance electron number (Li et al. 1984).

For titanium alloys working at high temperature as rotating components in jet engines, once the balance of microstructure stability and hot strength has been achieved, high temperature creep during long-term service may arise as a major problem. It was recognized that high temperature diffusion may be the main cause of creep under relatively low stress, and the interaction between vacancies and alloying atoms is related to diffusion. However, this interaction is not measurable experimentally. First-principles calculations have been carried out to investigate the interactions between vacancies and alloying elements (Hu et al. 2001) (Fig. 2). Analysis shows alloying elements attractive to vacancy may reduce the diffusion rate, while those repulsive or noninteractive with vacancies, such as Fe, Co, or Ni, diffuse much faster and need to be restricted. The calculated interaction energy variation corresponds well with diffusion rate and creep measurements. This can be used as a guideline for alloying element selection for creep resistance design. These efforts led to the successful design of Ti55, a titanium alloy for applications upto 550 °C (Guan et al. 1998), and Ti60 at 600 °C and beyond (Wang et al. 2014d).

Oxidation may become a major concern as temperature increases further. Since Ti is very reactive, once a TiO_2 layer is formed, it is very stable at low temperature either in air or in water, even seawater. At high temperature however, the TiO_2 layer becomes no longer protective, a layer of Al_2O_3 may provide better protection to the titanium alloy underneath. The oxidation resistance of these oxides depends on both defect structure in the oxidation layer and the bonding of the layer with matrix, and a detailed understanding needs electronic calculation of the bonding and defect properties (Wang et al. 2015; ▶ Chap. 4, "Titania and Its Outstanding Properties: Insights from First Principles Calculations").
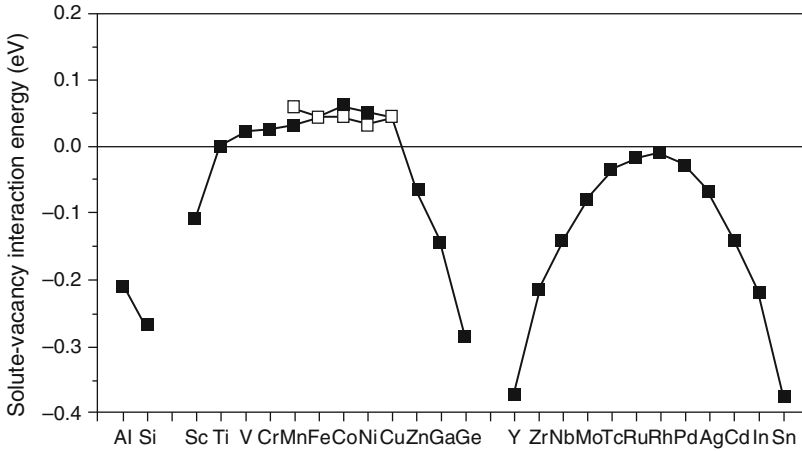
Fig. 2 Interaction energies between solutes and vacancy in α-Ti, where positive interaction means repulsive to vacancy, and negative attractive (Hu et al. 2001)
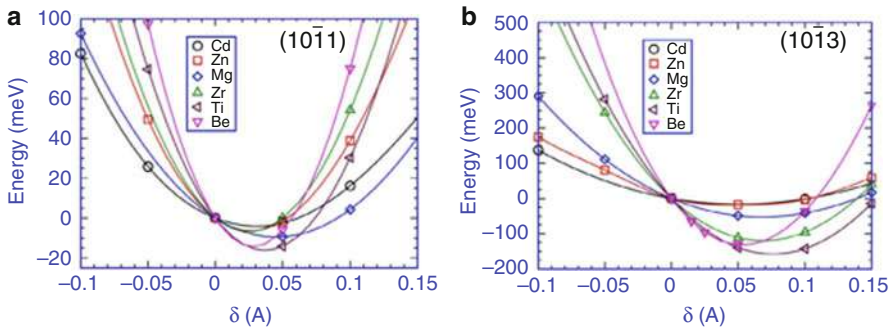
Fig. 3 Variation in excess energy of the simulation cell as a function of δ for six metals: (**a**) (10$\bar{1}$1) twin boundary; (**b**) (10$\bar{1}$3) twin boundary (Kumar et al. 2015)

Early exploration of electronic scale calculation has been utilized in other titanium alloys, such as for anticorrosion design with calculated bond order and energy level of d electrons for alloying elements selection (Morishita et al. 1991). First-principles calculation has also been carried out to predict the alloying effects on elastic modulus, which has promoted the design of low modulus titanium alloys for bio-implants and other applications (Song et al. 1999; Hao et al. 2007). Twin-associated boundaries were systematically investigated in a series of hexagonal metals including Ti, with energy and solubility calculated for different types of interfaces under different stress (Fig. 3) (Kumar et al. 2015). Recent high-throughput computation carried out to investigate the alloying effects on the energy trend of stacking fault of various types in Ti system with a view of understanding the strengthening effect in some engineering alloys (Wang et al. 2018). Apart from

the major elements, Al, V, Nb, Zr, Mo, Sn, etc., which are mostly substitutional in lattice, some interstitial elements, e.g., O, C, N or B, are either alloyed or exist as impurities. It was found that despite their limited concentration, some impurities may have stronger effect than the major elements, such as higher hardening rate or weakening effect, but the reason remains unclear. The strong hardening effects have been clarified recently by combination of advanced characterization and first-principles calculation of the interaction between interstitial atoms O and dislocations (Yu et al. 2015). It can be expected that deeper understanding will be achieved via close collaboration between first-principles calculation and experiments, for other elements as well.

3.2   The Ordering Behavior in Intermetallic Alloys

Above 600 °C, conventional disordered titanium alloys are not capable of long time service, and ordered intermetallic phase such as α_2 and γ phases can be introduced either as strengthening phase or as matrix to increase the working temperature further to 650 °C or even 900 °C. The ordering status in intermetallics has important effects on many properties like electric, magnetic, and mechanical properties via different mechanisms, such as affecting dislocation dissociation or twinning tendency. For the temperature dependence of ordering, a self-consistent scheme based on first-principles calculation and statistical mechanics was proposed to evaluate the point defect concentration and order-disorder transformation temperature in binary intermetallics (Hu et al. 2004). The dependence of defect concentration and ordering status of multielement compounds needs both new model and accurate defect energy.

Although intermetallic compounds have many advantages at high temperature, most of them are very brittle, partly due to the partially covalent bonding between atoms. Various alloying elements were added in order to improve the room temperature ductility and other properties, but a detailed understanding relies on electronic structure calculations that require information of atomic position. The site occupation of alloying elements in ordered intermetallic compounds TiNi, TiAl, and Ti_3Al, etc., was studied on the basis of the bond order or binding energy of alloying elements with matrix (Xu et al. 1997a, b). A new method was proposed to predict the site occupation behavior of alloying elements in ordered phases, where two straight lines were used in the diagram to separate the alloying behavior into three regions (Fig. 4). For elements lying between the two lines, the site occupation is composition dependent and therefore may be adjusted. This dependence has been observed experimentally by ALCHEMI measurement (Tadaki et al. 1995; Hao et al. 1999).

Due to different ordering tendencies, the elements may have "site competition" between them, and the lattice may have a "stepped ordering" behavior, as more alloying elements were introduced (Xu et al. 1997a, b). These new findings can be utilized in fine-tuning the mechanical or other properties of intermetallic alloys.
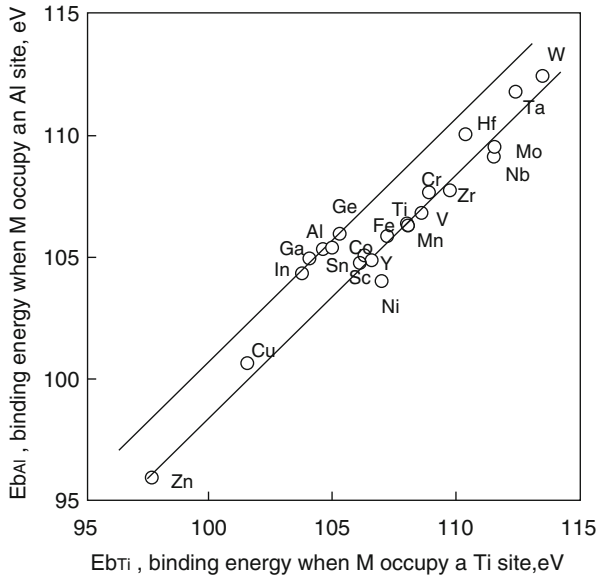
Fig. 4 Site occupation diagram in Ti$_3$Al. The elements above and below the lines occupy Al and Ti sites, respectively; the occupancy of those in between the lines depends on alloy composition (Xu et al. 1997a)

4   Atomistic Simulation of Deformation Mechanisms

Atomic-scale simulation has become a powerful tool to study atomic processes between electronic and mesoscale simulations. In metals and alloys, point defect, dislocation, twinning, grain and phase boundary behavior are investigated with atomic-scale simulations in order to understand corresponding mechanical properties. In titanium alloys, despite their strong anisotropy and complex microstructure, atomic-scale simulations have been extensively used in modeling various defects, to reveal the detailed deformation and fracture mechanisms under different conditions. For titanium alloys in jet engines, fatigue resistance is critical and is the most difficult to understand, especially for dwell fatigue, which often results in large scattering of fatigue life, while for TiAl alloys, its brittleness and long-term microstructure stability are both important. How defects accumulate and fatigue damage proceeds in materials finally causing crack and fracture is a challenging problem, which could not be solved solely by experimental observation, partly because crack nucleation is often subsurface, and the crack embryo is very small. Various deformation mechanisms, formation and interaction among defects, and transformation mechanisms had been revealed with atomic resolution.

4.1   Dislocation Structure and Dipole Transformation in Titanium and TiAl

Dislocations are the most important plastic deformation carrier in metals and alloys. To model single dislocations, the key is the reproduction of correct core structures. This requires a good description of the interactions between atoms, i.e., interatomic potential, and a proper simulation box either large enough or with carefully designed size and shape to minimize the influence of image dislocations (Cai et al. 2000). For instance, with modified embedded-atom method (EAM) (Ghazisaeidi and Trinkle 2012) and mean-field EAM (Liu et al. 2014) interatomic potentials, respectively, the core structures of screw dislocations in α-titanium and of edge dislocations in β-titanium are studied. Qualitative agreement is achieved with ab initio calculations on the core configurations (Fig. 5a), whereas detailed atomic structures depend on potential, e.g., the dislocation dissociates on the basal plane according to the stacking fault energy predicted by the potential. With ab initio calculations, dislocation cores in Ti and Zr were modeled and different glide behaviors were observed (Clouet et al. 2015), which explains the locking and easy glide in Ti and Zr, respectively (Fig. 5b).

There has been extensive work on modeling dislocations in titanium aluminides (Simmons et al. 1997; Panova and Farkas 1998; Porizek et al. 2003; Woodward and Rao 2004; Katzarov et al. 2007; Xu et al. 2014). The bond-order potential was employed to study the glide behavior of 1/2[<110] ordinary dislocations. The MD simulation using an EAM potential found the asymmetric core configurations and motion of <011] superdislocations, and the core structure depends sensitively on dislocation characteristics (Xu et al. 2014). Relevant work on multicomponent titanium alloys has been rare. In particular, a 4-element potential was constructed using mean field concept (Liu et al. 2014) and was able to obtain dislocation cores in a β-type titanium alloy, Ti2448 (Hao et al. 2007).

Dislocation interaction of various types such as dipolar or forest interaction is crucial for the understanding of strength and hardening of metals, while dislocation dipole annihilation is the most important processes during fatigue. It was found 40 years ago that dislocation density saturates as fatigue proceeds, leading people to consider that the dipolar dislocation within a certain height disappeared completely after annihilation (Essmann and Mughrabi 1979). Simulations show that dipoles can form through the meeting of parallel dislocations with opposite Burgers vectors or the motion of jogged screw dislocations under single or multiple slip, or even form homogeneously if the lattice was sheared strongly (Veyssiere et al. 2009; Xu et al. 2009). It was also shown that dipoles may weaken the lattice, and the lattice strength decreases as the density of dipoles increases, therefore causing the lattice to crack easily (Gu et al. 2010). The behavior of nonscrew dislocation dipoles in a series of metallic systems has been systematically investigated (Wang et al. 2008a, 2009c, 2013a, 2014b). In titanium (Wang et al. 2014b) and TiAl (Wang et al. 2013a), narrow dipoles were found to transform into individual defects, producing debris depending on dipole height, orientation, and temperature. Reconstructed configurations are formed at low temperature, while vacancy clusters, stacking fault tetrahedra, and
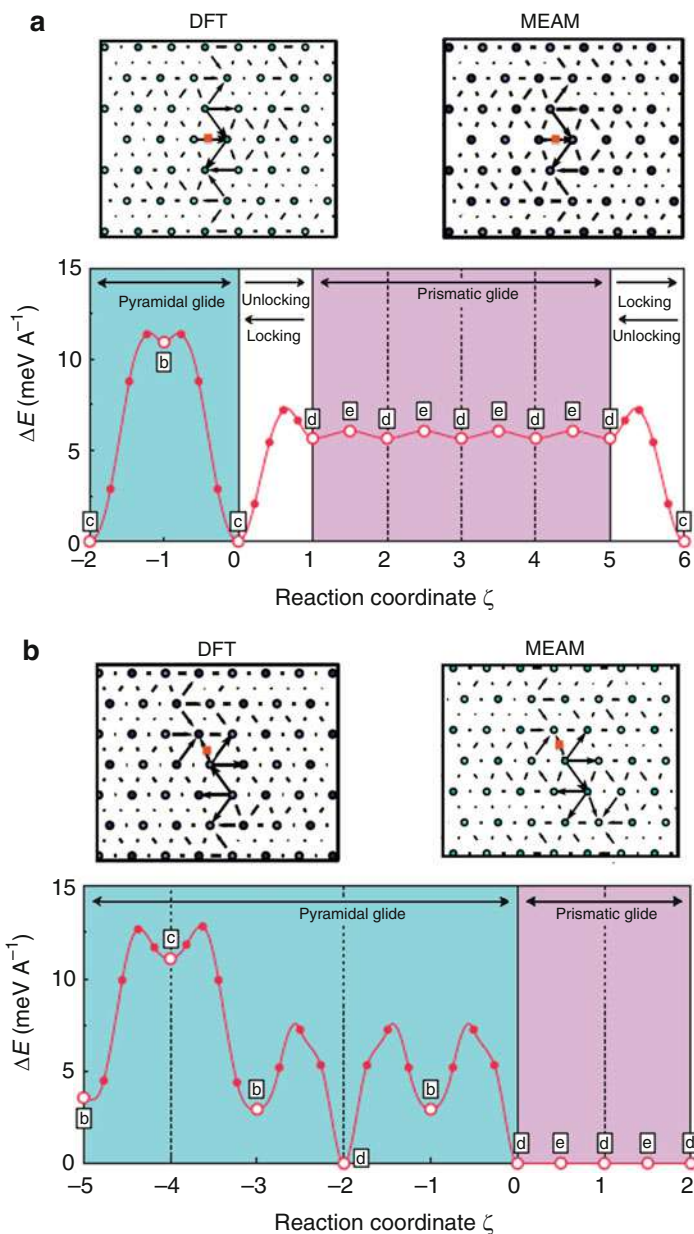
Fig. 5 (**a**) Typical core structures of 1/3[1$\underline{2}$10] screw dislocation in titanium (Ghazisaeidi and Trinkle 2012). (**b**) Different glide energy barriers in Ti (left) and Zr (right) that relate to locking and easy glide, respectively (Clouet et al. 2015)
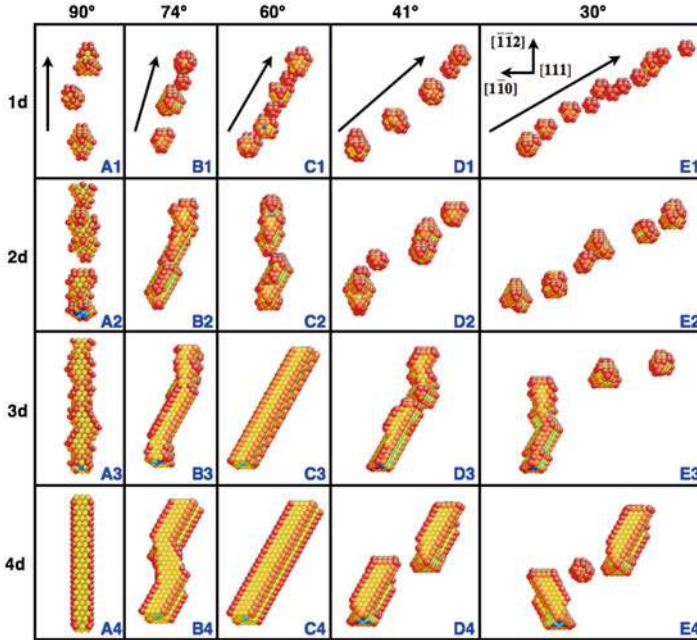
Fig. 6 [111] projection of vacancy dipoles after 1700 K-aging for 1 ns in TiAl, with orientation from 30° to 90° and heights of 1*d* to 4*d*. Initial orientations are indicated with arrows in the first line. (Wang et al. 2013a)

vacancy or interstitial type loops may form at high temperature (Fig. 6). An SFT formation mechanism from deformation was proposed on the basis of atomistic simulation and later observed experimentally (Wang et al. 2011c, b, d, 2013c). Employing saddle-point search methods, activation energies of the atomic processes therein are obtained and the lifetime of the above by-products is estimated, showing the extreme stability of certain clusters and loops on the experimental timescale (He et al. 2018). Such point defects may interact with dislocations hindering their motion and form a basis for work-hardening and crack nucleation during monotonic deformation and fatigue. For defects on the surface, they may have a stronger impact on the deformation and fracture compared with internal ones of similar size. Depending on the defect shape, location, and orientation, they may serve as the nucleation site for dislocation, twinning, or crack (Wu et al. 2016).

4.2   Deformation Twinning in Pure Ti and TiAl

Due to the limited slip systems in α-Ti and TiAl, twinning becomes important to accommodate plastic deformation. Despite the extensive investigations on twinning in Mg and Zr, there have been relatively few modeling investigations on Ti on the
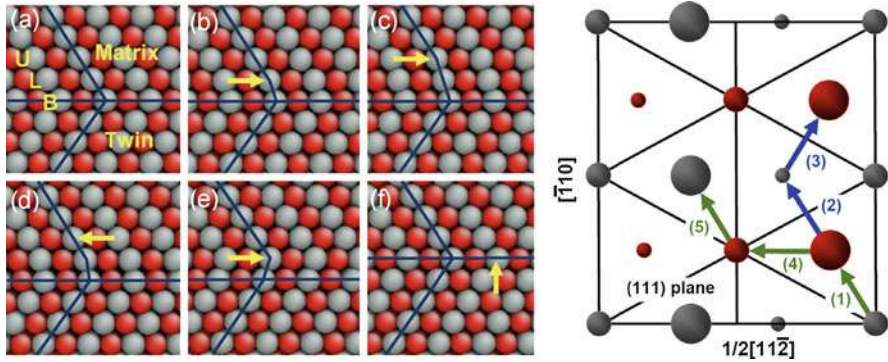
Fig. 7 A new twin growth model in TiAl through a five-step process involving the synchro-shear of two atomic layers by a pesudotwinning shear (Xu et al. 2008a)

atomic scale (Serra and Bacon 1986; Oberson et al. 2011; Wang et al. 2011f, 2013b, 2014c; Li et al. 2012; Ren et al. 2014; Zong et al. 2014). High throughput MD simulations have been carried out for different type of interfaces under various loads, in search for possible new mechanism to increase the deformability of these systems. For TiAl, shear deformation under different pressure was simulated. It was found the pseudo-twinning shear deformation mechanisms are sensitive to the hydrostatic pressure under which shear is imposed. Under hydrostatic tension, true twin can still proceeds in TiAl with $L1_0$ structure even when sheared along pseudo-twining directions, by a new mechanism involving the movement of 5 partials on two nearby slip plane (Fig. 7), while under compression, the same shear produces a pseudo-twin with an $L1_1$ structure (Xu et al. 2008a). First principles calculations show that some elements such as Nb, Ta may promote the $L1_1$ pseudo-twinning during deformation and therefore may shed some light on increasing the ductility of TiAl alloys (Xu et al. 2008b).

It should be noted that the pressure effect is very difficult to examine experimentally, especially during dynamic loading, while MD simulation is especially good for that. In fact, historically it is believed that twin is mainly formed under pressure, because it is often observed in the explosive impacted samples. Our MD simulation in Mo shows that tension prefers twin formation, and the impacted sample may form twin during the tension period, as evidenced by the fact that twins are often observed around the hole formed in laser impacted sample, and more twins were observed often in tension samples at the position that breaks the last (Xu et al. 2004).

4.3 Grain Boundary-Mediated Deformation in Ti

Critical components in aircraft or jet engine are mostly wrought processed with complex microstructure, where dislocation-grain boundary (GB) interaction is crucial in determining their mechanical properties, especially in the case of fatigue
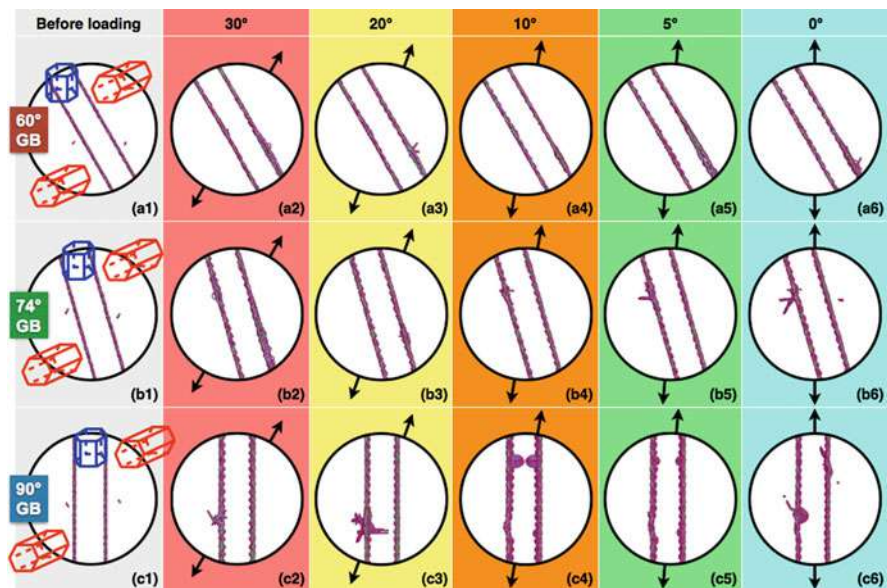
Fig. 8 Yielding behavior for different GB-dislocation interactions under loading in various directions (Wang et al. 2016a)

(Leyens and Peters 2003). Atomic scale information of dislocation-GB interaction not only provides penetrating understanding of plastic behaviors but also helps to establish constitutive models, thus contributing to the elucidation of relevant property dependence. Especially in a series of α-titanium alloys, previous experiments and phenomenological simulations indicated that dislocation-GB interaction has potential, however, unclarified, importance during fatigue (Dunne et al. 2012).

Employing a bicrystal model with soft-hard orientation, the behavior of different grain boundary (GB) interaction with dislocations is systematically investigated (Wang et al. 2012, 2016a). Depending on the GB orientation and the loading direction, yielding initiates through grain boundary cracking, dislocation emission, or grain protrusion (Fig. 8). The residual dislocation weakens the GB and the critical stress is found sensitive to the loading direction. The largest orientation sensitivity occurs when GB plane is parallel to the *c*-axes of both grains. A new plastic deformation mode is identified, which produces grain protrusion from the soft to the hard grain through lattice reorientation, and provides relatively large tensile strain of 8.3% with activation energy of 0.04 eV/atom. Lattice analysis shows that the protruded part has a special twin relation with the matrix grain. High throughput computations are carried out to search for the alloying element, which may benefit this new twinning mechanism (Zhou et al. 2018).

The atomic mechanism revealed through the atomic simulation may shed some light on improving the properties of the titanium-based alloys.

5 Mesoscale Simulation of Microstructure Evolution Under Various Conditions

The mechanical properties of titanium alloys such as strength, stiffness, fatigue and damage tolerance depend largely on their microstructures. Various microstructures can be obtained from heat treatment or thermo-mechanical processing. Accurate prediction and control of microstructure are of great importance for achieving outstanding mechanical properties (Banerjee and Williams 2013). However, the relations among the microstructure, processing, and performance are not clear, even the description of the multiscale hierarchical structure is not simple. On the simulation, alongside the rapid development of computational power in the last 20 years, many mesoscale methods have been developed, such as CA (Ding and Guo 2002; Pourian et al. 2015), Monte Carlo (MC) (Zhang et al. 2012; Tan et al. 2017), phase field (PF) (Chen 2002; Steinbach 2013), CPFEM (Raabe and Roters 2004; Roters et al. 2010) and level set (LS), smoothing modeling (Park et al. 2014) (Vermaak et al. 2014; Xu et al. 2015), and have been successfully utilized to simulate microstructural evolution under different conditions. These models and simulations have been able to characterize microstructure, reveal mechanism of complex microstructural evolution during thermo-mechanical processing, and have promoted microstructure design for certain application and accelerated the optimization of titanium alloys, some of which are exemplified below.

5.1 Microstructure Evolution During Heat-Treatment

Rich microstructure can be obtained in titanium alloys by solid-state transformation during heat-treatment. In order to understand how various microstructures form, the influence of several factors such as interfacial energy, elastic strain energy, chemical driving forces, and the interplay among these factors are analyzed based on phase field and other simulations.

The β to α transformation is the main transformation in titanium alloys most microstructures rely on; however, because the lattice change and diffusion occur jointly, it is not clear how the transformations happen under different conditions. Systematic simulations were carried out in the last decade for understanding and design such transformations. The nucleation of precipitates was found to depend on interfacial energy, elastic strain energy, and chemical driving force. Elastic interaction determined by the lattice parameters of the two phases was identified as the major controlling factor of forming the lath shape during β to α transformation. Figure 9 shows the microstructure formed and phase evolution during the martensitic transformation in Ti alloys, calculated with the newly developed generic phase-field model for displacive-type transformations in polycrystals under different condition (Heo and Chen 2014a).

The microstructure of titanium alloys depends not only on the fraction of the two main phases, α and β, but also on the transformation path. α may form directly
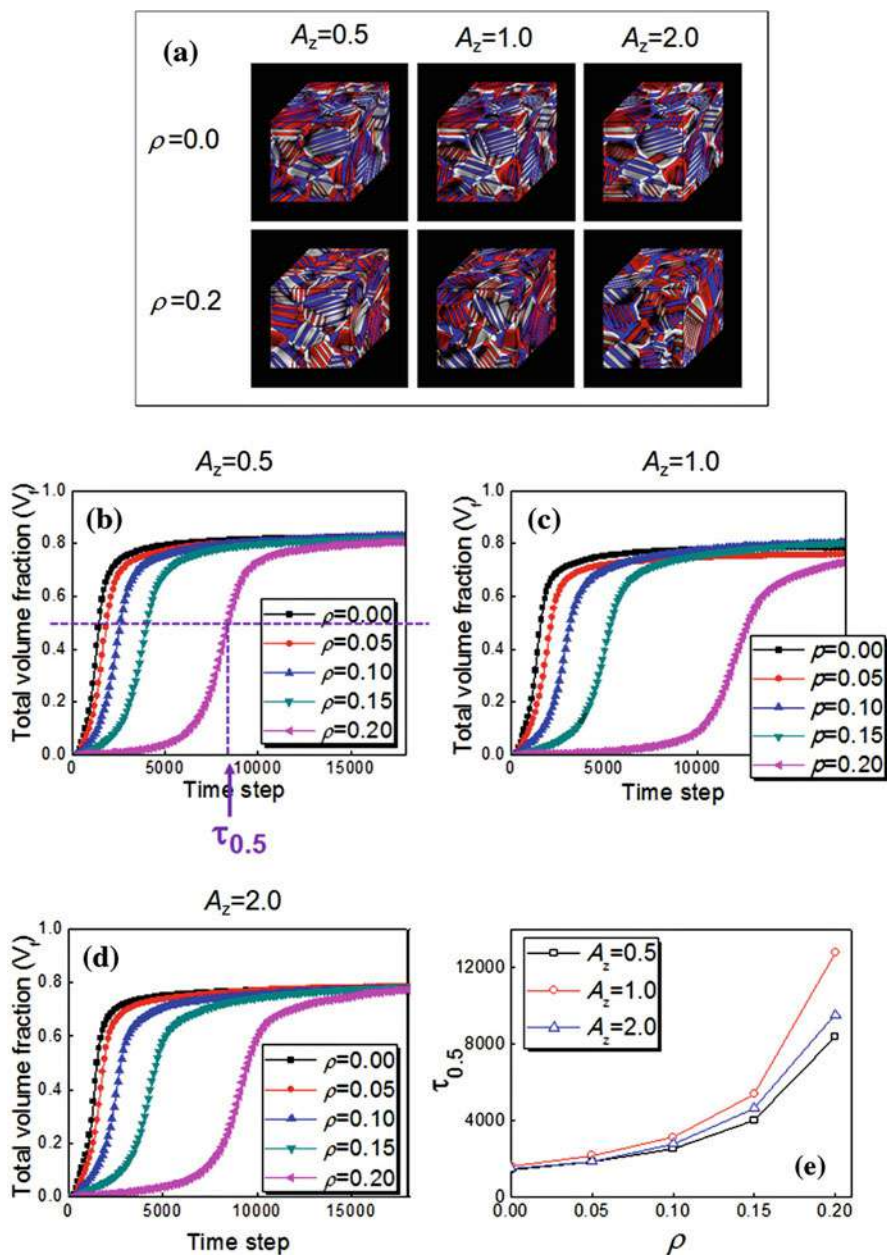
Fig. 9 (**a**) Microstructures of Ti alloy for two misfit strain relaxation with different elastic anisotropy. (**b**–**e**) Evolution of β fraction with different q and A_z (Heo and Chen 2014a)

from β phase, or from defect mediated nucleation, or by some intermediate phases, such as α'' martensite and metastable ω phase, or from concentration modulation by prior-ω phase, or even spinodal decomposition. The detailed path chosen and microstructure formation process can only be observed through simulation considering these factors (Ni and Khachaturyan 2009; Nag et al. 2012; Shi et al. 2016) (Heo et al. 2014b).

Thermodynamic data are crucial for microstructure simulation and phase equilibrium estimation. If we could consider kinetics further, unconventional kinetic pathways can be predicted, based on systematic analysis of phase transformation in multicomponent alloy, such as pseudospinodal decomposition involving both displacive transformation and long-range diffusion. Such pathways may lead to different microstructure under different treatment (Ji et al. 2016). Transformation strain not only affects the microstructure and mechanical properties but also has a strong impact on many functional properties such as electric conductivity, magnetic moment, and even multiferrous behavior coupling electro-magnetic and mechanical strain, with special strain stabilized domain structure and certain properties, which may results in components with different functions (Hu et al. 2013).

Another important alloy type of interest is the γ-TiAl-based ordered intermetallics (recently named gammalloy). Extensive studies were carried out to control their microstructure and improve their ductility for aero-space applications (Kim and Kim 2018). Different commercial alloys were developed for various components including turbine blade for jet engine and exhaust valve for car engine. The lamellar structure has good strength and toughness at ambient temperature, but the long-term stability of the lamellar structure during service is a concern for components such as turbine blades. The high density of interface in lamellar structure increased the energy of the structure. During high-temperature service, it may transform to bulk structure with lower energy, and then the properties from the lamellar structure will be degraded and the blade may become extremely brittle and weak. Recently, it is recognized that lamellae with high density of interfaces with lower energy may be a good choice. Twin boundary is such an interface with much lower energy compared with others; therefore, lamellar structure containing high fraction of twin boundary will benefit long-term stability. In this context, the effects of several factors on twin boundary fraction in lamellar TiAl were studied by phase-field simulations (Teng et al. 2012). The influences of coherency stress, interfacial energy, and undercooling on the nucleation and colony growth were analyzed. Twin boundary fraction was found to increase with increasing coherency strain energy and increasing energy difference among different types of interfaces but decrease with the increase of driving force for phase transformation (undercooling) (Fig. 10). Depending on the relative contributions of elastic strain energy, simulation found that gamma phase nucleation events could be collective, correlated, or independent. Further simulations show that external stress can also affect the twin boundary fraction strongly and could be utilized for microstructure control (Teng et al. 2015). It is expected that by detailed adjusting the above parameters, and controlling the nucleation and growth carefully, relatively stable microstructure can be achieved.
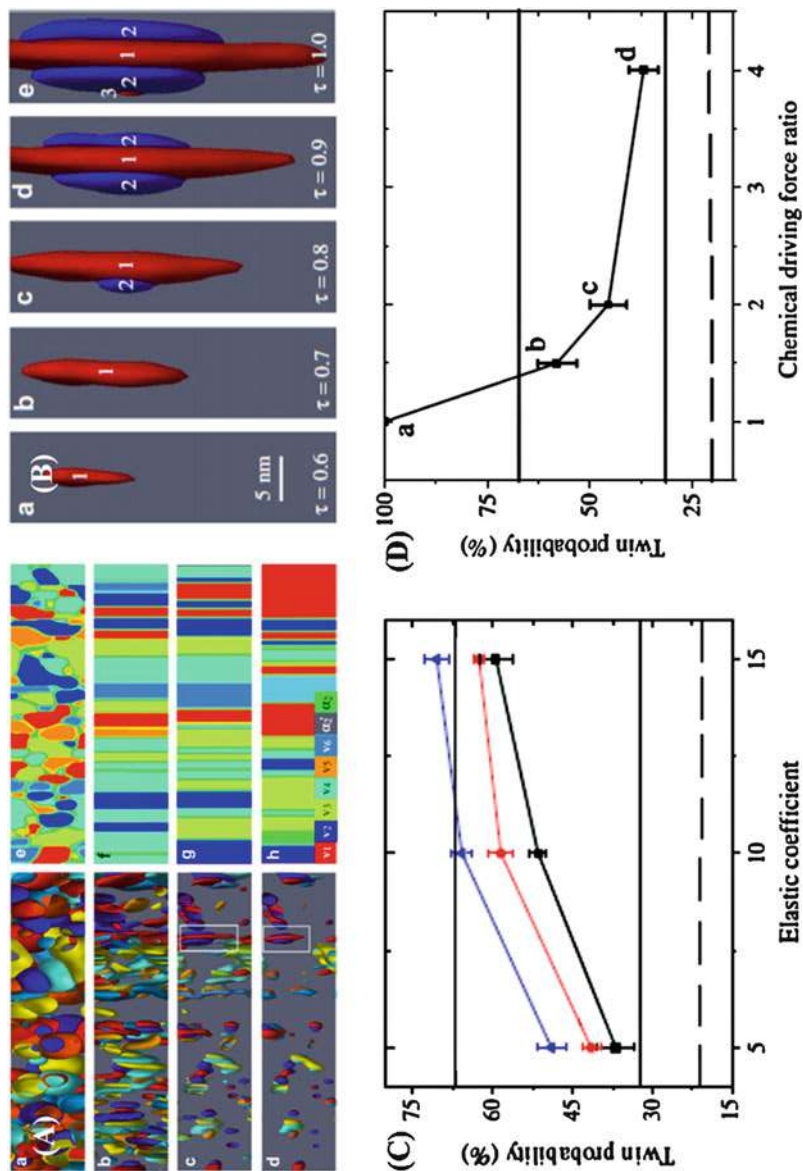
Fig. 10 (**a**) The effect of elastic strain energy contributions on the microstructure at the nucleation and final stages. (**b**) Correlated nucleation at medium elastic energy. (**c**) Twin boundary probabilities as a function of the elastic energy amplitude. (**d**) Twin boundary probabilities as a function of chemical driving force ratio (Teng et al. 2012)

The idea above for heat-treatment route design may also be applied to other titanium alloys. In order to understand how α plate is formed, phase field model considering inhomogeneous and anisotropic elasticity was developed to study the growth of acicular precipitates observed in Widmanstätten structures under isothermal conditions (Cottura et al. 2014). When the plate-like precipitate grow large, the elastic energy anisotropy plays a key role even during diffusion-controlled process. When only one soft direction exists, stationary growth is obtained after a transient stage, while if several soft directions are in competition, sharp concomitant morphological and kinetic transitions are observed. To understand what controls texture formation, the morphology and variant selection (VS) of grain boundary allotriomorphs and Widmanstätten sideplates are investigated using 3D phase field model (Qiu et al. 2016). Quantitative analysis indicates that precipitate morphology and VS are determined by the interplay among elastic interaction between nucleating precipitates and GB dislocation networks, while growth anisotropy is determined by the relative inclination of the habit plane with respect to GB dislocations, density of nucleation sites and coalescence during growth, and spatial confinement from simultaneously nucleated neighboring variants of dissimilar types.

The recently designed ultralow modulus Ti2448 alloy has very unique mechanical properties (Hao et al. 2012): when quenched to 193 K, martensitic transformation (MT) may become a major deformation mechanisms (Zhu et al. 2017). Both crystallographic analysis and phase field simulations have revealed the diverse strain accommodating modes, especially the excellent geometrical compatibilities of twins and herringbone structures, providing the MT with great flexibility for arbitrary shape changes and defect structures generated during plastic deformation, which may have contributed to the exceptional mechanical properties of the alloy. Phase field simulations have also been used to study the effects of many other factors on the microstructure, such as sideplate formation from grain boundary in Ti-6Al-4 V alloy (Wang et al. 2009b), effect of interfacial energy anisotropy (Yang et al. 2012), and the effect of initial particle size distribution (Wang et al. 2009a). Apart from growth of precipitates, the dissolution and coarsening process were also simulated in order to gain understanding of the kinetics of these processes. These may help with the heat treatment route design (Wang et al. 2009a, 2011a).

5.2 Microstructure Evolution During Thermo-Mechanical Processing

Titanium ingots are usually broken down and deformed into various shapes through different thermo-mechanical processes, such as forging, rolling, or extrusion, into required rod, sheet, or plate. Further forming can be accomplished by isothermal forging or friction welding. These processing operations play a significant role to produce microstructure with outstanding properties. However, the whole process is long and complicated, with each step having influence on later ones, the link between microstructure and processing being not clear. With the fast progress of
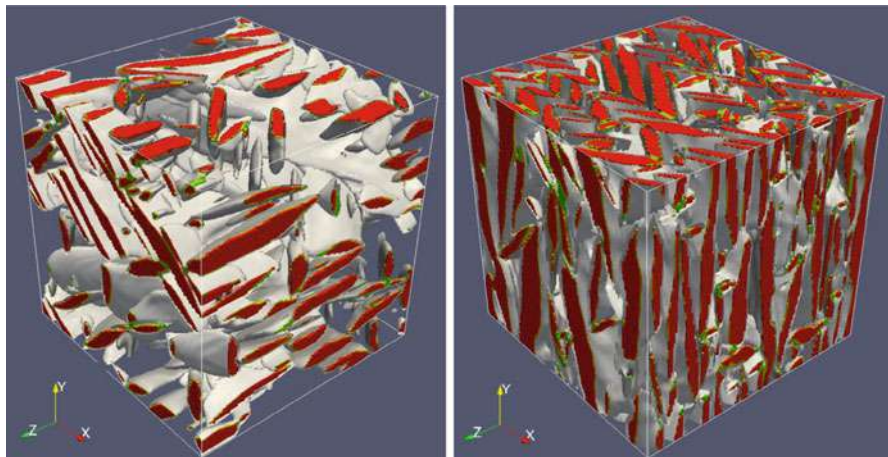
Fig. 11 Microstructure with VS caused by prestrain of uniaxial tension or compression (50 MPa) along [101]$_\beta$ (Shi and Wang 2013)

microstructure simulation method recently, the difficulties could be partly overcome by computer simulation.

Simulations considering the influence of certain factors separately is feasible and may be superior compared to experiment. Phase field model was developed to predict VS and microstructural evolution of α precipitates during β to α transformation (Fig. 11). It is found that the microtexture with VS is determined by the interplay between external stress and internal stress from precipitation (Shi and Wang 2013). Further simulation in polycrystalline Ti-6Al-4 V (Shi et al. 2015) found that local stresses lead to obvious VS and, under certain prestrains, polycrystalline β sample could end up with a relatively weak α texture even if it has a strong starting β phase texture. Simulation also shows that due to the strong anisotropy of α precipitates, the 2D sampling may introduce "pseudo VS," thus when analyzing texture data, one must bear this in mind. The findings may shed some light on the control of transformation texture by thermomechanical processing.

Besides PF, CA is also used for modeling microstructure evolution, e.g., to simulate the micromechanical behavior in a Ti alloy under dwell-fatigue loading (Pourian et al. 2016). Simulations show clearly the load transfer from a deformed cell to its neighbors, during cyclic loading, with hard cells experiencing higher stress until they start to deform plastically and eventually generate a crack. Although CA model is simple, it could describe high level of complexity and allow the statistical description of polycrystals by considering stress-strain transfers between neighbors. Another simplified CA model is developed to simulate the 3D local behavior of nontextured microstructures in elasticity with reasonable results compared to more sophisticated FEM (Pourian et al. 2014).

The CPFEM is an integrated theoretical framework containing micromechanical models and concepts under a unified theory of crystal plasticity. It is a powerful

tool for solving mechanical events during hot working. A microstructure-sensitive fatigue study in Ti-6Al-4V and Ti-18 alloy by CPFEM (Smith et al. 2016) suggests that reduced lamellar colony sizes, α phase, and transverse texture can enhance the resistance to formation and early growth of fatigue cracks. The CPFE investigation of the micromechanical behavior of Ti-6Al-4V shows that the nanoindentation results depend sensitively on grain orientations (Han et al. 2015). The CPFE simulation of the deformation of titanium sheet under different strain shows that twinning affected significantly the anisotropies in Lankford value and work hardening under compression, and cause tension-compression asymmetries (Hama et al. 2017).

A simulation can be coupled with another to restore the thermo-mechanical progress, e.g., 3D CA-CPFE coupling (Li et al. 2016), with dynamical recrystallization (DRX) evolution built into CPFE framework that allows the investigation of isothermal compression in the β phase field. The DRX-induced reduction of slip resistance and GB migration cause a relatively large subsequent plastic deformation around GB, while small strain is found within recrystallized grains. Finally, simulation combining Monte Carlo with stochastic CPFE in a self-consistent viscoplasticity framework shows that microstructural features such as grain size distribution and crystallite orientations govern the uncertainty in the mechanical response of the polycrystalline materials (Askari et al. 2015).

Some special models are introduced to simulate the microstructure evolution during processing and service conditions. A deterministic model was established to account for the effect of initial crack size, microtextured region size, and aspect ratio on the fatigue life of titanium alloys (Pilchak 2014b). The results indicate that dwell and cyclic fatigue life depend mostly on microtextured region size and initial crack size, respectively. The creation of a process-structure-property database for dual-phase titanium alloys by a synthetic microstructure generation method and a mesh-free fast Fourier transform-based micromechanical model has been worked out (Ozturk and Rollett 2017). It is found that the response is most sensitive to α fraction and prior β texture. Increase in α fraction enhanced the tensile strength, as the α phase is harder. Increased prior β (001) texture decreases tensile strength. Microstructural evolution was modeled by LS method to explore the lath spheroidization phenomenon in α/β titanium alloy, with consideration of interface diffusion and the motion driven by curvature (Polychronopoulou et al. 2016), but since deformation was not accounted for in detail, further study is still needed. Thorough understanding may rely on further integration of new technology and simulation (Pilchak et al. 2014a).

5.3 Microstructure Evolution During Advanced Forming Processes

Many new processes emerged in the last decade, some of which may in some way be superior to conventional forming, and can replace them in certain circumstances. Powder metallurgy may produce very uniform composition since it has no

solidification process causing macrosegregation, while additive manufacturing may produce original shape from draft, without mold. For some advanced forming techniques that combine high energy beam melting and solidification process in small volume, the details are difficult to reveal solely by experiments, and their microstructure evolution and defect formation are not easy to control. More complicated model considering various factors needs to be developed and implemented.

Additive manufacturing or 3D printing is an emerging technology based on computer-aided design and manufacture. It can shape complex structure quickly and overcome the limitations of traditional processes. The microstructure generated by this novel technology has been investigated by experiments, and simulated by MC, FE, and CA, or coupling of these methods. Recently, a Potts MC-based model implemented in the SPPARKS framework has been utilized to simulate melting, solidification, and microstructural evolution in the fusion and heat-affected zones of welds (Theron et al. 2017) (Fig. 12). It was shown that weld speed is a key process parameter. Together with shape parameters describing various possible pool shapes and heat-affected zones representing different heating method, it facilitated the design of many possible welding processes.

The evolution of grain structure is simulated using 3D MC model of grain growth in the heat-affected zone of CP-Ti welds (Yang et al. 2000). The calculated grain sizes for various heat inputs were comparable with the results of corresponding experiments. The mean grain size in regions equidistant from the fusion plane increased along the circumferential direction from the top surface to the weld root.

Besides these models, 3D model is proposed to analyze heat transfer and fluid flow in weld pool during electron beam freeform fabrication of Ti-6Al-4 V (Tang et al. 2014). Vigorous fluid flows were found in the front part of weld pool with thermal-capillary force as the main driving force, and frequent droplet impacting influences heat and mass transfer in the weld pool. Scanning electron beam could decrease temperature gradient and magnitude of fluid flow in the weld pool. The grain structure of Ti-6Al-4V after laser melting is explained using a new model combining 3D FE and CA validated by experiments (Dezfoli et al. 2017). The growth direction of elongated grains is controlled by heat flow direction, which is a function of laser scanning speed, and the grain size decreases with increasing laser speed.

Recently a new AM process, hybrid deposition and microrolling (HDMR), is introduced that integrates arc deposition and microrolling. A 2D CA coupling finite volume method is developed for simulating microstructure evolution including solidification and dynamic recrystallization during HDMR forming (Zhou et al. 2016). Rolling reduction was found to control the dynamic recrystallization region and size of recrystallized grains. Computational analysis was also made on linear friction welding of titanium alloys, and the effect of processing parameters evaluated (Li et al. 2010; Yang et al. 2015b). Apparently the simulation on these complicated new processes needs more efforts.
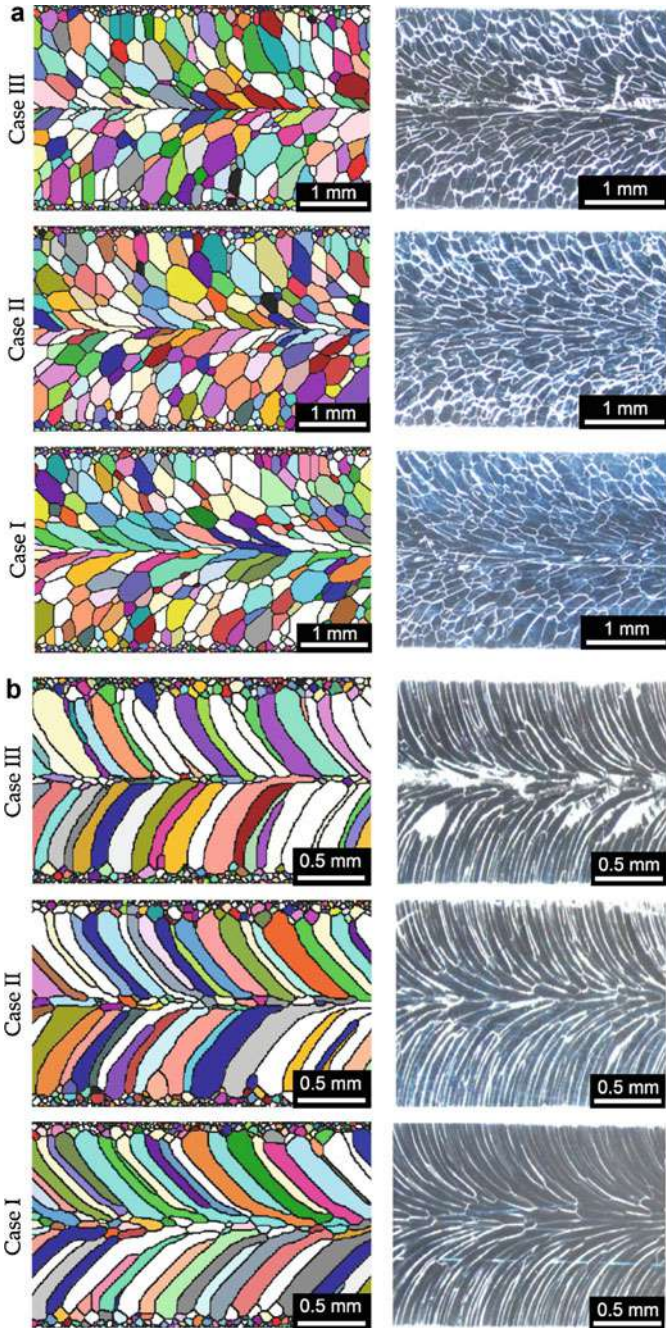
Fig. 12 (**a**) Comparison between simulated and experimental microstructures along the top and (**b**) the bottom surface of a weld (Theron et al. 2017)

6   Finite Element Simulation for Understanding and Optimization of Forming Process

Titanium alloy components are fabricated traditionally through melt-casting, forging, rolling, extrusion, powder metallurgy, and recently additive manufacturing, and many factors should be controlled to achieve the best performance. In addition, optimizations are made to shorten processing routes, alongside performance enhancement and defect control. In order to obtain a consistent quality product at a minimal cost, various simulations are carried out for each manufacturing process in order to reduce the long and costly trial-and-error design iteration and lower the cost at production stage (Avitzur 1968; Semiatin 1984; Gegel et al. 1988; Schey 1997). FEM is the most popular technique to improve the designer's capability to evaluate and optimize various aspects of forming process (Luo et al. 2007; Lacki et al. 2010; Wu and Zhang 2014; Wang et al. 2016b).

6.1   Development of FEM for Various Processes Simulation

Early process simulations relied on simple analytical techniques to estimate forming load, workability limits, and mold design features, then with complex analytical equations converted to charts for easy use. FEM for metal forming was introduced in early-1970, and has impacted industry since mid-1980. Due to its capability and industry demand, FEM has evolved into true 3D models, with robust functionalities since late-1980 (Kobayashi et al. 1989; Malas and Seetharaman 1992; Rowe et al. 2005), and succeeded in reducing R&D cost for various processes for titanium alloys, from vacuum arc remelting (VAR), to forming analysis for processing window control, mold design evaluation, and defect prevention.

The melting of titanium alloy is the most complicated process for simulation, an example being the most popular VAR technology. This process involves melting, liquid flow, diffusion, solidification, and electro-magnetic stirring and cooling, and cannot be solved by a single simulation; models should be established focusing on each aspect. Robust and efficient model for transient analysis of electromagnetic field, flow, heat transfer, macro-segregation, and inclusion behavior was established for the VAR process, and simulations in the last two decades have promoted effectively the development of this technology (Mitchell et al. 1991; Hernandez-Morales and Mitchell 1999; Kelkar et al. 2004; Mitchell 2005; Mitchell and Reed 2009) (Kelkar et al. 2007), see Fig. 13. Recent development of continuous casting of titanium rod and plate ingot for direct rolling may reduce significantly the cost of titanium products (Yu 2013) but is still challenging for various alloys, and simulations are strongly needed.

After melting, ingots are broken down into large rod in the high temperature β phase field and then forged at progressively lower temperature, mostly down to the two-phase field. A typical process for making components for aviation application needs a dozen passes of forging and rolling, and may undergo isothermal forging to

Time = 540 min
Ingot Length = 2.49 m
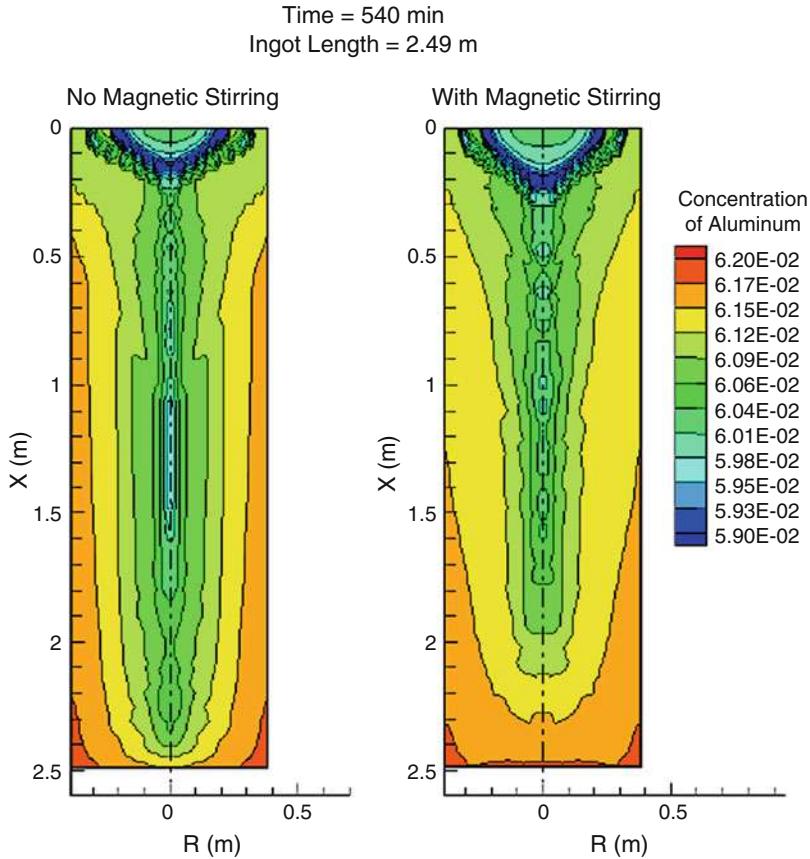
Fig. 13 Simulated variation of Al concentration in Ti-6Al-4V VAR ingot with and without magnetic stirring (Kelkar et al. 2007)

final shape, and then heat-treated to desired microstructure. Shortening processing route is challenging work for production cost reduction without sacrificing performance.

For hot forming technology like extrusion, lubrication is crucial for protection of the alloys and mold, since titanium is extremely reactive at high temperature and reacts with most molds if in direct contact. A continuous film of viscous glass can be good lubricant and separator between titanium and mold; however, the viscosity control is not an easy task, due to the very high temperature gradient across the contact. The glass viscosity is too low and will leak out if temperature is too high, while it becomes hard and brittle otherwise; in such circumstances not only is the lubricant effect lost, but the product may be scratched and destroyed. FEM simulation was carried out to investigate the temperature and viscosity distribution within glass lubricant film under different extrusion condition and, based on that,

lubricant was designed and successful extrusion was achieved. For the canned extrusion of TiAl, the situation becomes more complicated: apart from lubricant, a right dwell time in air is needed for the container to achieve a similar strength to that of the TiAl billet inside. FEM simulations provide key information to glass lubricant design and extrusion process, and lead to successful extrusion after only very limited number of trial (Bai et al. 2010).

6.2 Integrated Simulation for Complete Manufacturing Process Optimization

Most titanium alloy components go through a long manufacturing cycle, which typically includes melting, ingot breakdown, forging, heat-treating, welding, machining, and assembling and installation. The cast microstructure in titanium alloy ingot is usually very coarse, can be broken down through cogging, upsetting and drawing, and forged into semi-finished microstructure, and then die-forged into final products. For each step, the results of last procedure will have strong influence on the next, such as in temperature and microstructure distribution, strain or even defect state; therefore, integrated simulation of the whole process is the dream of the process design engineer. However, since the numerous details need to be considered in each step of the long and complicated process, and tremendous amount of computation is involved in each step, there is no truly integrated simulation throughout. Numerous efforts on these integrations were made in the last decades to produce software enabling the integration of two or more procedures, such as sequential forging or rolling simulation with output of the last simulation into the next as input, or integrate two or more scales, such as FEM simulation with consideration of microstructure changes. FEM simulation was carried out for forging, rolling, extrusion, casting, and powder metallurgy process, with partial consideration of integration of scales e. g., CPFEM, Yada model are used to predict microstructure during deformation, and phase field simulation of microstructure evolution considers various factors under different condition. Since a complete material flow and thermal mechanical history is important to model the behavior properly during production or during service, it is ideal to have integration of different subprocess in a group, if not in one simulation, to gain understanding of microstructure and performance evolution throughout the entire manufacturing process.

The forging of TA15 titanium alloy cylindrical ingot was investigated using 3D FEM simulation. It was found that the deformation strain at each pass influences strain distribution strongly: very small strain of each strike results in near surface deformation, while very large strain forms localized strain in the center. The deformation beneath hammer and between two hammers have opposite distribution, sometimes with the latter having much larger accumulated strain. The forging process was optimized on the basis of FEM strain analysis (Yu et al. 2010). After the ingot was forged into bars, they go through rolling passes or through extrusion into

tube or rod with profiled cross-section. The rolling stress is high in the two-phase field, and the microstructures are sensitive to temperature and strain, which make the processing window narrow, and the pass design and process control difficult. Special Y-shaped roller and unique caliber are designed based on quantitative FEM analysis of the effects of each parameter on temperature and deformation characteristics within rolled bars. Continuous rolling succeeded in making the wire for fasteners with required uniform microstructure and texture.

The constitutive relations are crucial for FEM simulation, and are usually determined experimentally, using physical simulation like Gleeble thermo-mechanical simulator, which gives the stress-strain relations at different temperature and strain rate by controlled tension or compression. The flow stress-strain data should cover the whole process window of actual forming process, including temperature, strain, and strain rate at each position of the components simulated. Large extrapolation of data range usually results in unexpected error. During Gleeble test, factors introducing error should be considered carefully, including the temperature change due to heat loss into anvils and air, the friction between specimen and anvils, and the temperature rise caused by plastic deformation (Meng et al. 2018). FEM simulations on Ti60 and TiAl alloys showed that the temperature distribution in specimen is not uniform during hot compression, resulting in significant deformation inhomogeneity and unignorable error in the flow stress-strain relation, invalidating the uniform strain assumption commonly employed to extract the constitutive relation from Gleeble tests (Yu et al. 2012). Based on FEM simulations with iterative corrections, a scheme to refine the constitutive relations from Gleeble tests was proposed, through which more accurate constitutive relations can be obtained (Yu et al. 2019).

6.3 Integrated Simulation of Microstructure Evolution During Hot Processing

Currently, most FEM modeling considers only isotropic and unchanged behavior of materials. In fact, many factors can alter the material, such as recrystallization, texture formation, phase transformation, grain size evolution, precipitate size and distribution, and even damage, could all potentially affect deformation behavior (Banerjee and Williams 2013; Collins et al. 2013). Understanding the microstructure/properties evolution during thermal mechanical processing and its associated effect on mechanical properties, such as flow stress, modulus, and creep behavior, is very crucial to the model accuracy since these properties are generally the most important input to FEM analysis.

Artificial intelligence method play an increasingly important role as big data (both experimental and simulation) of titanium alloy composition/processing/microstructure/properties are accumulated. The microstructural evolution of titanium alloy under isothermal and other hot forging was predicted using artificial neural networks (ANN) and finite element (FE) simulation (Kim et al. 2009). The change in phase fraction, grain size, and dynamic globularization fraction were

modeled considering various hot working conditions, based on input parameters such as initial specimen/die/environment temperature, steady-state phase fraction, and elapsed time during forging. Resulted ANN models were coupled into FE simulation, with grain size variation and a globularization model developed for hot forging, and validated on Ti-6Al-4V hot-worked under various conditions.

Apart from phase fraction, texture is also very important for mechanical properties. For example, the existence of microtexture is believed to be the root cause of the large fatigue life scattering during dwell fatigue for some titanium alloys such as Ti6242 as rotating components in jet engine (Kassner et al. 1999; Sinha et al. 2006; Rugg et al. 2007; Dunne and Rugg 2008; Germain et al. 2008; Pilchak et al. 2009; Qiu et al. 2014). The mechanism of texture evolution in TA15 sheets during hot tension along the rolling direction was investigated by both experiment and CP simulation (Zhao et al. 2017). Four main orientations in undeformed TA15 sheets were found to evolve in different ways. CPFEM simulation shows that basal slip would rotate <0001> axis off the tensile direction, while prismatic slip rotates $[10\bar{1}0]$ axis oppositely. This lattice rotation results in the transition of orientation C and D, while keeping the orientation of A and B steady. Such understanding on texture change may shed some light on the process design for texture optimization in titanium alloys for critical application such as rotating components in jet engine. Various models and methods to describe the deformation character of various microstructures were developed, while more effort is still needed to couple these seamlessly with fabrication process simulation (Li et al. 2019).

Beyond traditional titanium research community, there are much broader ranges of new idea we may learn from, to advance titanium alloy development. For better mechanical properties, we need to overcome the conflicts between the strength and toughness of material or find a balance between strength and ductility (Ritchie 2011). New strategies for this can be employed, including analyzing the mechanism of each property with the help of simulation, and learning from nature to achieve good strength and damage tolerance (Gao 2006; Lu 2010). With the help of multiscale simulation and new characterization technique, more possibilities of microstructure modification can be found, e.g., ultra-fine microstructure can be obtained by carefully designing thermal-mechanical processing or heat treatment route through nonconventional transformation pathways utilizing some metastable phases such as ω or O″ phase (Wang et al. 2008b, 2010; Ping 2014; Zheng et al. 2016a, b, c). Through specially engineered coherency of internal boundaries and stability control, both high strength and conductivity can be achieved (Lu et al. 2009). Ultra-strong steel was designed by introducing high-density of nanoscale precipitates with minimum lattice misfit (Jiang et al. 2017), or even forming amorphous phase with the help of martensitic transformation (Zhang et al. 2018). New approach to achieve smaller grains with strong GB may be realized with the help of deeper understanding of the segregation and embrittlement of GB by integrated simulation (Chookajorn et al. 2012; Kalidindi et al. 2015; Polyakov et al. 2016; Lejcek et al. 2017).

7   Perspective and Challenges

7.1   Perspectives

1. The performance requirement for titanium alloys is becoming more application specific, so materials are now not an independent choice as before. In order to enhance performance further to fit new design, application-specific developments are needed for critical components. It is the whole processing technology that ensures the successful and economic application of material. Therefore, the development of new materials needs to take a different route, to achieve better performance balance for each usage, using simulation as much as possible, and make early contact between materials researcher, manufacturer and system designer, and communicate what they really need and what they can offer.

2. Developing a new material needs to go through composition selection, microstructure and properties optimization, and processing optimization, etc., then a series of component design is tried, to obtain the best component design. In order to accelerate this process, in the future, the different stages of materials development and the system design can proceed in parallel, with the help from multiscale computation. The performance of components is simulated well in advance before materials are made, and materials designer needs to work together with component designer to adjust the composition, microstructure, and processing route according to the requirement of the specific application. More challenging problem can only be solved efficiently by strong collaboration between computational specialist and experimentalist, and between materials and system designer (Rugg 2014).

3. Future computations should focus on bottleneck problems in each application and serve either to identify the controlling factor in processing or to reveal working mechanism which is not easy to uncover solely by experiments. Materials for the compressor of jet engine require high strength, toughness, fatigue, creep and oxidation resistance under high temperature, etc., but from the first to the end stage, the temperature and pressure increase gradually, and the bottleneck may shift from fatigue to creep, or even oxidation, so the materials design should adjust accordingly.

4. Traditionally our processing route is designed to achieve uniform microstructure throughout components; as the demanding for materials increases, position-specific inhomogeneous microstructure and composition (functional gradient microstructure, or architectured microstructure) may be designed. With special arrangements, microstructure such as gradient/intermixed grain size/composition/phase may enhance the strength-ductility compatibility, and be beneficial to the performance of components (Lu 2014; Zeng et al. 2016; Jiang et al. 2017). Computation at different scales may help on designing better materials and processes of fabrication.

5. With the advancement of computation, various integrations are possible, either of different scales, e.g., FEM with microstructure changes, or MD with experimental timescale, or linking different processes, such as passing details of component along different stages of manufacturing. New advances in computing such as artificial intelligence should also be employed in both simulation and analysis.

7.2   Challenges

1. The rapid development of high-techs need more advanced titanium alloys, often urgently, but cannot tolerate its long insertion time of 10–20 years like before. The MGI and ICME approach should be adopted throughout materials composition design until fabrication, to accelerate new materials development (Kubel 2013; Glavicic and Venkatesh 2014). How to communicate between different parties and what information to share need some new regulations, involving not only science and technology, but also economy or even politics, especially for the strategically important materials like titanium alloys.
2. Computations focused on more specific processes need better input, such as proper boundary conditions and constitutive relation. However, realistic determination of processing parameters, such as constitutive relation under different stresses, elastic properties of phases with different composition/temperature, and friction/thermal conductivity coefficient at interfaces between different parts during hot working, may not be an easy task. New technologies for achieving such parameters are needed for titanium alloys studies, including high throughput computation/characterization method, or from some data sharing mechanism (Curtarolo et al. 2012; Christodoulou 2013; Jain et al. 2013; Agren 2014; Wang et al. 2014a, e; Zhao 2014).
3. Multiscale modeling and simulation have been utilized in the last few decades to solve basic properties, such as energy and modulus of phases, and gradually to more complicated problems, such as strength, toughness, fatigue, and stress corrosion cracking, which are much more challenging. Modeling and simulation should be fully utilized and designed carefully to help identify the controlling factors, and to promote our understanding, and help reduce the number of experiment and cost of the development, and finally to help materials to get out of the "death valley."
4. Simulations have been successful in promoting manufacturing of different types, such as casting, ingot melting, forging, rolling, extrusion, and welding. New processing technologies may need more help from the simulation community, e.g., powder metallurgy, additive manufacturing, and friction stir welding. In each of these processes, the microstructure changes and properties variation during processing are difficult to control and largely unknown, and simulation may do more to solve these challenging problems.

References

Agren J (2014) The materials genome and CALPHAD. Chin Sci Bull 59:1635

Anahid M, Ghosh S (2013) Homogenized constitutive and fatigue nucleation models from crystal plasticity FE simulations of Ti alloys. Part 2: macroscopic probabilistic crack nucleation model. Int J Plast 48:111

Appel F, Paul JDH, Oehring M (2011) Gamma titanium aluminide alloys: science and technology. Wiley-VCH Verlag GmbH & Co. KGaA

Appel F, Clemens H, Fischer FD (2016) Modeling concepts for intermetallic titanium aluminides. Prog Mater Sci 81:55

Askari H, Maughan MR, Abdolrahim N, Sagapuram D, Bahr DF, Zbib HM (2015) A stochastic crystal plasticity framework for deformation of micro-scale polycrystalline materials. Int J Plast 68:21

Avitzur B (1968) Metal forming: processes and analysis. McGraw-Hill, New York

Bai CG, Xu DS, Lei JF, Liu YY, Xu Z, Liu MH, Lv PC, Duan SJ, Yang R (2010) FEM simulation of glass lubrication during section extrusion of titanium alloy. Chin J Nonferrous Metals 20:s857

Banerjee D, Williams JC (2013) Perspectives on titanium science and technology. Acta Mater 61:844

Cai W, de Koning M, Bulatov VV, Yip S (2000) Minimizing boundary reflections in coupled-domain simulations. Phys Rev Lett 85:3213

Chen LQ (2002) Phase-field models for microstructure evolution. Annu Rev Mater Res 32:113

Chen LQ (2014) Phase-field method and Materials Genome Initiative (MGI). Chin Sci Bull 59:1641

Chookajorn T, Murdoch HA, Schuh CA (2012) Design of stable nanocrystalline alloys. Science 337:951

Christodoulou J (2013) Integrated computational materials engineering and materials genome initiative: Accelerating Materials Innovation. Adv Mat Proc: 28

Clark SJ, Segall MD, Pickard CJ, Hasnip PJ, Probert MJ, Refson K, Payne MC (2005) First principles methods using CASTEP. Zeitschrift Fur Kristallographie 220:567

Clemens H, Mayer S (2013) Design, processing, microstructure, properties, and applications of advanced intermetallic TiAl alloys. Adv Eng Mater 15:191

Clouet E, Caillard D, Chaari N, Onimus F, Rodney D (2015) Dislocation locking versus easy glide in titanium and zirconium. Nat Mater, advance online publication

Collins PC, Koduri S, Welk B, Tiley J, Fraser HL (2013) Neural networks relating alloy composition, microstructure, and tensile properties of alpha/beta-Processed TIMETAL 6-4. Metall Mater Trans A 44:1441

Cottura M, Appolaire B, Finel A, Le Bouar Y (2014) Phase field study of acicular growth: role of elasticity in Widmanstätten structure. Acta Mater 72:200

Curtarolo S, Setyawan W, Wang SD, Xue JK, Yang KS, Taylor RH, Nelson LJ, Hart GLW, Sanvito S, Buongiorno-Nardelli M, Mingo N, Levy O (2012) A distributed materials properties repository from high-throughput ab initio calculations. Comput Mater Sci 58:227

Dezfoli ARA, Hwang W-S, Huang W-C, Tsai T-W (2017) Determination and controlling of grain structure of metals after laser incidence: theoretical approach. Sci Rep 7:41527

Ding R, Guo ZX (2002) Microstructural modelling of dynamic recrystallisation using an extended cellular automaton approach. Comput Mater Sci 23:209

Dunne FPE, Rugg D (2008) On the mechanisms of fatigue facet nucleation in titanium alloys. Fatigue Fract Eng Mater Struct 31:949

Dunne FPE, Wilkinson AJ, Allen R (2007b) Experimental and computational studies of low cycle fatigue crack nucleation in a polycrystal. Int J Plast 23:273

Dunne FPE, Rugg D, Walker A (2007a) Lengthscale-dependent, elastically anisotropic, physically-based hcp crystal plasticity: application to cold-dwell fatigue in Ti alloys. Int J Plast 23:1061

Dunne FPE, Walker A, Rugg D (2007c) A systematic study of hcp crystal orientation and morphology effects in polycrystal deformation and fatigue. Proc R Soc A 463:1467

Dunne FPE, Kiwanuka R, Wilkinson AJ (2012) Crystal plasticity analysis of micro-deformation, lattice rotation and geometrically necessary dislocation density. Proc R Soc A 468:2509

Essmann U, Mughrabi H (1979) Annihilation of dislocations during tensile and cyclic deformation and limits of dislocation densities. Philos Mag A 40:731

Gao HJ (2006) Application of fracture mechanics concepts to hierarchical biomechanics of bone and bone-like materials. Int J Fract 138:101

Gegel HL, Malas JC, Doraivelu SM, Shende V (1988) Modeling techniques used in forging process design, forming and forging. ASM Handbook 14:417

Germain L, Gey N, Humbert M, Vo P, Jahazi M, Bocher P (2008) Texture heterogeneities induced by subtransus processing of near alpha titanium alloys. Acta Mater 56:4298

Ghazisaeidi M, Trinkle DR (2012) Core structure of a screw dislocation in Ti from density functional theory and classical potentials. Acta Mater 60:1287

Ghazisaeidi M, Trinkle DR (2014) Interaction of oxygen interstitials with lattice faults in Ti. Acta Mater 76:82

Ghosh S, Anahid M (2013b) Homogenized constitutive and fatigue nucleation models from crystal plasticity FE simulations of Ti alloys. Part 1: macroscopic anisotropic yield function. Int J Plast 47:182

Ghosh S, Chakraborty P (2013a) Microstructure and load sensitive fatigue crack nucleation in Ti-6242 using accelerated crystal plasticity FEM simulations. Int J Fatigue 48:231

Girshick A, Pettifor DG, Vitek V (1998) Atomistic simulation of titanium – II. Structure of 1/3 <1210> screw dislocations and slip systems in titanium. Philos Mag A 77:999

Glavicic MG, Venkatesh V (2014) Integrated computational materials engineering of titanium: current capabilities being developed under the metals affordability initiative. JOM 66:1310

Gu XY, Xu DS, Wang H, Yang R (2010) Lattice weakening by edge dislocation core under tension. Model Simul Mater Sci Eng 18:065004

Guan SX, Kang Q, Wang QJ, Liu YY, Li D (1998) Influence of long-term thermal exposure on the tensile properties of a high-temperature titanium alloy Ti-55. Mater Sci Eng A 243:182

Hafner J (2008) Ab-initio simulations of materials using VASP: density-functional theory and beyond. J Comput Chem 29:2044

Hama T, Kobuki A, Takuda H (2017) Crystal-plasticity finite-element analysis of anisotropic deformation behavior in a commercially pure titanium Grade 1 sheet. Int J Plast 91:77

Han F, Tang B, Kou H, Li J, Feng Y (2015) Experiments and crystal plasticity finite element simulations of nanoindentation on Ti-6Al-4V alloy. Mater Sci Eng A 625:28

Hao YL, Xu DS, Cui YY, Yang R, Li D (1999) The site occupancies of alloying elements in TiAl and Ti_3Al alloys. Acta Mater 47:1129

Hao YL, Li SJ, Sun BB, Sui ML, Yang R (2007) Ductile titanium alloy with low Poisson's ratio. Phys Rev Lett 98:216405

Hao YL, Zhang ZB, Li SJ, Yang R (2012) Microstructure and mechanical behavior of a Ti–24Nb–4Zr–8Sn alloy processed by warm swaging and warm rolling. Acta Mater 60:2169

He Y, Liu Z, Zhou G, Wang H, Bai CG, Rodney D, F Appel DSX, Yang R (2018) Dislocation dipole-induced strengthening in intermetallic TiAl. Scr Mater 143:98

Hennig RG, Trinkle DR, Bouchet J, Srinivasan SG, Albers RC, Wilkins JW (2005) Impurities block the alpha to omega martensitic transformation in titanium. Nat Mater 4:129

Hennig RG, Lenosky TJ, Trinkle DR, Rudin SP, Wilkins JW (2008) Classical potential describes martensitic phase transformations between the alpha, beta, and omega titanium phases. Phys Rev B 78:054121

Heo TW, Chen LQ (2014a) Phase-field modeling of displacive phase transformations in elastically anisotropic and inhomogeneous polycrystals. Acta Mater 76:68

Heo TW, Shih DS, Chen LQ (2014b) Kinetic pathways of phase transformations in two-phase Ti alloys. Metall Mater Trans A 45:3438

Hernandez-Morales B, Mitchell A (1999) Review of mathematical models of fluid flow, heat transfer, and mass transfer in electroslag remelting process. Ironmaking Steelmaking 26:423

Hu JM, Yang TN, Chen LQ, Nan CW (2013) Engineering domain structures in nanoscale magnetic thin films via strain. J Appl Phys 114:164303

Hu QM, Xu DS, Li D (2001) First-principles investigations of the solute-vacancy interaction energy and its effect on the creep properties of alpha-titanium. Philos Mag A 81:2809

Hu QM, Xu DS, Yang R, Li D, Wu WT (2003) First-principles investigations of ordering in binary alpha-Ti solid solutions. Philos Mag 83:217

Hu QM, Yang R, Hao YL, Xu DS, Li D (2004) Concentrated point defects in and order-disorder transition temperature of intermetallic compounds. Phys Rev Lett 92:4

Hu Q-M, Vitos L, Yang R (2014) Theoretical investigation of the omega-related phases in TiAl-Nb/Mo alloys. Phys Rev B 90:054109

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: The materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:011002

Ji YZ, Heo T, Zhang F, Chen LQ (2016) Theoretical assessment on the phase transformation kinetic pathways of multi-component Ti alloys: application to Ti-6Al-4V. J Phase Equilib Diffus 37:53

Jiang SH, Wang H, Wu Y, Liu XJ, Chen HH, Yao MJ, Gault B, Ponge D, Raabe D, Hirata A, Chen MW, Wang YD, Lu ZP (2017) Ultrastrong steel via minimal lattice misfit and high-density nanoprecipitation. Nature 544:460

Kalidindi AR, Chookajorn T, Schuh CA (2015) Nanocrystalline materials at equilibrium: a thermodynamic review. JOM 67:2834

Kassner ME, Kosaka Y, Hall JA (1999) Low-cycle dwell-time fatigue in Ti-6242. Metall Mater Trans A 30:2383

Katzarov IH, Cawkwell MJ, Paxton AT, Finnis MW (2007) Atomistic study of ordinary (1)/(2) < 110] screw dislocations in single-phase and lamellar gamma-TiAl. Philos Mag 87:1795

Kelkar KM, Mok J, Patankar SV, Mitchell A (2004) Computational modeling of electroslag remelting processes. J Phys IV 120:421

Kelkar KM, Patankar SV, Mitchell A, Kanou O, Fukada N, Suzuki K (2007) Computational modeling of the Vacuum Arc Remelting (VAR) process used for the production of ingots of titanium alloys. In: Niinomi M (ed) Ti-2007: Proceedings of the 11th world conference on Titanium, pp. 1279–1282

Kim JH, Reddy NS, Yeom JT, Hong JK, Lee CS, Park N-K (2009) Microstructure prediction of two-phase titanium alloy during hot forging using artificial neural networks and FE simulation. Met Mater Int 15:427

Kim YW (1998) Strength and ductility in TiAl alloys. Intermetallics 6:623

Kim Y-W, Kim S-L (2018) Advances in gammalloy materials-processes-application technology: successes, dilemmas, and future. JOM 70:553

Kobayashi S, Oh SI, Altan T (1989) Metal forming and the finite element method. Oxford University Press, New York

Kubel E (2013) AM, ICME, and MGI gaining momentum. Adv Mater Processes 171:2

Kumar A, Wang J, Tome CN (2015) First-principles study of energy and atomic solubility of twinning-associated boundaries in hexagonal metals. Acta Mater 85:144

Lacki P, Adamus J, Motyka M (2010) FEM simulation of superplastic forming of a spherical cap made of Ti-6Al-4V. Inżynieria Materiałowa 31:720

Lejcek P, Sob M, Paidar V (2017) Interfacial segregation and grain boundary embrittlement: An overview and critical assessment of experimental data and calculated results. Prog Mater Sci 87:83

Leyens C, Peters M (2003) Titanium and titanium alloys, fundamentals and applications. Wiley, Weinheim

Li B, El Kadiri H, Horstemeyer MF (2012) Extended zonal dislocations mediating <11$\bar{2}$2><11$\bar{2}\bar{3}$> twinning in titanium. Philos Mag 92:1006

Li D, Liu YY, Wan XJ (1984) On the thermal stability of Ti alloys: I the electron concentration rule for formation of Ti3X-phase; II the behaviour of transition elements in Ti3X-phase formation; III the criterion for thermal stability and its application. Acta Metall Sin 20(6):375–397

Li H, Sun X, Yang H (2016) A three-dimensional cellular automata-crystal plasticity finite element model for predicting the multiscale interaction among heterogeneous deformation, DRX microstructural evolution and mechanical responses in titanium alloys. Int J Plast 87:154

Li WY, Ma TJ, Li JL (2010) Numerical simulation of linear friction welding of titanium alloy: Effects of processing parameters. Mater Des 31:1497

Li XX, Xu DS, Yang R (2019) Crystal plasticity finite element investigation of the high temperature deformation consistency in dual-phase titanium alloy. Acta Metall Sin 55:928

Li Y-J, Hu Q-M, Xu D-S, Yang R (2011) Strengthening of γ-TiAl-Nb by short-range ordering of point defects. Intermetallics 19:793

Liao Y-S, Chyuan S-W, Chen J-T (2004) FEM versus BEM. IEEE Circuits Devices Mag 20:25

Liu YX, Wang H, Wu HN, Xu DS, Yang R (2014) A mean-field interatomic potential for a multi-component β-type titanium alloy. Comput Mater Sci 95:414

Lu K (2010) The Future of Metals. Science 328:319

Lu K (2014) Making strong nanomaterials ductile with gradients. Science 345:1455

Lu K, Lu L, Suresh S (2009) Strengthening materials by engineering coherent internal boundaries at the nanoscale. Science 324:349

Luo YS, Yu M, Peng XH (2007) The 3D modeling of dies based on UG and numerical simulation of the heat rheological forming of titanium alloy vane disk. Key Eng Mater 340–341:725

Lütjering G, Williams JC (2007) Titanium, 2nd edn. Springer, New York

Malas JC, Seetharaman V (1992) Using material behavior models to develop process control strategies. JOM 44:8

Manonukul A, Dunne FPE (2004) High- and low-cycle fatigue crack initiation using polycrystal plasticity. Proc R Soc Lond A 460:1881

Matthies HG, Kayser-Herold O (2005) Least-squares FEM literature review

Mitchell A (2005) Solidification in remelting processes. Mater Sci Eng A 413:10

Mitchell A, Reed RC (2009) The influence of remelting processes on the mechanical properties of forged alloys. High Temp Mater Processes 28:285

Mitchell A, Tripp DW, Cockcroft SL, Entrekin CH (1991) Modeling the Thermal Fields in Electron-Beam Remelted Ingots. Proceedings of the conference on electron bean melting and refining – state of the art 1991, p 248, Bakish Materials Corp, Englewood

Meng QG, Bai CG, Xu DS (2018) Flow behavior and processing map for hot deformation of ATI425 titanium alloy. J Mater Sci Technol 34:679

Morishita M, Ashida Y, Chikuda M, Morinaga M, Yukawa N, Adachi H (1991) Active Corrosion Rate for Ti-Based Alloys in Aqueous Corrosion and Its Correlation with the Bond Order Obtained by Electron Theory. ISIJ Int 31:890

Nag S, Zheng Y, Williams REA, Devaraj A, Boyne A, Wang Y, Collins PC, Viswanathan GB, Tiley JS, Muddle BC, Banerjee R, Fraser HL (2012) Non-classical homogeneous precipitation mediated by compositional fluctuations in titanium alloys. Acta Mater 60:6247

Ni Y, Khachaturyan AG (2009) From chessboard tweed to chessboard nanowire structure during pseudospinodal decomposition. Nat Mater 8:410

Oberson PG, Wyatt ZW, Ankem S (2011) Modeling interstitial diffusion controlled twinning in alpha titanium during low-temperature creep. Scr Mater 65:638

Ozturk T, Rollett AD (2017) Effect of microstructure on the elasto-viscoplastic deformation of dual phase titanium structures. Comput Mech 61:55

Panova J, Farkas D (1998) Atomistic simulation of dislocation core configurations in TiAl. Philos Mag A 78:389

Park CL, Voorhees PW, Thornton K (2014) Application of the level-set method to the analysis of an evolving microstructure. Comput Mater Sci 85:46

Pilchak AL (2014b) A simple model to account for the role of microtexture on fatigue and dwell fatigue lifetimes of titanium alloys. Scr Mater 74:68

Pilchak AL, Bhattacharjee A, Rosenberger AH, Williams JC (2009) Low delta K faceted crack growth in titanium alloys. Int J Fatigue 31:989

Pilchak AL, Li J, Rokhlin SI (2014a) Quantitative comparison of microtexture in near-alpha titanium measured by ultrasonic scattering and electron backscatter diffraction. Metall Mater Trans A Phys Metall Mater Sci 45a:4679

Ping DH (2014) Review on ω phase in body-centered cubic metals and alloys. Acta Metallurgica Sinica (English Letters) 27:1

Polyakov MN, Chookajorn T, Mecklenburg M, Schuh CA, Hodge AM (2016) Sputtered Hf-Ti nanostructures: a segregation and high-temperature stability study. Acta Mater 108:8

Polychronopoulou D, Bozzolo N, Pino Muñoz D, Bruchon J, Shakoor M, Millet Y, Dumont C, Freiherr von Thüngen I, Besnard R, Bernacki M (2016) Introduction to the level-set full field modeling of laths spheroidization phenomenon in α/β titanium alloys. MATEC Web Conf 80:02003

Porizek R, Znam S, Nguyen-Manh D, Vitek V, Pettifor DG (2003) Atomistic studies of dislocation glide in γ-TiAl. Defect properties and related phenomena in intermetallic alloys, MRS Proceedings, vol 753. p 181, Warrendale, MRS

Pourian MH, Pilvin P, Bridier F, Bocher P (2014) Heterogeneous elastic behavior of HCP titanium polycrystalline aggregates simulated by cellular automaton and finite element. Comput Mater Sci 92:468

Pourian MH, Pilvin P, Bridier F, Bocher P (2015) Modeling the elastoplastic behaviors of alpha Ti-alloys microstructure using cellular automaton and finite element methods. Comput Mater Sci 99:33

Pourian MH, Bridier F, Pilvin P, Bocher P (2016) Prediction of crack initiation sites in alpha Ti-alloys microstructures under dwell-fatigue using cellular automaton simulation method. Int J Fatigue 85:85

Qiu JK, Ma YJ, Lei JF, Liu YY, Huang AJ, Rugg D, Yang R (2014) A comparative study on dwell fatigue of Ti-6Al-2Sn-4Zr-xMo (x=2 to 6) alloys on a microstructure-normalized basis. Metal Mater Trans A 45:6075

Qiu J, Feng X, Ma Y, Lei J, Liu Y, Huang A, Rugg D, Yang R (2016) Fatigue crack growth behavior of beta-annealed Ti–6Al–2Sn–4Zr–xMo (x = 2, 4 and 6) alloys: influence of microstructure and stress ratio. Int J Fatigue 83(Part 2):150

Raabe D, Roters F (2004) Using texture components in crystal plasticity finite element simulations. Int J Plast 20:339

Raabe D, Sander B, Friák M, Ma D, Neugebauer J (2007) Theory-guided bottom-up design of [beta]-titanium alloys as biomaterials based on first principles calculations: theory and experiments. Acta Mater 55:4475

Ren J, Sun Q, Xiao L, Ding X, Sun J (2014) Phase transformation behavior in titanium single-crystal nanopillars under [0001] orientation tension: a molecular dynamics simulation. Comput Mater Sci 92:8

Ritchie RO (2011) The conflicts between strength and toughness. Nat Mater 10:817

Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Acta Mater 58:1152

Rowe GW, Sturgess CEN, Hartley P (2005) Finite-element plasticity and metal forming analysis. Cambridge University Press, Cambridge

Rugg D (2014) Materials for future gas turbine applications. Mater Sci Technol 30:1848

Rugg D, Dixon M, Dunne FPE (2007) Effective structural unit size in titanium alloys. J Strain Anal Eng Des 42:269

Schey J (1997) Manufacturing processes and their selection. Mater Sel Des 20:687

Semiatin SL (1984) Workability in forging In: Dieter GE (ed) Workability testing techniques. American Society for Metals, p 197

Serra A, Bacon DJ (1986) Computer simulation of twin boundaries in the h.c.p. metals. Philos Mag A 54:793

Shi R, Wang Y (2013) Variant selection during alpha precipitation in Ti-6Al-4V under the influence of local stress - A simulation study. Acta Mater 61:6006

Shi R, Dixit V, Fraser HL, Wang Y (2014) Variant selection of grain boundary α by special prior β grain boundaries in titanium alloys. Acta Mater 75:156

Shi R, Zhou N, Niezgoda SR, Wang Y (2015) Microstructure and transformation texture evolution during α precipitation in polycrystalline α/β titanium alloys – a simulation study. Acta Mater 94:224

Shi R, Wang D, Wang Y (2016) Modeling and simulation of microstructure evolution during heat treatment of titanium alloys. ASM International, Materials Park

Simmons JP, Rao SI, Dimiduk DM (1997) Atomistics simulations of structures and properties of 1/2[110] dislocations using three different embedded-atom method potentials fit to γ-TiAl. Philos Mag A 75:1299

Sinha V, Mills MJ, Williams JC (2006) Crystallography of fracture facets in a near-alpha titanium alloy. Metall Mater Trans A 37:2015

Smith BD, Shih DS, McDowell DL (2016) Fatigue hot spot simulation for two Widmanstätten titanium microstructures. Int J Fatigue 92:116

Soler JM, Artacho E, Gale JD, Garcia A, Junquera J, Ordejon P, Sanchez-Portal D (2002) The SIESTA method for ab initio order-N materials simulation. J Phys Condens Matter 14:2745

Song Y, Xu DS, Yang R, Li D, Wu WT, Guo ZX (1999) Theoretical study of the effects of alloying elements on the strength and modulus of beta-type bio-titanium alloys. Mater Sci Eng A 260:269

Steinbach I (2013) Phase-field model for microstructure evolution at the mesoscopic scale. Annu Rev Mater Res 43:89

Tadaki T, Nakata Y, Shimizu K (1995) J Phys Paris 5:C8–C81

Tan Y, Liu Y, Maniatty AM, Zheng C, Wen JT (2017) Monte Carlo grain growth modeling with local temperature gradients. Model Simul Mater Sci Eng 25:065003

Tang Q, Pang S, Chen B, Suo H, Zhou J (2014) A three dimensional transient model for heat transfer and fluid flow of weld pool during electron beam freeform fabrication of Ti-6-Al-4-V alloy. Int J Heat Mass Transf 78:203

Teng CY, Zhou N, Wang Y, Xu DS, Du A, Wen YH, Yang R (2012) Phase-field simulation of twin boundary fractions in fully lamellar TiAl alloys. Acta Mater 60:6372

Teng CY, Du A, Xu DS, Wang Y, Yang R (2015) Effect of external stress on gamma nucleation and evolution in TiAl alloys. Intermetallics 65:1

Theron MR, John AM, Veena T (2017) A Monte Carlo model for 3D grain evolution during welding. Model Simul Mater Sci Eng 25:064006

Trinkle DR, Hennig RG, Srinivasan SG, Hatch DM, Jones MD, Stokes HT, Albers RC, Wilkins JW (2003) New mechanism for the alpha to omega martensitic transformation in pure titanium. Phys Rev Lett 91:025701

Vermaak N, Michailidis G, Parry G, Estevez R, Allaire G, Bréchet Y (2014) Material interface effects on the topology optimization of multi-phase structures using a level set method. Struct Multidiscip Optim 50:623

Veyssiere P, Wang H, Xu DS, Chiu YL (2009) Local dislocation reactions, self-organization and hardening in single slip. Dislocations 2008(3):012018

Wang BD, Dai JH, Wu X, Song Y, Yang R (2015) First-principles study of the bonding characteristics of $TiAl(111)/Al_2O_3(0001)$ interface. Intermetallics 60:58

Wang D, Zhang Z, Zhang JA, Zhou YM, Wang Y, Ding XD, Wang YZ, Ren XB (2010) Strain glass in Fe-doped Ti-Ni. Acta Mater 58:6206

Wang D, Shi RP, Zheng YF, Banerjee R, Fraser HL, Wang YZ (2014a) integrated computational materials engineering (ICME) approach to design of novel microstructures for Ti-alloys. JOM 66:1287

Wang G, Xu DS, Ma N, Zhou N, Payton EJ, Yang R, Mills MJ, Wang Y (2009a) Simulation study of effects of initial particle size distribution on dissolution. Acta Mater 57:316

Wang G, Xu DS, Yang R (2009b) Phase field simulation on sideplates formation in Ti-6Al-4V alloy. Acta Phys Sin 58:S343

Wang G, Xu DS, Payton EJ, Ma N, Yang R, Mills MJ, Wang Y (2011a) Mean-field statistical simulation of grain coarsening in the presence of stable and unstable pinning particles. Acta Mater 59:4587

Wang H, Xu DS, Yang R, Veyssière P (2008a) The transformation of edge dislocation dipoles in aluminium. Acta Mater 56:4608

Wang H, Xu DS, Yang R, Veyssière P (2009c) The transformation of narrow dislocation dipoles in selected fcc metals and in γ-TiAl. Acta Mater 57:3725

Wang H, Xu DS, Yang R, Veyssière P (2011b) The formation of stacking fault tetrahedra in Al and Cu: I. Dipole annihilation and the nucleation stage. Acta Mater 59:10

Wang H, Xu DS, Yang R, Veyssière P (2011c) The formation of stacking fault tetrahedra in Al and Cu: II. SFT growth by successive absorption of vacancies generated by dipole annihilation. Acta Mater 59:1

Wang H, Xu DS, Yang R, Veyssière P (2011d) The formation of stacking fault tetrahedra in Al and Cu: III. Growth by expanding ledges. Acta Mater 59:19

Wang H, Xu DS, Yang R, Rugg D, Huang AJ (2012) The influence of loading directions on the failure of the Bicrystal in alpha-Ti with grain-boundary penetrating dislocations. In: Proceedings of the 12th world conference on titanium, vol II. Science Press, Beijing, p 1219

Wang H, Xu DS, Rodney D, Veyssière P, Yang R (2013a) Atomistic investigation of the annihilation of non-screw dislocation dipoles in Al, Cu, Ni and γ-TiAl. Model Simul Mater Sci Eng 21:025002

Wang H, Xu DS, Yang R (2014b) Defect clustering upon dislocation annihilation in α-titanium and α-zirconium with hexagonal close-packed structure. Model Simul Mater Sci Eng 22:085004

Wang H, Xu DS, Yang R (2014c) Atomic modelling of crack initiation on twin boundaries in α-titanium under external tensile loading along various orientations. Philos Mag Lett 94:779

Wang H, Zhou G, Qiu JK, Ye LH, Xu DS, Lei JF, Rugg D, Huang AJ, Yang R (2016a) Grain boundary-mediated lattice reorientation in a-titanium to promote plastic deformation in hard-oriented grains. In: Proceedings of the 13th world conference on titanium. TMS, Pittsburgh, 1893–1896

Wang H, Li Y, Wang G (2016b) Finite element modeling and simulation towards metallic microstructure of titanium alloy based on Abaqus. Key Eng Mater 667:292

Wang J, Beyerlein IJ, Hirth JP, Tomé CN (2011f) Twinning dislocations on {1011} and {1013} planes in hexagonal close-packed crystals. Acta Mater 59:3990

Wang J, Yadav SK, Hirth JP, Tomé CN, Beyerlein IJ (2013b) Pure-shuffle nucleation of deformation twins in hexagonal-close-packed metals. Mater Res Lett 1:126

Wang JW, Narayanan S, Huang JY, Zhang Z, Zhu T, Mao SX (2013c) Atomic-scale dynamic process of deformation-induced stacking fault tetrahedra in gold nanocrystals. Nat Commun 4:2340

Wang QJ, Liu JR, Yang R (2014d) High temperature titanium alloys: status and perspective. J Aeronaut Mater 34:1

Wang WY, Zhang Y, Li J, Zou C, Tang B, Wang H, Lin D, Wang J, Kou H, Xu D (2018) Insight into solid-solution strengthened bulk and stacking faults properties in Ti alloys: a comprehensive first-principles study. J Mater Sci 53:7493

Wang Y, Ren X, Otsuka K, Saxena A (2008b) Temperature-stress phase diagram of strain glass Ti48.5Ni51.5. Acta Mater 56:2885

Wang Z, Yang XY, Zheng YF, Yong QL, Su H, Yang CF (2014e) Integrated materials design and informatics platform within the materials genome framework. Chin Sci Bull 59:1755

Westbrook JH (1960) Mechanical properties of intermetallic compounds. Wiley, New York

Wooding SJ, Bacon DJ, Phythian WJ (1995) A computer-simulation study of displacement cascades in alpha-titanium. Philos Mag A 72:1261

Woodward C, Rao SI (2004) Ab-initio simulation of (a/2)< 110] screw dislocations in gamma-TiAl. Philos Mag 84:401

Wu HB, Zhang SJ (2014) 3D FEM simulation of milling process for titanium alloy Ti6Al4V. Int J Adv Manuf Technol 71:1319

Wu HN, Xu DS, Wang H, Yang R (2016) Molecular Dynamics Simulation of Tensile Deformation and Fracture of gamma-TiAl with and without Surface Defects. J Mater Sci Technol 32:1033

Xu DS, Song Y, Li D, Hu ZQ (1997a) Site occupancy of alloying elements in Ti3Al compounds. Mater Sci Eng A 234:230

Xu DS, Song Y, Li D, Hu ZQ (1997b) Site occupancy of alloying elements in TiNi compounds. Philos Mag A 75:1185

Xu DS, Chang JP, Li J, Yang R, Li D, Yip S (2004) Dislocation slip or deformation twinning: confining pressure makes a difference. Mater Sci Eng A 840:387–389

Xu DS, Wang H, Yang R, Veyssière P (2008a) Molecular dynamics investigation of deformation twinning in γ-TiAl sheared along the pseudo-twinning direction. Acta Mater 56:1065

Xu DS, Wang H, Li YJ, Yang R (2008b) Deformation of TiAl by the formation of L1(1) pseudo-twin. In: Kim YW, Morris D, Yang R, Leyens C (eds) Structural aluminides for elevated temperatures. Warrendale, TMS, p 119

Xu DS, Wang H, Yang R, Veyssiere P (2009) Point Defect Formation by Dislocation Reactions in TiAl. Dislocations 2008(3):012024

Xu DS, Wang H, Yang R, Sachdev AK (2014) MD simulation of asymmetric nucleation and motion of < 011] superdislocations in TiAl. Chin Sci Bull 59:1725

Xu Z, Li D, Xu W, Devaraj A, Colby R, Thevuthasan S, Geiser BP, Larson DJ (2015) Simulation of heterogeneous atom probe tip shapes evolution during field evaporation using a level set method and different evaporation models. Comput Phys Commun 189:106

Yang JL, Xiao CY, Xia SD, Wang KL (1993) Site Preference of Alloying Additions in Intermetallic Compounds. J Phys Condens Matter 5:6653

Yang M, Wang G, Teng C, Xu D, Zhang J, Yang R, Wang Y (2012) 3D Phase field simulation of effect of interfacial energy anisotropy on sideplate growth in Ti-6Al-4V. Acta Metall Sin 48:148

Yang R (2015a) Advances and challenges of tIaL base alloys. Acta Metall Sin 51:129

Yang XW, Li WY, Ma TJ (2015b) Computational analysis of linear friction welding process and micromechanical modeling of deformation behavior for medium carbon steel. J Cent South Univ 22:3275

Yang Z, Sista S, Elmer JW, DebRoy T (2000) Three dimensional Monte Carlo simulation of grain growth during GTA welding of titanium. Acta Mater 48:4813

Yu DJ, Bai CG, Xu DS, Li WQ, He SL, Lei JF, Liu YY, Yang R (2010) FEM simulation of forging of Ti alloy. Chin J Nonferrous Metals 20:s500

Yu DJ, Teng CY, Xu DS, Bai CG, Wei GZ, Yang R (2012) FEM simulation of hot compression of TiAl on Gleeble simulator. Proceedings of the 12th world conference on titanium, Science Press, Beijing, 1549

Yu DJ, Xu DS, Wang H, Zhao ZB, Wei GZ, Yang R (2019) Refining constitutive relation by combination of finite element simulations and Gleeble experiments. J Mater Sci Technol 34. https://doi.org/10.1016/j.jmst.2018.12.026

Yu KO (2013) Advancing titanium by continuous casting, IUMRS-ICAM. Qingdao

Yu Q, Qi L, Tsuru T, Traylor R, Rugg D, Morris JW, Asta M, Chrzan DC, Minor AM (2015) Origin of dramatic oxygen solute strengthening effect in titanium. Science 347:635

Zeng Z, Li X, Xu D, Lu L, Gao H, Zhu T (2016) Gradient plasticity in gradient nano-grained metals. Extreme Mech Lett 8:213

Zhang L, Rollett AD, Bartel T, Wu D, Lusk MT (2012) A calibrated Monte Carlo approach to quantify the impacts of misorientation and different driving forces on texture development. Acta Mater 60:1201

Zhang L, Zhang H, Ren X, Eckert J, Wang Y, Zhu Z, Gemming T, Pauly S (2018) Amorphous martensite in β-Ti alloys. Nat Commun 9:506

Zhao J, Lv L, Liu G (2017) Experimental and simulated analysis of texture evolution of TA15 titanium alloy sheet during hot tensile deformation at 750°C. Procedia Eng 207:2179

Zhao JC (2014) High-throughput experimental tools for the materials genome initiative. Chin Sci Bull 59:1652

Zhao YK, Ma HM, Zhang L, Bai CG, Xu DS, Ai HJ (2006) FEM analysis of effect of elastic modulus of artificial tooth to the stress concentration in the jawbone. West China J Stomatol 24:89

Zheng YF, Williams REA, Sosa JM, Alam T, Wang YZ, Banerjee R, Fraser HL (2016a) The indirect influence of the omega phase on the degree of refinement of distributions of the alpha phase in metastable beta-Titanium alloys. Acta Mater 103:165

Zheng YF, Williams REA, Wang D, Shi RP, Nag S, Kami P, Sosa JM, Banerjee R, Wang YZ, Fraser HL (2016b) Role of omega phase in the formation of extremely refined intragranular a precipitates in metastable ss-titanium alloys. Acta Mater 103:850

Zheng YF, Banerjee D, Fraser HL (2016c) A nano-scale instability in the beta phase of dilute Ti-Mo alloys. Scr Mater 116:131

Zhou XM, Zhang HO, Wang GL, Bai XW, Fu YH, Zhao JY (2016) Simulation of microstructure evolution during hybrid deposition and micro-rolling process. J Mater Sci 51:6735

Zhou G, Ye LH, Wang H, Xu DS, Meng CG, Yang R (2018) A First-Principles Study on Basal/Prismatic Reorientation-Induced Twinning Path and Alloying Effect. Acta Metall Sin 54:603

Zhu J, Wu H, Wang D, Gao Y, Wang H, Hao Y, Yang R, Zhang T-Y, Wang Y (2017) Crystallographic analysis and phase field simulation of transformation plasticity in a multifunctional β-Ti alloy. Int J Plast 89:110

Zong H, Lookman T, Ding X, Luo S-N, Sun J (2014) Anisotropic shock response of titanium: Reorientation and transformation mechanisms. Acta Mater 65:10

Quantitative Rheological Model for Granular Materials: The Importance of Particle Size

8

Ken Kamrin

Contents

Abstract

One of the great puzzles in describing the constitutive behavior of flowing granular media is the role of grain size. Even for relatively simple, stiff, round grains, the length scale of grains themselves is large enough to influence the continuum rheology, causing departures from local model predictions in the presence of flow heterogeneity and thin geometries. Here we present the development of a particular nonlocal model, the nonlocal granular fluidity (NGF) model, which has shown to be quantitatively predictive and able to capture this effect in the various forms it emerges. We show that the model correctly gives steady flow fields of grains over multiple orders of magnitude in many different 2D and 3D geometries without refitting. The same model captures the

K. Kamrin (✉)
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: kkamrin@mit.edu

size-dependent strengthening of thin granular bodies, a phenomenon that may at first glance appear separate from that of predicting steady flow profiles. NGF also correctly captures the so-called secondary rheology of grains, wherein faraway flow alters the apparent yield stress everywhere. The underlying physics of the nonlocal rheology is gradually becoming clearer, as it has recently been shown that the nonlocal fluidity field, which is a key field variable in the model, is a true kinematic state variable. We also discuss the remaining open questions and future directions.

1   Introduction

The problem of determining accurate constitutive relations for granular media is a challenging and long-lasting problem going back to the days of Coulomb. Numerous models have been proposed – far too many to attempt to summarize here. See Andreotti et al. (2013), Mitchell et al. (2005), and Goddard (2014) for a rather current review of models. The civil engineering community has traditionally modeled granular media using principles of continuum solid mechanics, which emphasizes the role of density, evolving internal variables, and pressure on the plastic yield condition. This modeling approach tends to focus on the behavior of natural grains, typically soils, undergoing moderate levels of total strain. On the other hand, one can consider problems of developed granular flow, where strains are enormous and one may consider something closer to a fluid-like treatment of the material. This viewpoint has been studied in recent decades by physicists and fluid mechanicians from both discrete-particle and continuum perspectives. Here, even with idealized simple grains, the issue of determining a precise "Navier–Stokes Equations" of granular flow has been elusive.

This chapter describes recent developments geared more toward the latter perspective, with the aim of coming up with a geometrically general, quantitatively accurate model that predicts the rheological behavior of systems of simple particles. The emphasis will be on developed flow problems and conditions related to existence of developed flows. The chapter proceeds for the most part chronologically in the order the developments were made. We begin by describing the state of the art as of the early 2000s, in particular the *inertial rheology* of granular flow. This model is fully local, and we discuss a number of key granular phenomena that cannot be represented under this approach, which leads to well-known accuracy losses. We then describe the reasoning behind adding a particular nonlocal term into the constitutive relation to rectify these shortcomings of the local model. The resulting model, which is denoted *nonlocal granular fluidity* (NGF), is then tested in a variety of flow geometries and shown to produce accurate predictions over several order of magnitude of flow speed, in both 2D and 3D. The same model is then tested against other unusual manifestations of grain-size cooperativity known in the granular materials literature, including the effect known as *secondary rheology* and the so-called H_{stop} effect. The latter demonstration is accompanied by a theoretical description of how the steady form of NGF is derived from an appropriate dynamic

form. We close with ongoing efforts to explain the microscopic underpinnings of NGF, including studies to relate its parameters back to grain-level descriptors, and results that show the model's key order parameter field, the *granular fluidity*, is in fact well-defined from microscopically observable grain behaviors.

2   Local Model

Over the last 15 or so years, significant progress has been made in determining the rheology of isotropically shaped, stiff grains in simple shearing (da Cruz et al. 2005; Jop et al. 2006). The model that has arisen, known as the "inertial rheology," could be seen as a generalization of Bagnold's scaling law (Bagnold 1954) for constant volume shearing. The model has been applied to a variety of problems, including silo drainage (Staron et al. 2012a, b; Kamrin 2010; Dunatunga and Kamrin 2015), heap flows (Jop et al. 2006, 2007; Kamrin 2010), granular column collapse (Lagrée et al. 2011; Dunatunga and Kamrin 2015; Valette et al. 2017), and projectile impact into dense granular media (Dunatunga and Kamrin 2017). A major feature of the inertial rheology is that its form can be deduced primarily from dimensional considerations, as we review now.

Consider a (planar) simple shear cell, containing quasi-monodisperse grains having characteristic size d, density ρ_s, surface friction μ_{surf}, restitution coefficient e, and "shape" s. The grains are presumed stiff enough that the stiffness k is not relevant to the rheology. Suppose a wall pressure P is imposed on walls of the cell and a tangential wall velocity is prescribed such that, at steady state, the granular strain rate is $\dot{\gamma}$. From these input conditions, a shear stress τ arises as the stress needed to sustain the steady motion. The system can be described by the following dimensionless groups:

$$\mu = \tau/P, \quad I = \dot{\gamma}\sqrt{\frac{\rho_s d^2}{P}}, \quad \mu_{\text{surf}}, \quad e, \quad s \tag{1}$$

where μ is the *stress ratio* and I is the *inertial number*. Thus, by Buckingham's theorem, the rheology of the grains must be expressible as

$$\mu = \hat{\mu}(I, \mu_{\text{surf}}, e, s). \tag{2}$$

For a fixed material, we can absorb the grain properties and tersely re-express the relation above as

$$\mu = \hat{\mu}(I). \tag{3}$$

This functional relationship is known as the *inertial rheology*. For $\mu_{\text{surf}} > 0$, it commonly fits well to a form

$$\mu = \mu_s + \frac{\Delta\mu}{\Delta\mu/bI + 1} \tag{4}$$

where μ_s, b, and $\Delta\mu$ are material constants. It is approximately linear in I when $I \ll 1$ but asymptotes to a value $\mu \to \mu_s + \Delta\mu$ as $I \to \infty$. It is also common, when interested in slow to moderate flows, to expand the above to linear order near $I \sim 0$ and use

$$\mu = \mu_s + bI. \tag{5}$$

If $\mu < \mu_s$ there is no flow. Hence, μ_s acts as a critical internal (bulk) friction coefficient. On the other hand, the more the applied stress ratio μ exceeds μ_s, the higher the relative flow rate I.

3   Departures from Locality

Although the inertial rheology is quite successful at describing homogeneous, steady-state simple shear flows in systems large compared to d, it was noticed fairly soon after its introduction that the relationship is violated in a number of important situations. In general, the $\hat{\mu}(I)$ relation is not satisfied in inhomogeneous steady flow geometries, particularly in regions of slowly flowing material (MiDi 2004). Moreover, a variety of other, seemingly independent phenomena stand in contrast to the outcomes that would be given by $\hat{\mu}(I)$. For example, it is observed that as geometries shrink in size, granular materials act as if they are stronger, i.e., a larger than expected stress ratio μ is needed to maintain a steady flow. This occurs in thin flows down inclined chutes (Pouliquen 1999), hourglass nozzles (Beverloo et al. 1961), and other geometries (Liu and Henann 2018) whose size scale impinges on the size scale of $< O(100)$ grains. As another unusual phenomenon, a "secondary rheology" is observed in granular flows (Reddy et al. 2011), wherein the motion of a boundary causes even the static looking parts of a domain far from the said boundary to lose their yield stress and flow under arbitrarily small forcing of a probe.

All of these behaviors are intrinsically nonlocal; the particle size is producing cooperativity in the dynamics over a grain-based length scale, and this causes the rheology at a point to divert from a relationship between local stress and local strain rate, as would be implied by Eq. 3. Whatever the remedy is, it involves explicitly accounting for the mean grain size as a length scale within the rheology. Previous size-dependent modeling approaches exist, such as Cosserat (micropolar) continua (Mohan et al. 2002), kinetic theory (Jenkins and Berzi 2010), self-activation (Pouliquen and Forterre 2009), partial fluidization (Aranson and Tsimring 2001), nonlocal inertial-number modeling (Bouzid et al. 2013) and the void/spot models (Kamrin and Bazant 2007). The key advantages of the theory we present below are its relative simplicity, as it depends on few parameters, and its ability to accurately capture the different departures from local mechanics exhibited by grains.

4   Nonlocal Granular Fluidity Model

The sieve that leads to the nonlocal model presented here lies in a set of data gathered by Koval and co-workers in 2009, which gives very clear indications as to how the local inertial relation is violated in inhomogeneous flow geometries (Koval et al. 2009). The authors use the discrete element method (DEM) to compute steady flows of quasi-monodisperse ($\sim$20% polydispersity) frictional disks in an annular Couette geometry. The Couette's exterior wall pressure is controlled but cannot displace in the θ direction, and its inner wall is rigid with rotational velocity prescribed (see Fig. 1a). All walls are fully rough. The output flow is characterized by a large speed near the inner wall that decays rapidly as r increases. By varying the grain size, as measured through R/d, and inner wall speed v_{wall}, many different flows are obtained. By coarse-graining stress and velocity in space and sampling many different locations, a scatter plot of μ and I is obtained for each test, and how well μ vs. I remains well-collapsed onto one curve can be tested. Since locally the flow in this geometry looks everywhere like steady simple shear, any departure from the one-to-one μ vs. I relation observed in homogeneous shear can *only* be explained through nonlocal effects that call upon higher flow gradients.

As shown in Fig. 1, for low values of I, which correspond to positions far from the inner wall, the plots from different tests do not converge onto one curve. In fact, as long as $R/d < \infty$, we see many data points where flow is occurring, $I > 0$, but the stress level is below what the local model would predict is the yield condition, i.e., $\mu < \mu_s$. In the tests where wall speed is varied, those tests where the wall speed is very slow all have the same range of μ present in the flow field, however the different slow wall speeds each have entirely different ranges of I. Here, the stress fields are nearly unchanged between the different tests, however the flow fields are all scaled by the speed of the wall. This rate-independent phenomenon in slow creeping zones, which runs directly counter to the rate-dependent nature of the inertial rheology, must be reconciled with nonlocal modeling.

The approach we take herein is a granular extension of a mathematical model proposed in 2007 for emulsions called the kinetic elastoplastic (KEP) model (Goyon et al. 2008; Bocquet et al. 2009). Though the KEP model was originally justified with micro-statistical arguments, our motivation for modifying the KEP model for granular media is primarily one of practicality, as its mathematical form possesses features that at the outset appear uniquely capable of rectifying the basic departures from locality apparent in the Koval data, yet the model remains fairly simple. We present the model now, which we refer to as the *nonlocal granular fluidity* (NGF) model.

Define the granular fluidity field g as an order parameter that exists throughout the granular domain. The fluidity plays a key role in the constitutive relation as it is the field that specifies how local loading causes local flow, via

$$\dot{\gamma} = g\mu. \tag{6}$$
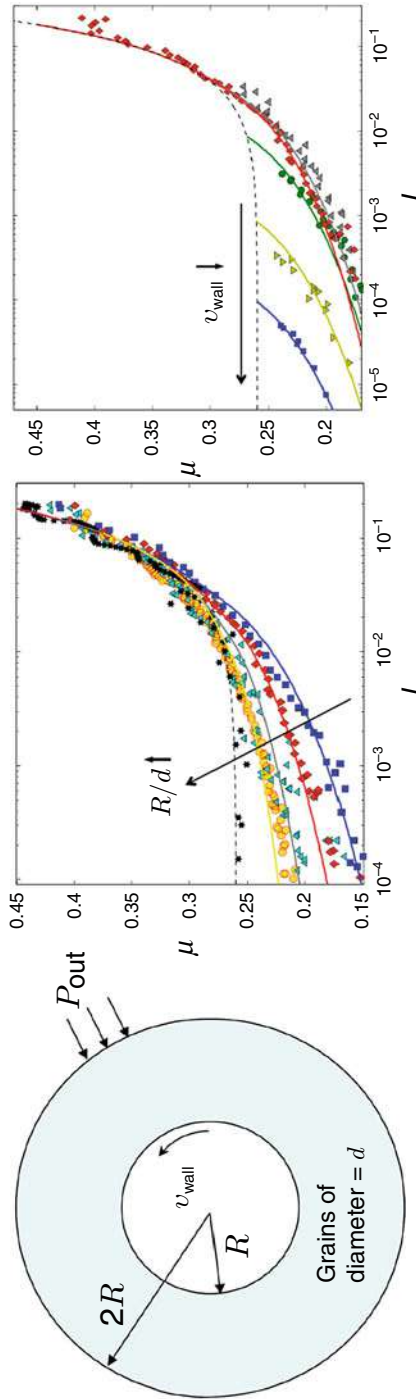
Fig. 1 (Left) The annular Couette flow setup. (Middle) Locus of μ vs. I in many steady DEM flows (symbols) each simulation having different R/d but all having same v_{wall}. (Right) Locus of μ vs. I for DEM flows with fixed R/d but various V_{wall}. Dotted line is local rheology prediction. Colored lines are predictions of NGF. (Images modified from Kamrin and Koval 2012 and Henann and Kamrin 2014a)

Larger g means more susceptibility to flow, whereas $g = 0$ represents solid material. If the inertial rheology were correct, g would have a formula, denoted g_{loc}, given by Eq. 5 in the slow flow regime:

$$g_{\text{loc}}(\mu, P) = \hat{\mu}^{-1}(\mu, P)/\mu = \frac{\langle \mu - \mu_s \rangle}{b\mu} \sqrt{\frac{P}{\rho_s d^2}} \qquad (7)$$

where $\langle \cdot \rangle$ denotes the Macaulay bracket. The proposed nonlocal rheology in its steady-state form states the true fluidity g is the solution to the following PDE:

$$g = g_{\text{loc}}(\mu, P) + \xi^2 \nabla^2 g. \qquad (8)$$

Equations 6 and 8 constitute the NGF constitutive model in steady, scalar form. Here, ξ is a length scale explicitly scaled by the mean grain size d. It can be seen in the above that the true fluidity is obtained as the solution of a PDE, for which g_{loc} serves as a source term. The true fluidity field, g, is a diffused version the field of g_{loc}. This allows regions to flow ($g > 0$) even if those regions have $\mu < \mu_s$. Secondly, in regions where $\mu < \mu_s$, g_{loc} vanishes and the PDE reduces to $g = \xi^2 \nabla^2 g$, which is linear in g and allows a solution to be scaled by a constant and still be a solution. This fact implies the system transitions to a rate-independent form in creeping zones whose stress is under the local yield criterion. If the flow is homogeneous or the length scale ξ vanishes, the $\nabla^2 g$ term vanishes, and one is left with the inertial rheology. All these checks pass the sieve of Koval's data.

The length scale above is given by $\xi = d\frac{A}{\sqrt{|\mu - \mu_s|}}$ as inspired by the form in Bocquet et al. (2009). It turns out this formula can be derived by considering the dynamic version of the fluidity PDE, which will be considered in an upcoming section. The divergence of the length scale as $\mu \to \mu_s$ was directly observed in DEM simulations in Kamrin and Koval (2012). The parameter A is the only new material parameter in the model. It is called the *nonlocal amplitude*; it is dimensionless and order one in size and depends only on grain information excluding grain size. Lastly, the system is closed by applying a set of boundary conditions to the new field. Since the fluidity PDE is of source-diffusion type, the choice of boundary conditions is less important than in boundary-only-sourced PDEs. In the 2D demonstrations provided next, $g = g_{\text{loc}}$ is applied at all boundaries.

The solid lines in Fig. 1 display the solutions of the NGF model in the annular Couette geometry, using $A = 0.8$. The parameters μ_s and b are taken from simple shear tests of the same DEM disks in Koval et al. (2009). It is observed that the NGF model correctly captures all data points from the DEM simulations. Contrary to the local model, NGF captures all the low-I data, which does not converge on a single μ vs. I curve.

5   Steady Flows

In light of the rheological check above, we conducted robustness tests to see how well this same model predicts motion of the same material in different geometries. Using the DEM disk material from before, we conducted tests of flow down a vertical chute and shear flow in the presence of gravity (Kamrin and Koval 2012). In both geometries, the gravity G is varied through several values. Results are shown in Fig. 2. It can be seen that the NGF model predicts the steady flow field correctly over several orders of magnitude without any model recalibration. Of note it captures the exponential-type decay of the fields and correctly finds that there are *no* regions where the flow rate is zero; however small it may get, it is nonvanishing *everywhere* as long as flow exists *somewhere*. This is in contrast to the inertial rheology which predicts sharp flow cutoffs.
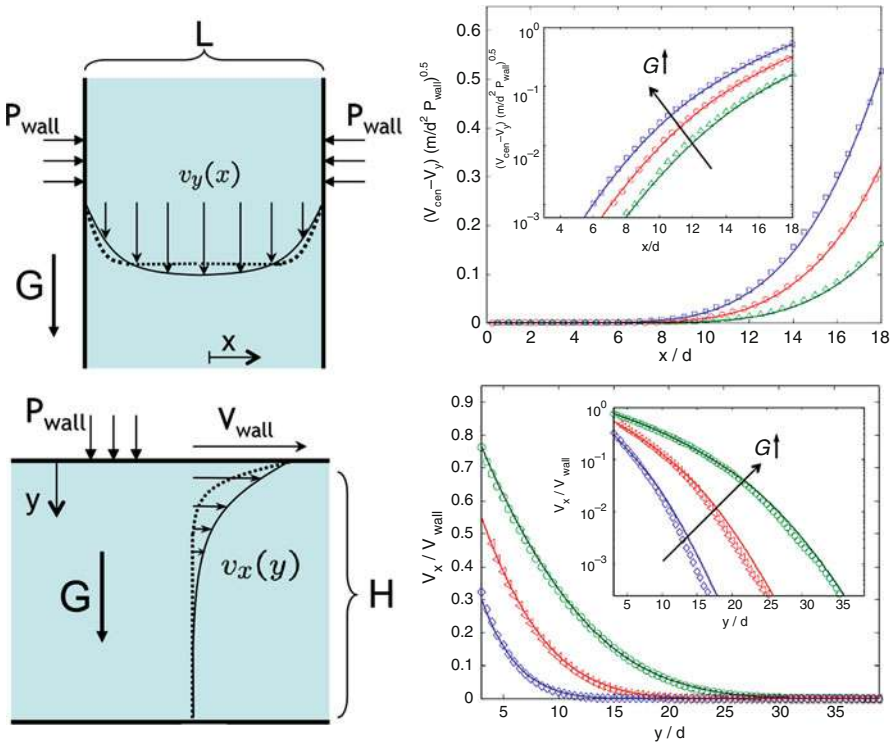
Fig. 2 (Top) (left) Chute flow geometry (gravity G indicated). Dotted line shows a cartoon of local model prediction with sharp flow-rate cutoffs at particular locations, unlike typical observations (solid line). (right) Direct flow field comparisons (measured relative to velocity at $x = 0$). Symbols are DEM solutions; lines are corresponding solutions to NGF model. In axis, m is the mean particle mass. Inset in semilog to highlight slow flowing areas. (Bottom) Same but for shearing with gravity present. (Images modified from Kamrin and Koval 2012; Henann and Kamrin 2013, 2014a, and Zhang and Kamrin 2017).

With verification that the model correctly predicts 2D flows in discrete simulations, we move on to show the same model predicts 3D flows in real experiments. We extend the scalar form of the model to a general form by presuming codirectionality of the stress deviator and the strain-rate tensor, which is a relatively good approximation in developed flows of simple grains (Koval et al. 2009; Rycroft et al. 2009; Weinhart et al. 2012; Henann and Kamrin 2013). If we also presume the rate of change of free volume has vanished at steady state, we obtain the following system:

$$\boldsymbol{\sigma} = -P\mathbf{1} + 2\frac{P}{g}\mathbf{D} \qquad \text{stress constitutive relation} \qquad (9)$$

$$\mathrm{tr}(\mathbf{D}) = 0 \qquad \text{constant-volume constraint} \qquad (10)$$

$$g = g_{\mathrm{loc}}(\mu, P) + \xi^2 \nabla^2 g \qquad \text{nonlocal rheology} \qquad (11)$$

where $\mathbf{D}$ is the deformation-rate tensor, $\boldsymbol{\sigma}$ is the Cauchy stress, and μ and P are now defined from stress invariants: $P = -\mathrm{tr}\boldsymbol{\sigma}/3$, $\tau = \sqrt{\sigma' : \sigma'/2}$, and $\mu = \tau/P$. Here, $'$ denotes the deviator. The function for the length scale ξ carries the same definition as before, though now its arguments are defined from these stress invariants. Coupling the above to the standard equations of momentum balance closes the system.

To test the 3D model, we simulate flows in a "split-bottom Couette" cell (Fenistein and van Hecke 2003). This geometry, pictured in Fig. 3, was chosen primarily because no existing continuum model had correctly predicted its flow fields; it is a notorious benchmark case (Dijksman and van Hecke 2010). The geometry is essentially an annular trough whose floor is split along a circle. Grains fill the trough and are set into motion by rotating the outer part of the trough with respect to the inner. The tantalizing observation in this geometry is that the flow field on the free surface is amazingly close to an error function as long as the fill height is relatively shallow. That is, the surface flow field $\omega = v_\theta / r$ fits almost exactly to

$$\omega(\lambda) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}(\lambda), \quad \lambda = \frac{r - R_c}{W} \qquad (12)$$

where R_c is the shear-zone center and W is the shear-zone width. It is observed that as filling height H increases, R_c shrinks and W grows, but the error-function shape is robust.

To simulate the NGF model in 3D, a user element routine in Abaqus was written (Henann and Kamrin 2016), with Neumann boundary conditions on the g field. As for material parameters, we consider only experiments that used glass beads and surmise a value of $A = 0.48$ this material. The other material constants (μ_s and b) are taken directly from the literature on simple shear of glass beads (Jop et al. 2005). As can be seen in Fig. 3, the NGF model captures the experimental flow data very well. It concurs in the way that smaller fill heights H produce thinner shear zones on the free surface and the shear-band center moves inward as H increases. Figure 3 also shows data from 20 simulated flows, varying H and d in the NGF
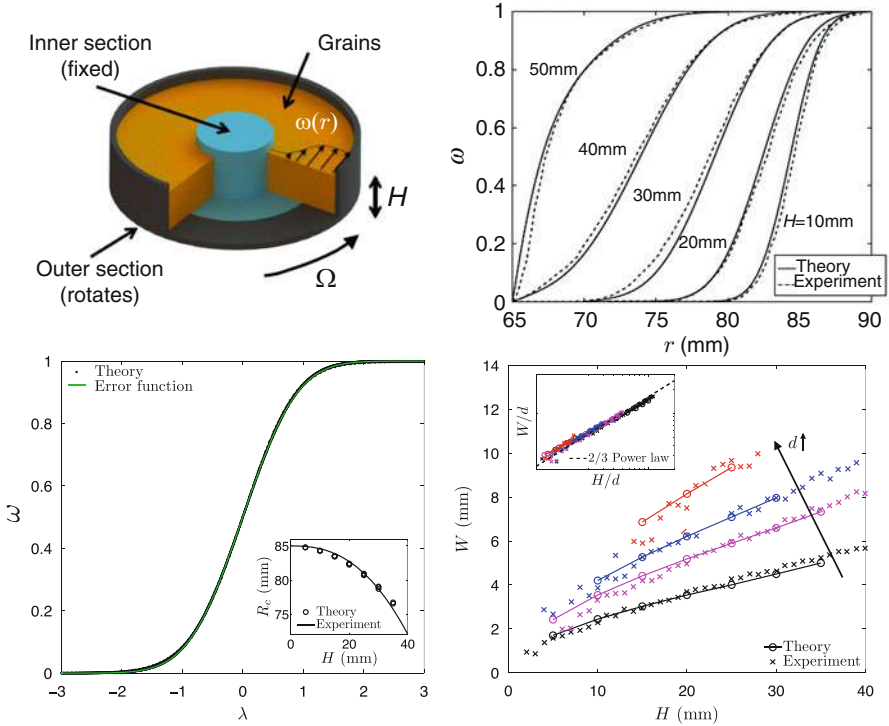
Fig. 3 (Top left) The split-bottom Couette geometry and free-surface flow field $\omega(r)$. (Top right) Comparison of the flow fields of 0.35 mm glass beads to solutions of a single calibration of NGF. (Bottom left) Twenty combinations of H and d reveal a universal error-function shape predicted by NGF (dust of black dots). Inset shows model captures shear-zone center as well. (Bottom right) Model captures correct interrelationship between W, H, and d compared to over 150 experimental measurements. (Images modified from Kamrin and Koval 2012; Henann and Kamrin 2013, 2014a, and Zhang and Kamrin 2017)

model, within the shallow range. When the output data is rescaled in terms of R_c and W, we see that in fact every flow curve is an almost exact error function.

We note that the split-bottom geometry brings about a complex interrelationship between three important lengths: the grain size d, the fill height H, and the shear-zone width at the free surface W. A stringent test of the NGF model is to check if it correctly captures this interrelationship. Figure 3, bottom right, shows data from over 150 experiments of glass beads in a split-bottom cell, varying d and H and outputting the resulting W (Fenistein et al. 2004). The NGF model reproduces the behavior very well, showing that it captures the subtle way in which the various lengths in the problem affect the length scales of the flow field. While this process of flow going from a Heaviside solution on the floor to an error function at the free surface may resemble "diffusion," experiments reveal the power law between H and W is closer to 2/3 than a diffusive 1/2. The NGF model concurs that the power law is about 2/3 (see inset).
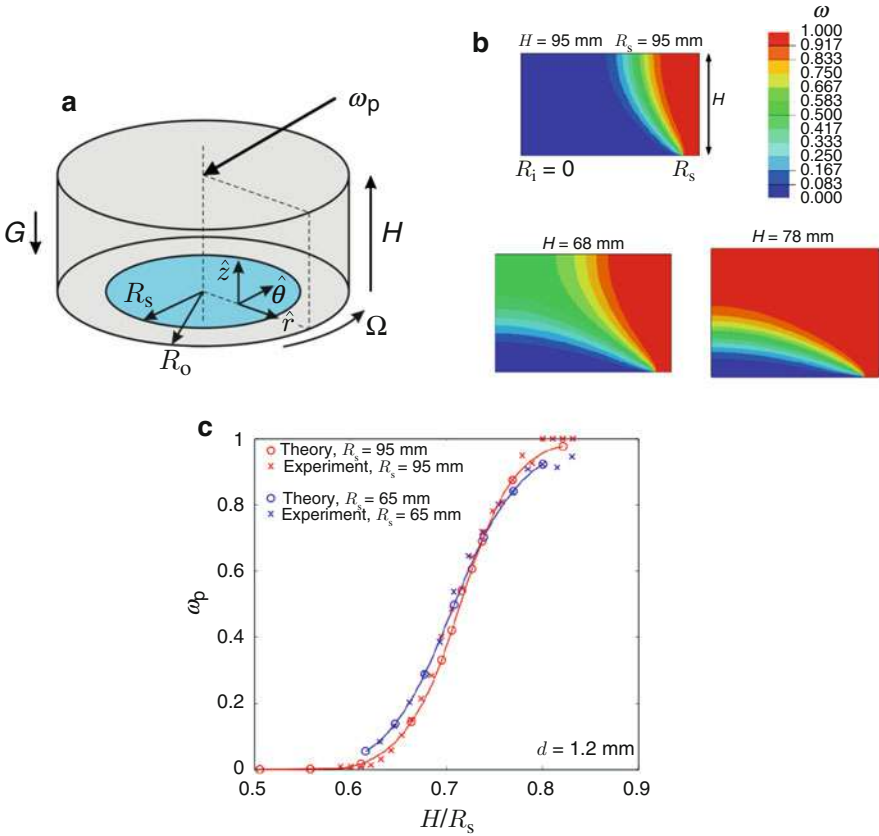
Fig. 4 (**a**) Schematic of the modified split-bottom cell. (**b**) Finite-element results of the NGF model in the rz-plane. Revolution rate ω shown. (**c**) Experiment (Fenistein et al. 2006) vs. NGF prediction for ω_P for two different modified split cells. (Images modified from Henann and Kamrin 2013, 2014a)

After the initial experimental results in the split-bottom Couette cell, modified split-bottom experiments were conducted where the inner wall was completely removed (see Fig. 4a) so that deeper layers could be studied without the flow field being interrupted by the inner wall (Fenistein et al. 2006). A wide transition in the flow behavior occurs as H increases – when H is still rather small, the flow looks like that of a standard split-bottom cell (Fig. 4b, top) but transitions to a flow in which a dome of grains is essentially "stuck" on the stationary platform, and the other material moves around it (Fig. 4b, right). The transition can be characterized by the normalized rotation rate, ω_P, of the grain located at the center of the free surface. Figure 4c shows experimental data of ω_P versus H for glass beads in two modified split cells, differing in the split radius R_c. The NGF model was then simulated in both these geometries, and the results are shown as well. The

predictions of the model match the experiments quite well. The fact that the model is able to sense correctly the small difference in the solution as the split radius varies is encouraging.

The same NGF model was solved in other geometries for which data on glass bead flows exist, including 3D annular shear cells, plate dragging with lateral gravity, and heap flows. The NGF model makes similarly quantitative predictions in these cases with no model recalibration (Henann and Kamrin 2013; Liu and Henann 2017).

6   Secondary Rheology

As noted in the introduction, not only is the particle length scale needed to correctly predict flow fields, but a number of unusual granular phenomena appear tied to the importance of grain size. Multiple experiments have established that when an intruder is placed in apparently quiescent material far away from zones of pronounced flow, it will move as long as *any* finite force is applied to it (Nichol et al. 2010; Reddy et al. 2011; Wandersman and Van Hecke 2014) . That is, flow *somewhere* effectively removes the yield stress *everywhere*, even in zones that appear virtually stationary. Conversely, if there is no existing "primary flow," a critical force must be exceeded to move an intruder through the grains. The term *secondary rheology* was coined to describe this effect, i.e., how the rheology of a probe changes due to the presence of a primary flow far away.

As a clear example, consider the setup shown in Fig. 5a, which was used in the experiments of Reddy et al. (2011). Here, an annular shear cell has a rod intruder dipped into the grains at some distance L from the inner wall, and a force F can be applied to it. When the inner wall is spinning, $\Omega > 0$, it creates a primary flow that is concentrated near the inner wall, and for much of the domain the grains are nearly static. When the wall is stationary, $\Omega = 0$, a critical force F_c must be applied
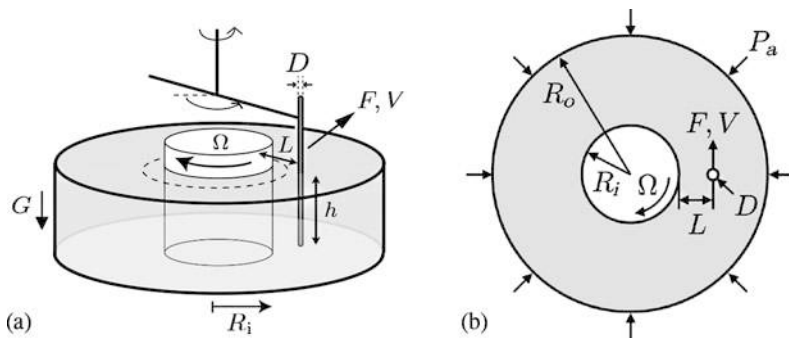
Fig. 5 (**a**) Schematic of the secondary rheology experiment of Reddy et al. (2011), (**b**) 2D version we consider. (Image from Henann and Kamrin 2014a)

to the rod to cause it to move. However, when the wall is moving, any force on the rod allows it to move; it translates with intermittent jerks about a robust mean speed. The mean motion of the intruder is observed to have several robust features: (i) it is rate-independent in that the speed of the rod is directly proportional to the speed of the moving wall for fixed L and F; (ii) at a fixed wall speed and L, the rod speed is exponentially dependent on the rod force (in the regime $F/F_c < 1$); (iii) the rod creeps faster when it is placed closer to the region of primary flow, i.e., smaller L. It is the fact that "close" in the third point is measured relative to grain size that we see this process as yet another manifestation of the grain-size effect, i.e., nonlocality.

In fact, through the equations of NGF, we can gain some insight into what is happening here. When the wall is stationary, the rod forcing alone must be large enough to create non-zero g_{loc} in the surrounding grains in order for non-zero fluidity to emerge in the granular material and permit rod motion. Because g_{loc} vanishes when stresses are too small, a finite critical force F_c emerges in order for rod motion to occur. However, when the inner wall is moving, the region adjacent to the inner wall has a non-zero g_{loc}, which now plays the role of a fluidity source in the NGF equation. The fluidity generated at the inner wall diffuses throughout the system, causing $g > 0$ everywhere including the region surrounding the rod. This permits rod motion under any force. The NGF PDE induces a decaying fluidity as one moves away from fluidity sources, which explains why larger L means slower rod speed, point (iii) above. The rate independence of the intruder, point (i), arises because, as previously discussed, the fluidity system reduces to a linear PDE in regions where g_{loc} vanishes. As long as the wall is slowly moving, the primary flow has $g_{loc} > 0$ only within a tiny ring about the inner wall, and the g_{loc} value is proportional to the wall speed. Hence, the fluidity field everywhere is scaled directly by the wall speed.

To check these ideas more directly, we simulated the NGF model in an analogous geometry in 2D for ease (Henann and Kamrin 2014a); see Fig. 5b. Here, instead of a rod, a rigid disk is pushed within a 2D annular cell of NGF material. As is evident in Fig. 6, all the signatures of secondary rheology emerge in solutions of the equations in this geometry. We observe the presence of a yield force F_c when the primary flow is off ($\Omega = 0$), but this critical force vanishes when the primary flow is on. We also see the rate independence of the rod motion when the primary flow is on, the exponential-type relationship relating intruder motion and intruder force, and diminishing of intruder speed as distance from the primary flow increases.

7   Fluidity Dynamics and Thin-Domain Strengthening

As we have noted throughout, the NGF PDE represented in Eq. 8 is for quasi-steady-state flow. As was shown in Henann and Kamrin (2014b), the steady-state system can be obtained as an approximate steady solution to a more primitive, dynamic form that describes the evolution of g through time. This dynamic form is justified thermomechanically with the aid of a virtual power argument. Like the steady-state
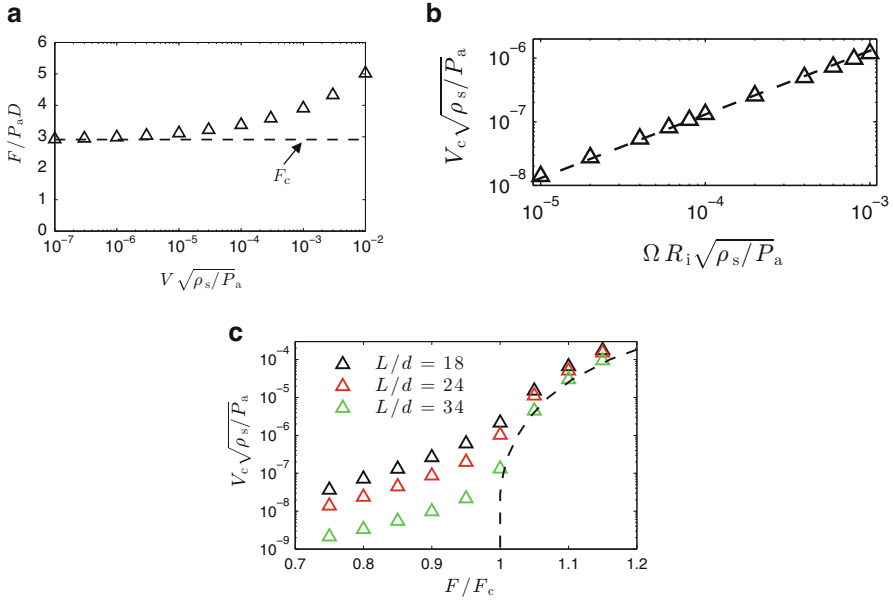
Fig. 6 NGF model results relating to secondary rheology. (**a**) When the inner wall is stationary, the force F on the rod must exceed a finite value F_c for any non-zero rod speed V to occur. (**b**) The rod's creep speed is proportional to the inner wall speed ΩR_i at fixed L and fixed $F < F_c$. (**c**) At fixed wall speed, an exponential-type behavior relating creep speed of the rod and rod force is observed when $F < F_c$. Rod motion decreases as L increases. Dashed line is the drag force data from part (**a**). (Images from Henann and Kamrin 2014a; Kamrin and Koval 2014)

form, the dynamic version of NGF is customized to the desired form of the local rheology; this is what the model will give in a uniform, steady, planar shear test. For example, if one desires Eq. 5 for the local response, the corresponding dynamic NGF equation is

$$t_0 \dot{g} = A^2 d^2 \nabla^2 g - (\mu_s - \mu)g - b\sqrt{\frac{\rho_s d^2}{P}}\mu g^2. \tag{13}$$

for some positive t_0, whose actual value is irrelevant to the upcoming studies. It is important to see how the above form relates to Eqs. 5 and 8 as it does not arise merely by setting the left-hand side to zero. The dynamic form above can have multiple steady solutions – note that $g = 0$ is always a solution. However, in flowing systems this is an *unstable* solution, and instead a solution with $g > 0$ emerges as time increases. So the steady-state form we have been discussing is in fact an approximate system satisfied only by the *stable* steady solutions of Eq. 13. It is arrived at as follows (more details in Henann and Kamrin (2014b)).

Let $g_{\text{hom}}(\mathbf{x}, t)$ be defined to satisfy Eq. 13 in the case of homogeneous flow. That is,

$$t_0 \dot{g}_{\text{hom}} = -(\mu_s - \mu)g_{\text{hom}} - b\sqrt{\frac{\rho_s d^2}{P}} \mu g_{\text{hom}}^2. \qquad (14)$$

Performing a fixed point analysis of this ODE reveals that when $\mu < \mu_s$, the only stable solution is $g_{\text{hom}} = 0$. In contrast, when $\mu > \mu_s$, a bifurcation occurs, and the only stable solution is $g_{\text{hom}} = \frac{\mu - \mu_s}{b\mu}\sqrt{\frac{P}{\rho_s d^2}}$. We can join these two solutions over the range of μ with the aid of the Macaulay bracket, $\langle \cdot \rangle$, giving the following general formula for the stable fixed points, g_{hom}^{ss}, of g_{hom}:

$$g_{\text{hom}}^{ss} = \frac{\langle \mu - \mu_s \rangle}{b\mu}\sqrt{\frac{P}{\rho_s d^2}}. \qquad (15)$$

We recognize this formula is precisely that of Eq. 5 for g_{loc}, which confirms that the intended local response emerges as the steady stable solution under homogeneous flow constraints.

Next, we approximate that in inhomogeneous cases at steady state, the fluidity is a small deviation away from $g_{\text{loc}} = g_{\text{hom}}^{ss}$. That is, $g = g_{\text{loc}} + \delta$ for some small δ. Wherever $\mu < \mu_s$, g_{loc} vanishes, so $g \approx \delta$ and up to first order in δ Eq. 13 becomes

$$0 = A^2 d^2 \nabla^2 \delta - (\mu_s - \mu)\delta \quad \rightarrow \quad 0 = A^2 d^2 \nabla^2 g - (\mu_s - \mu)g. \qquad (16)$$

Wherever $\mu > \mu_s$, to first order we have $g^2 = (g_{\text{loc}} + \delta)^2 \approx g_{\text{loc}}^2 + 2g_{\text{loc}}\delta = g_{\text{loc}}^2 + 2g_{\text{loc}}(g - g_{\text{loc}}) = -g_{\text{loc}}^2 + 2g_{\text{loc}}g$, and thus Eq. 13 becomes

$$0 = A^2 d^2 \nabla^2 g - (\mu_s - \mu)g - b\sqrt{\frac{\rho_s d^2}{P}} \mu(-g_{\text{loc}}^2 + 2g_{\text{loc}}g)$$

$$= A^2 d^2 \nabla^2 g - (\mu_s - \mu)g - (\mu_s - \mu)(g_{\text{loc}} - 2g)$$

$$= A^2 d^2 \nabla^2 g - (\mu - \mu_s)(g - g_{\text{loc}}). \qquad (17)$$

Equations 16 and 17 can be put into a single relation over the range of μ:

$$0 = A^2 d^2 \nabla^2 g - |\mu - \mu_s|(g - g_{\text{loc}}) \qquad (18)$$

which we recognize to be identical to the steady NGF PDE (8) with $\xi = d\frac{A}{\sqrt{|\mu - \mu_s|}}$ as claimed previously.

The approximations made in deriving the steady form, Eq. 8, from Eq. 13 are important to recognize. One interesting problem area for which the approximations do not hold is that of thin-body strengthening. In this case, bodies that are thin can experience μ significantly greater than μ_s and not flow, corresponding to a situation where g_{loc} is rather large but $g = 0$. This violates the small δ assumption of the

$g \approx g_{\text{loc}} + \delta$ approximation. This fact is of no concern, since we can instead use the more primitive, dynamic form to address these cases. We will do so now.

The specific problem we will consider is the well-known "H_{stop} effect" (Pouliquen 1999). A layer of grains of thickness H on a rough planar surface will flow if tilted high enough and stop if the tilt angle is reduced. Due to static determinacy, the stress distribution in a tilted layer is fairly straightforward to compute. For Φ the packing fraction, θ the tilt angle from horizontal, and G the acceleration of gravity, one finds $\mu = \tan\theta$ and $P = \rho_s \Phi G z \cos\theta$, where z measures depth orthogonally from the free surface. Because μ is dependent solely on tilt angle, local mechanics would suggest that H has no effect on the critical angle that stops a flowing layer. However, many experiments confirm this is not the case (MiDi 2004) even while confirming the stress distribution above and a near uniform Φ. In fact, thinner layers act stronger. That is, a flowing layer at a given tilt can be made to stop by decreasing its thickness, H. The critical arrest curve that determines the thickness at which flow stops is denoted $H_{\text{stop}}(\theta)$. It is directly proportional to the grain size, which indicates that this phenomenon is a consequence of nonlocal rheological effects. It could be said that what is happening here is that for thin layers, the entire system is experiencing cooperativity with the fixed bottom boundary, which promotes diminished motion in the granular system.

Because the H dependence of the flow threshold often permits stopped flow for large μ, approaching the maximum value of $\mu_2 = \mu_s + \Delta\mu$, for added precision we use the dynamic form whose local response is Eq. 4. It is

$$t_0 \dot{g} = A^2 d^2 \nabla^2 g - \Delta\mu \frac{\mu_s - \mu}{\mu_2 - \mu} g - b \sqrt{\frac{\rho_s d^2}{P}} \mu g^2. \tag{19}$$

Because NGF does not (yet) account for hysteresis effects, it is sufficient to consider what H and θ combinations render the no-flow, $g = 0$ solution, stable in this geometry. In these cases, a layer perturbed to flow would return to a stopped solution. The $H_{\text{stop}}(\theta)$ curve represents the phase separator between the stable and unstable states. We linearize the above PDE about $g = 0$, which amounts simply to neglecting the g^2 term, leaving us with

$$t_0 \dot{g} = A^2 d^2 \nabla^2 g - \Delta\mu \frac{\mu_s - \mu}{\mu_2 - \mu} g. \tag{20}$$

Since we are interested in thin layers, where most material is near a boundary, it is more important now than before to ensure we have a precise boundary condition for g. DEM evidence from Silbert et al. (2003) suggests that at the rough floor, $z = H$, the shear rate vanishes, and near the free surface, $z = 0$, the shear-rate gradient vanishes. Because μ is spatially constant, these observations, in light of Eq. 6, suggest the correct conditions are $g(z = H) = 0$ and $\frac{\partial g}{\partial z}\big|_{z=0} = 0$. The lowest mode solution to Eq. 20 under these boundary conditions is

$$g(z, t) = C \exp\left[\left(-\frac{\pi^2 A^2 d^2}{4H^2} + \Delta\mu \frac{\tan\theta - \mu_s}{\mu_2 - \tan\theta} \right) \frac{t}{t_0} \right] \cos\left(\frac{\pi z}{2H} \right). \tag{21}$$

for arbitrary C. Based on the sign of the term in the exponential, we see that a perturbed layer solidifies, i.e., returns to $g = 0$, if

$$H < \frac{\pi A d}{2} \sqrt{\frac{\mu_2 - \tan\theta}{\Delta\mu(\tan\theta - \mu_s)}} = H_{\text{stop}}(\theta). \tag{22}$$

We can compare this analytical solution using the *same* model parameters as before, plus a value of μ_2 for beads from the literature Jop et al. (2005). The result is shown in Fig. 7b. The model does an excellent job predicting the experimental flow stoppage curve (Kamrin and Henann 2015). The extent to which this curve is not a vertical straight line reflects the nontrivial size-dependent strengthening.
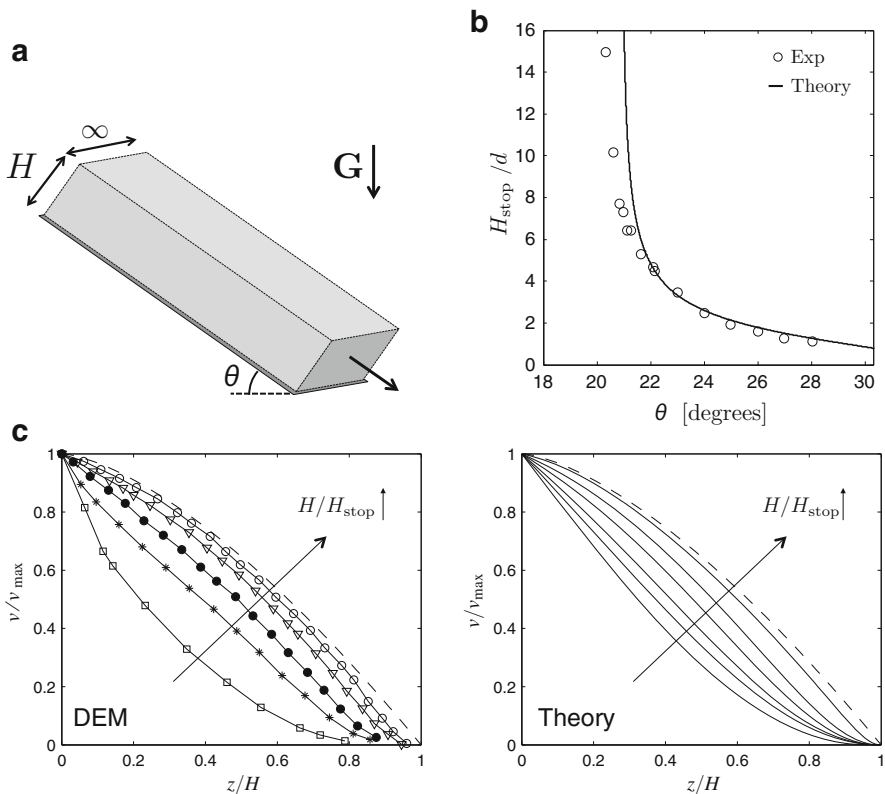
Fig. 7 (**a**) The inclined flow setup. (**b**) Theoretical vs. experimental H_{stop} curves for glass beads (Kamrin and Henann 2015; Pouliquen 1999). (**c**) The variation in flow profile shape for layers just above H_{stop} to thick layers at a common θ compared to DEM (Silbert et al. 2003). Dotted line is the solution from the local, inertial rheology, i.e., "Bagnold's profile"

Recent work in Liu and Henann (2018) has shown that the same dynamic NGF model correctly predicts grain-size-dependent flow thresholds in other geometries, including vertical chutes, shear flows with gravity, and annular shear flows.

We can also numerically run Eq. 19 to see what the shape of the flow profiles is for values of H above H_{stop}. Figure 7c shows a comparison of solutions to a sequence of DEM data sets from Silbert et al. (2003). It is notable that the model captures the concavity switch of shallow layers to deep ones that local modeling does not capture – the dotted line indicates the solution of the local law, Eq. 4, for any H.

8   Microscopic Physics of NGF Model

There exist a number of open questions on the physics underpinning the NGF model, relating to how the model connects back to the grain-scale dynamics. However, progress has been made in couple areas, which we discuss now.

8.1   Effect of Grain Information on Continuum Parameters

One issue is to seek how the NGF material parameters connect back to grain properties. There ought to be a direct correspondence between, say, the grain shape, surface friction, restitution, and stiffness and the resulting continuum model parameters, most importantly the nonlocal amplitude A.

In Kamrin and Koval (2014), a set of DEM simulations were used to determine the effect of surface friction, μ_{surf}, on the NGF continuum parameters. This was done using a sequence of DEM simulations where all other grain properties were fixed to the values used in Koval et al. (2009). The simulation produced flow fields in planar shear and in annular shear, the former is enough to determine g_{loc}, and the latter is used to determine A. Some results of the study are shown in Fig. 8. The grains used were relatively stiff disks with coefficient of restitution of 0.1% and 20% polydispersity.

Interestingly, the value of A does not appear to depend much on particle surface friction unless μ_{surf} is very small (<0.1). Over the entire range tested, A remains order one. These observations suggest the cooperativity distance of granular motion, which A essentially measures, is driven strongly by the geometry of the non-affine motion of particles densely traversing over other particles, more so than the details of the specific contact interaction.

8.2   Microscopic Meaning of the g Field

It is tempting, though not physically instructive, to state that g is defined by inverting the flow rule Eq. 6 to yield $g = \dot{\gamma}/\mu$. While this definition may be true, it confuses the *role* g plays in the flow rule with the *meaning* of g. Granted, one need not

Fig. 8 Dependence of local bulk friction coefficient μ_s and nonlocal amplitude A on particle surface friction. (Image from Kamrin and Koval 2014)
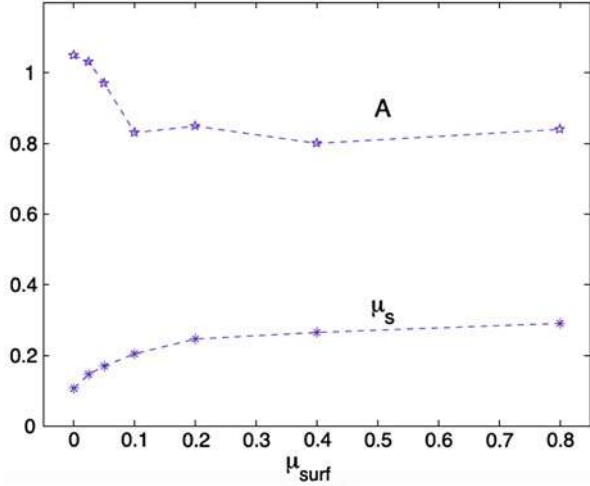
understand the meaning of g to use the NGF model as it is mathematically closed. However, a model that works ought to have a microphysical foundation, and we stipulate that in fact g is a real kinematically measurable field, which can be extracted from particle motions. The treatment of g as a kinematic state variable was implicitly presumed in the virtual power derivation of Eq. 13 (see Henann and Kamrin 2014b), where g and ∇g were both presumed to play the role of kinematic variables over which power can be expended. But this analysis does not elucidate the specific formula for g based on particle dynamics. In Zhang and Kamrin (2017), such a formula was obtained. We hypothesized that two inputs that should influence the flow susceptibility are the packing fraction Φ and the velocity fluctuations δv of the particles about their mean velocity. It can be checked if these are sufficient to describe g by first applying a dimensional argument to produce an appropriately general function of these variables that has units of granular fluidity, inverse time. One obtains

$$g = \frac{\delta v}{d} F(\Phi) \qquad (23)$$

where F is some unknown function. To see if this form describes the g field well, we conducted numerous 3D DEM simulations of a common granular material (spherical particles) in multiple inhomogeneous flow geometries. We extracted the coarse-grained field variables δv, $g = \dot{\gamma}/\mu$, and Φ when the flows reached steady state. Plotting $gd/\delta v$ versus Φ in each case yields Fig. 9. The 20 different flows tested over 3 distinct configurations all appear to collapse onto a universal curve for the F function, which is roughly constant at low Φ but approaches 0 as Φ approaches random close packing.

By fitting F from this plot, we can use Eq. 23 to see the fluidity field spatially whenever data on particle fluctuations and packing can be obtained. For example,

Fig. 9 Normalized fluidity $gd/\delta v$ and packing fraction Φ cross-plotted from 20 different flows in 3 different flow configurations, revealing a universal relationship. Blue line is a fit for $F(\Phi)$ used in Fig. 10. (Images from Zhang and Kamrin 2017)
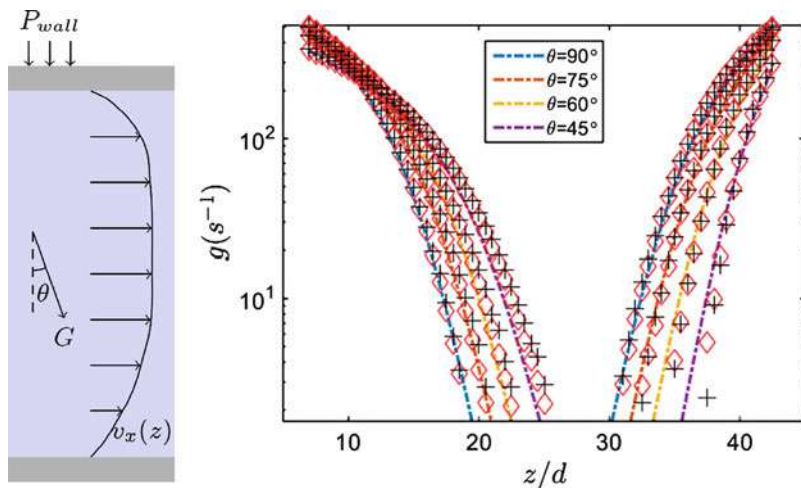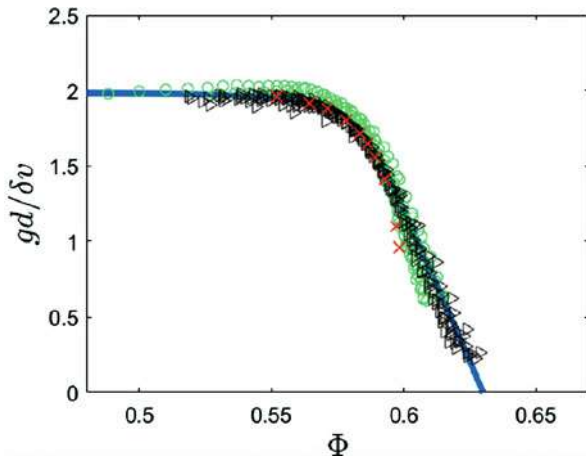
Fig. 10 Comparison of three ways to define the g field in the tilted incline chute geometry for various tilt angles with respect to gravity, G. DEM results of $\dot{\gamma}/\mu$ (diamonds), steady solutions of Eq. 19 (lines), and predictions of the microscopic formula, Eq. 23 (plusses). (Image from Zhang and Kamrin 2017)

Fig. 10 shows results for a tilted double-walled chute at four different tilt angles. DEM simulations were conducted, from which the inverted flow rule definition of the fluidity field $g = \dot{\gamma}/\mu$ can be extracted and plotted as a function of space. This was compared to two other ways to obtain the steady g field: by solving Eq. 19 to steady state, and by applying the new microscopic formula, Eq. 23. As can be seen in the figure, all three ways of obtaining the g field agree with each other over a large range of g.

With this understanding in tow, we can see that the NGF model is in fact a model that tracks the evolution of this conjoined microphysical field. Since granular temperature may be defined as $T \propto \delta v^2$, one might view the NGF PDE as some variant of a heat equation, though more study is needed to make this connection. The kinetic theory of granular flows (Jenkins and Savage 1983; Jenkins and Berzi 2010) may provide useful insights toward this aim; however, there is evidence that kinetic theory hypotheses are not all well-represented in our flow data. For example, even though $g = \dot{\gamma}/\mu = \frac{\delta v}{d} F(\Phi)$ is seen to hold generally across all flows, kinetic theory would suggest that other variables such as pressure P and viscosity η should collapse as $P = P(\Phi, T)$ and $\eta = \eta(\Phi, T)$. Our data does not show a strong collapse for these variables, except in dilute zones, where $\Phi < 0.57$. Of course, dilute flows have conditions more conducive to the basic collision-based hypotheses of kinetic theory. It appears that the g field is somewhat special in its ability to continually be represented in terms of δv and Φ even in the dense regime, across various inhomogeneous flows.

9   Conclusion and Future Work

The NGF model has shown the ability to describe granular flows of simple grains down to a fairly significant level of accuracy, over a variety of different cases. The model is notable for the range of effects it captures by adding, essentially, one new term to an existing local model, which brings about only one new material parameter, the dimensionless nonlocal amplitude A. Much of the future work on this approach breaks down into three areas: (i) going beyond dense, developed flows, (ii) going beyond simple grains, and (iii) understanding more microscale physics behind the model.

Regarding (i), there are current efforts to extend the model to permit a separated phase, so grains need not stay dense-packed. This model extension is needed to simulate the falling stream of grains in silo drainage, for example, and is needed to properly model size-dependent flow stoppage in nozzles where some grains fall away leaving an arched blockage. To go beyond the well-developed flow regime, it is important to recognize that even though Eq. 19 has dynamics, we only believe these dynamics to hold when the dynamics of other internal variables, like the contact statistics of the system, have steadied out. A model that allows for a full treatment of granular transients – including volume-strength coupling like in critical state models or the evolution of internal contact anisotropy variables like a fabric tensor – would need to be coupled to the NGF system. Also, while the value of t_0 in Eq. 19 was unimportant to computing the H_{stop} curve or steady flow profiles, capturing the correct time dependence of fluidity evolution would require a deeper study of t_0.

Regarding (ii), the most important additional effect to consider is polydispersity. The current model allows a small range of polydispersity; all experiments and DEM tests used herein have had some polydispersity, not exceeding 20%, and d has represented the *mean* particle size. The reason wider polydispersity has not yet been considered is because it tends to cause granular size segregation. Once

particle size is no longer uniformly distributed, *d* becomes a function of space, and thus a separate model is needed to track the spatial evolution of *d* before NGF can be applied to give the rheology. We also need to test the model with different grain shapes. As grains become more anisotropic in shape, it is possible the fluidity diffusion term becomes anisotropic based on mean grain orientation.

Regarding (iii), it remains important to continue the effort to derive the model in full from small-scale statistical principles. Even though NGF is thermomechanically justified and has a microscopically meaningful order parameter (Henann and Kamrin 2014b; Zhang and Kamrin 2017), we recall it was proposed primarily as a practical top-down constitutive form with potential to capture a number of puzzling observations. And indeed, the model has demonstrated accuracy in flow predictions, versatility across flow geometries, and the ability to capture other particle size effects beyond the prediction of flow fields. However, connecting to a bottom-up derivation is necessary to complete the picture.

References

Andreotti B, Forterre Y, Pouliquen O (2013) Granular media: between fluid and solid. Cambridge University Press, Cambridge

Aranson IS, Tsimring LS (2001) Continuum description of avalanches in granular media. Phys Rev E 64:020301

Bocquet L, Colin A, Ajdari A (2009) Kinetic theory of plastic flow in soft glassy materials. Phys Rev Lett 103(3):036001

Bagnold RA (1954) Experiments on a gravity free dispersion of large solid spheres in a newtonian fluid under shear. Proc R Soc Lond Ser A 225:9–63

Bouzid M, Trulsson M, Claudin P, Clément E, Andreotti B (2013) Nonlocal rheology of granular flows across yield conditions. Phys Rev Lett 111(23):238301

Beverloo WA, Leniger HA, de Velde JV (1961) The flow of granular solids through orifices. Chem Eng Sci 15:260

da Cruz F, Emam S, Prochnow M, Roux JN, Chevoir F (2005) Rheophysics of dense granular materials: discrete simulation of plane shear flows. Phys Rev E 72(2):021309

Dijksman JA, van Hecke M (2010) Granular flows in split-bottom geometries. Soft Matter 6(13):2901–2907

Dunatunga S, Kamrin K (2015) Continuum modelling and simulation of granular flows through their many phases. J Fluid Mech, Cambridge University Press 779:483–513

Dunatunga S, Kamrin K (2017) Continuum modeling of projectile impact and penetration in dry granular media. J Mech Phys Solids, Elsevier 100:45–60

Fenistein D, van Hecke M (2003) Wide shear zones in granular bulk flow. Nature 425:256

Fenistein D, van de Meent JW, van Hecke M (2004) Universal and wide shear zones in granular bulk flow. Phys Rev Lett 92(9):094301

Fenistein D, van de Meent JW, van Hecke M (2006) Core precession and global modes in granular bulk flow. Phys Rev Lett 96(11):118001

Goddard JD (2014) Continuum modeling of granular media. Appl Mech Rev 66(5):050801

Goyon J, Colin A, Ovarlez G, Ajdari A, Bocquet L (2008) Spatial cooperativity in soft glassy flows. Nature 454(7200):84–87

Henann DL, Kamrin K (2013) A predictive, size-dependent continuum model for dense granular flows. Proc Natl Acad Sci 110(17):6730–6735

Henann DL, Kamrin K (2014a) Continuum modeling of secondary rheology in dense granular materials. Phys Rev Lett 113(17):178001

Henann DL, Kamrin K (2014b) Continuum thermomechanics of the nonlocal granular rheology. Int J Plasticity 60:145–162

Henann DL, Kamrin K (2016) A finite element implementation of the nonlocal granular rheology. Int J Numer Methods Eng 108(4):273–302

Jenkins JT, Berzi D (2010) Dense inclined flows of inelastic spheres: tests of an extension of kinetic theory. Granul Matter 12(2):151–158

Jenkins JT, Savage SB (1983) A theory for the rapid flow of identical, smooth, nearly elastic particles. J Fluid Mech 130:187–202

Jop P, Forterre Y, Pouliquen O (2005) Crucial role of sidewalls in granular surface flows: consequences for the rheology. J Fluid Mech 541:167–192

Jop P, Forterre Y, Pouliquen O (2006) A constitutive law for dense granular flows. Nature 441:727

Jop P, Forterre Y, Pouliquen O (2007) Initiation of granular surface flows in a narrow channel. Phys Fluids 19(8):088102

Kamrin K, Bazant MZ (2007) Stochastic flow rule for granular materials. Phys Rev E 75:041301

Kamrin K (2010) Nonlinear elasto-plastic model for dense granular flow. Int J Plasticity, Elsevier 26(2):167–188

Kamrin K, Henann DL (2015) Nonlocal modeling of granular flows down inclines. Soft Matter 11(1):179–185

Kamrin K, Koval G (2012) Nonlocal constitutive relation for steady granular flow. Phys Rev Lett 108(17):178301

Kamrin K, Koval G (2014) Effect of particle surface friction on nonlocal constitutive behavior of flowing granular media. Comput Part Mech 1(2):169–176

Koval G, Roux JN, Corfdir A, Chevoir F (2009) Annular shear of cohesionless granular materials: from the inertial to quasistatic regime. Phys Rev E 79(2):021306

Lagrée P-Y, Staron L, Popinet S (2011) The granular column collapse as a continuum: validity of a two-dimensional Navier–Stokes model with a μ (I)-rheology. J Fluid Mech, Cambridge University Press 686:378–408

Liu D, Henann DL (2017) Non-local continuum modelling of steady, dense granular heap flows. J Fluid Mech 831:212–227

Liu D, Henann D (2018) Size-dependence of the flow threshold in dense granular materials. Soft Matter 14:5294–5305

MiDi G (2004) On dense granular flows. Eur Phys J E 14(4):341–365

Mitchell JK, Soga K et al (2005) Fundamentals of soil behavior, vol 3. Wiley, New York

Mohan LS, Rao KK, Nott PR (2002) A frictional Cosserat model for the slow shearing of granular materials. J Fluid Mech 457:377–409

Nichol K, Zanin A, Bastien R, Wandersman E, van Hecke M (2010) Flow-induced agitations create a granular fluid. Phys Rev Lett 104:078302. http://link.aps.org/doi/10.1103/PhysRevLett.104.078302

Pouliquen O (1999) Scaling laws in granular flows down rough inclined planes. Phys Fluids 11:542

Pouliquen O, Forterre Y (2009) A non-local rheology for dense granular flows. Philos Trans R Soc Lond A: Math Phys Eng Sci 367(1909):5091–5107

Reddy K, Forterre Y, Pouliquen O (2011) Evidence of mechanically activated processes in slow granular flows. Phys Rev Lett 106(10):108301

Rycroft CH, Kamrin K, Bazant MZ (2009) Assessing continuum postulates in simulations of granular flow. J Mech Phys Solids, Elsevier 57(5):828–839

Silbert LE, Landry JW, Grest GS (2003) Granular flow down a rough inclined plane: transition between thin and thick piles. Phys Fluids 15(1):1–10

Staron L, Lagrée P-Y, Popinet S (2012a) The granular silo as a continuum plastic flow: The hour-glass vs the clepsydra. Phys Fluids, AIP 24(10):103301

Staron L, Lagrée P-Y, Popinet S (2012b) Continuum simulation of the discharge of the granular silo. The European Physical Journal E, Springer 37(1):5

Valette R, Riber S, Hachem E (2017) Quantitative predictions of the μ (I) rheology in 2D and 3D granular column collapse: scaling laws and quasi-static vs. inertial regimes. EPJ Web Conf 140:11005

Wandersman E, Van Hecke M (2014) Nonlocal granular rheology: role of pressure and anisotropy. EPL (Europhys Lett) 105(2):24002

Weinhart T, Thornton AR, Luding S, Bokhove O (2012) From discrete particles to continuum fields near a boundary. Granul Matter. 14(2):289–294

Zhang Q, Kamrin K (2017) Microscopic description of the granular fluidity field in nonlocal flow modeling. Phys Rev Lett 118(5):058001

Mesoscale Mechanisms of Cement Hydration: BNG Model and Particle Simulations

9

Enrico Masoero

Contents

Abstract

Cement paste is the most widely used artificial material on Earth, causing massive CO_2 emissions. To address the problem, new chemical formulations are being explored, but these come with uncertainties around the kinetics of strength development and degradation. This kinetics is largely controlled by the precipitation of mesoporous solid phases from ionic aqueous solution. Modeling the precipitation process is a key step to understand and control the properties of future, more sustainable, cements. This manuscript starts by considering the classical boundary nucleation and growth (BNG) model of cement minerals precipitation. The mechanisms of nucleation and growth and the corresponding

E. Masoero (✉)
School of Engineering, Newcastle University, Newcastle upon Tyne, UK
e-mail: enrico.masoero@newcastle.ac.uk

parameters governing the model are discussed. The next step is to relate these mechanisms and parameters to the chemistry of the liquid and solid phases involved. This requires more fundamental models, and, to this end, nanoparticle simulations of cement mineral precipitation and morphology development are presented. These simulations are indeed shown to relate some input parameters of the BNG model to the chemistry of the paste. An outlook is finally provided, discussing some outstanding extensions of nanoparticle simulations that could deepen the current understanding of cement hydration.

1   Introduction

Hardened cement paste is the matrix of concrete, providing its binding ability and mechanical properties (Taylor 1997). With a yearly production of 2.5 billion tons, cement is the most widely used artificial material on Earth. This comes with important environmental concerns; in particular, for each ton of cement that is produced, one ton of CO_2 is released (notation: "cement" indicates the material before mixing with water; "cement paste" is cement mixed with water). On a global scale, this leads to an impressive 8% of anthropogenic CO_2 emissions being currently due to cement production alone (Olivier and Muntean 2014). Cement production is also expected to grow to more than 3.5 billion tons by 2050, which conflicts with the international target of drastically reducing greenhouse gas emissions by the same year (European Commission 2011; Scrivener et al. 2018).

Approximately 50% of cement-related CO_2 emissions are intrinsic to the chemical reactions taking place during the production of cement. This motivates research on new chemical compositions, different from today's ordinary cements (Juenger et al. 2011). New chemistries, however, entail uncertainties around controlling the temporal evolution of mechanical properties during cement hydration, viz., as the cement reacts chemically with water and the paste undergoes the "setting" transition from liquid to solid (Taylor 1997). There is hope that modeling and simulations could help predict and control the chemistry-dependent hydration kinetics of ordinary and new cements. Macroscale engineering models typically assume that the rate of hydration is a function of the microstructure of the paste, which evolves as the cement dissolves in water and new minerals precipitate. In such macroscale models, the relationship between rate and microstructure is summarized into "affinity" functions, which in principle depend on chemical composition, temperature, humidity, and other variables (Ulm and Coussy 1996). However, affinity functions are typically fitted to reproduce specific experimental results, thus not predictively linking cement chemistry with microstructural evolution and hydration rate.

This paper delves into the relationship between microstructure and hydration rate. In particular, the focus is on the first hours after mixing cement with water, when the structural features controlling hydration are at the micrometer length scale or below (Bullard et al. 2011). Section 2 provides a basic introduction to ordinary cement chemistry. Section 3 presents some typical experimental data on

the hydration rate of cement and on the concomitant evolution of pore structures at multiple length scales, from nano to macro. Section 4 presents the boundary nucleation and growth (BNG) model, widely used to relate the hydration rate to the evolving morphology of cement minerals precipitating from solution. The main assumptions and input parameters of the BNG model are discussed. Finally, Section 5 introduces nanoparticle simulations, which are a more recent and fundamental approach to modeling mineral precipitation and morphology development during cement hydration. Results from nanoparticle simulations show that the evolution of morphology and mechanical properties in a cement paste are manifestations of nanoscale processes driven by both chemical potentials and mechanical interactions. Opportunities and challenges for nanoparticle simulations to link the BNG model to the chemical composition and curing conditions of the cement paste are ultimately discussed.

2   Introduction to Ordinary Cement Chemistry

Ordinary cements are mostly made of calcium and aluminum silicates (Taylor 1997). They also contain significant amounts of other elements and impurities, leading to a very complex chemistry overall. This section will not address such complexity by contrast, only one key product of the calcium-silicate hydration is presented, alongside several chemical reactions that are necessary to contextualize the subsequent sections. Despite this simplification, the modeling techniques discussed in later sections will be generalizable to more complex chemical formulations.

The main mineral phase providing mechanical strength to ordinary cement pastes is calcium-silicate-hydrate, $(CaO)_x(SiO_2)(H_2O)_z$ (Grant et al. 2006), or simply C–S–H in cement oxide notation (where $C = CaO$, $S = SiO_2$, and $H = H_2O$). At the molecular level, water in C–S–H is partly bound as OH groups and partly adsorbed as highly confined H_2O molecules inside sub-nanometric spaces between calcium-silicate layers (Manzano et al. 2012). C–S–H is structurally heterogeneous and porous already at the scale of few nanometers; hence it is often modeled as a mesoporous agglomerate of 5–10 nm particles (Jennings 2008), leaving mesopores with characteristic width between 2 and 50 nm. These agglomerates are called "C–S–H gel." The x and z coefficients in the C–S–H formula indicate that the stoichiometry is not fixed, but it can change from one place to another within the same sample. This local composition is quantified by the calcium-to-silicon molar ratio, $x = Ca/Si$.

C–S–H forms by precipitation from aqueous ionic solution. A simple chemical reaction for this may be (Grant et al. 2006):

$$x\, Ca(OH)_2\, (aq) + H_2SiO_4\, (aq) + (z - x)\, H_2O = C{-}S{-}H + 2OH \quad (1)$$
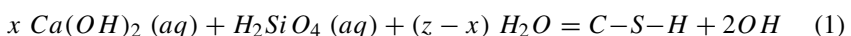
The solution is the result of the progressive dissolution of cement powder in water. In a typical ordinary cement, the main dissolving phase during the first weeks of

hydration is tricalcium silicate, Ca_3SiO_5, or simply C_3S in cement oxide notation (Grant et al. 2006):

$$C_3S + 3H_2O + 2OH = 3Ca(OH)_2\ (aq) + H_2SiO_4\ (aq) \qquad (2)$$
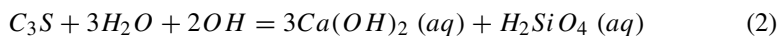
Producing C_3S is a CO_2 intensive process, mostly due to the thermal decomposition of calcium carbonate forming the CaO component of the C_3S. Reducing the reliance of cementitious materials on calcium oxide is possible, but this would lead to different reacting phases and thus alter the hydration reactions, the chemical composition of the solution, and the chemical composition and structure of the hydration product, including the C–S–H. All this explains the current interest in predicting how changes to the cement chemistry would affect the formation kinetics, structure, and mechanical properties of the hydration product and in particular of the C–S–H.

3 Early Hydration Kinetics: Experimental Results

"Early hydration" indicates the collection of chemical reactions taking place during approximately the first 24 h after mixing cement with water (Taylor 1997). These reactions control the fluid-solid transition of the paste, viz., its setting; hence they are crucial for the structure and mechanical performance of the hardened paste and concrete. The hydration rate is typically measured via the heat release rate, obtained from calorimetry experiment, with the dissolution of calcium silicates being the main source of heat. Typical evolutions of early hydration rates are shown in Fig. 1a.

Three interesting features emerge from Fig. 1:

1. The reaction accelerates during the first 10 h, as shown in Fig. 1a (except for a very rapid deceleration during the first minutes (Juilland et al. 2010), which will not be discussed here). This acceleration cannot be explained only by the evolving concentrations of ions in solution. Indeed, the undersaturation of the solution with respect to C_3S increases during early hydration, due to the precipitation of calcium hydroxide (Nicoleau and Nonat 2016; Bullard et al. 2015). However, the increase in undersaturation may only explain acceleration by a factor 2–3 compared to the rate at 1 h, not the order of magnitude in Fig. 1a. The observed strong acceleration is influenced by the evolving morphology of the cement paste microstructure, as discussed later in Sect. 4.
2. The reaction starts decelerating after *ca.* 10 h (see Fig. 1a). The deceleration cannot be simply explained by the C_3S being entirely consumed: indeed, only a small fraction of the C_3S is typically consumed when the deceleration starts (Garrault et al. 2006). Neither the deceleration can be explained by C–S–H filling all the available space for precipitation in the paste, this latter being proportional to the water-to-cement molar ratio, w/c. Indeed, Fig. 1a shows that the early hydration rate, including the deceleration stage, does not depend on
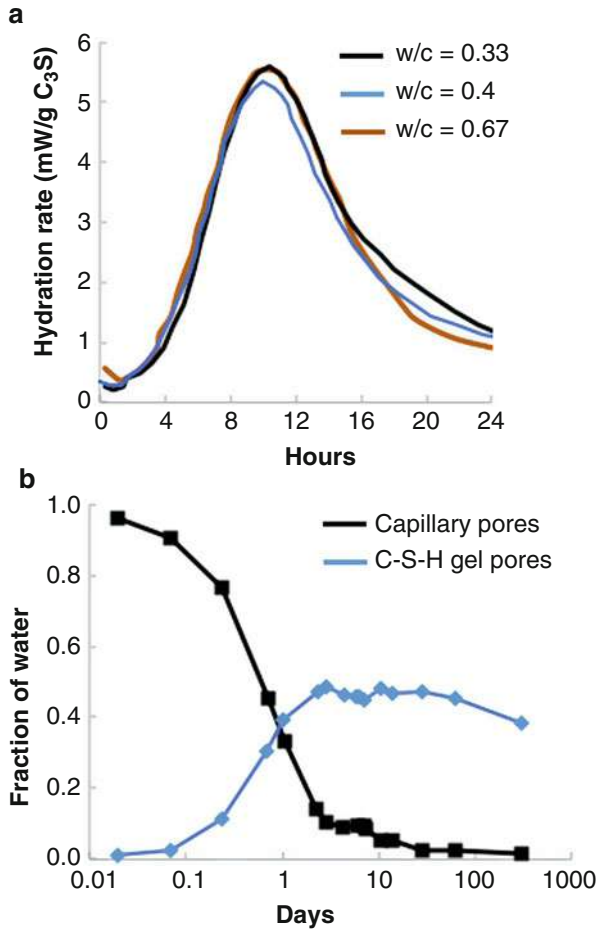
Fig. 1 (**a**) Heat release rate from isothermal calorimetry of hydrating C_3S pastes, which is analogous to the hydration of ordinary cements, except faster. (Data adapted from Thomas 2007). w/c is the mass ratio between water and cement (here C_3S) in the mix. (**b**) Evolution of fraction of water in pores with different sizes, in a hydrating cement paste with $w/c = 0.4$. (Data adapted from Muller et al. 2012). Data from deconvoluted ^{1}H NMR relaxometry spectra. Gel pores have width *ca.* between 1 and 10 nm; capillary pores are larger than that

w/c. Furthermore, Fig. 1b shows that after 1 day of hydration, there is still plenty of volume associated to large pores, where new C–S–H gel can precipitate. Possible explanations of the deceleration stage will be discussed in the next sections.

3. After the first week of hydration, the volume of gel pores in the C–S–H stabilizes and even starts to decrease (see Fig. 1b). Experiments have shown that the hydration of cement can continue logarithmically (hence with decreasing rate) for years and even decades (Bažant et al. 2015). The decrease in C–S–H gel

porosity thus indicates that the gel morphology is continuously evolving at the sub-micrometer scale, and specifically that the gel is getting denser (Muller et al. 2012; Königsberger et al. 2016).

4   The BNG Model: Boundary Nucleation and Growth

Boundary nucleation and growth (BNG) is the name by which the Cahn model of heterogeneous nucleation and growth is known in the cement chemistry literature (Cahn 1956; Thomas 2007). The model provides a framework to discuss the mechanistic origin of the acceleration and deceleration stages of early hydration in Fig. 1a. The BNG model describes the precipitation of C–S–H gel domains at the micrometer length scale, under the assumption that C–S–H precipitation is the rate-controlling process during early hydration. Recently, a model similar to BNG has been used instead to describe the dissolution of C_3S during early hydration, considering this latter as the rate-controlling process (Nicoleau and Nonat 2016). Such dissolution-based models will not be presented in this manuscript, whose focus is on C–S–H precipitation only.

4.1   BNG Model Formulation

A simple way to introduce the BNG model is to start with its homogeneous counterpart: the Avrami model of nucleation and growth (Avrami 1939, 1940; Thomas 2007). The Avrami model starts with considering a unit volume of bulk solution, supersaturated with respect to the precipitation of a solid phase. Under the so-called site saturation hypothesis, a number, N_v^0 of infinitely small nuclei of solid appear at time $t = 0$ (alternative formulations assuming instead a constant nucleation rate are also available in the literature (Scherer et al. 2012)). The model is "homogeneous" in the sense that the solution is the same everywhere and the nuclei can form anywhere with identical probability. As the time increases, the radii of the nuclei (initially null) grow with constant rate G in all directions; hence the domains of precipitating C–S–H gel are represented by growing spheres. G is called "linear growth rate," because it entails radii that increase linearly with time. The last ingredient of the Avrami model is the distinction between "extended" X^e and actual X volume fractions of precipitated solid phases. $X^e(t)$ is simply the cumulative volume of all the above-defined spheres at time t divided by the initial volume of the solution at time $t = 0$. If two spheres, i and j, overlap, $X^e(t)$ double-counts the overlapped volume, considering it as pertaining to both the impinging spheres i and j. $X(t)$ corrects for the overlapping by excluding this double-counting; thus $X(t) \leq X^e(t)$ always. If the spheres are at uncorrelated random locations inside an infinitely large volume of solution, the following relationship between X and X^e holds on average:

$$X = 1 - \exp(-X^e) \qquad (3)$$

The temporal evolution of X resulting from the Avrami model is (Scherer et al. 2012):

$$X = 1 - \exp\left(-\frac{4\pi G^3 N_V^0}{3} t^3\right) \tag{4}$$

The Cahn, or BNG, model is analogous to the Avrami model, except that (i) the initial nuclei appear only on randomly distributed planar surfaces inside the unit volume of solution and (ii) the domains of solid (here C–S–H gel) grow as hemispheres starting from these nuclei, still with constant rate of radial growth G. Two sources of impingement characterize this scenario: a tangential impingement between hemispheres growing from the same planar surface and a bulk impingement between hemispheres growing from different planar surfaces. Each of these two sources of impingement introduce an exponential term akin to Eq. 3. The resulting temporal evolution of X for the BNG model is:

$$X = 1 - \exp\left\{-2O_v^B \int_0^{Gt} \left[1 - \exp\left(-\pi N_S \left(G^2 t^2 - y^2\right)\right)\right] dy\right\} \tag{5}$$

In Eq. 5, y is a spatial coordinate starting from a generic planar surface on which the nuclei lie and increasing perpendicular to the surface itself. N_S is the number of nuclei per unit surface, and O_v^B is the surface available for nucleation per unit volume of solution at time $t = 0$.

4.2 BNG Model Applied to Cement Hydration

For a cement paste, O_v^B in Eq. 5 can be expressed as follows (if one neglects the change of volume of the cement grains due to their partial dissolution during early hydration):

$$O_v^B = \frac{S_c}{V_{w,0}} = S_c \frac{wc}{\rho_w} \tag{6}$$

S_c and $V_{w,0}$ are the initial cement surface and the initial volume of water per unit mass of anhydrous cement. ρ_w is the specific weight of water, and wc is the water-to-cement mass ratio at time $t = 0$.

The volume fraction of precipitated phases, X, is related to the cumulative heat released per unit mass of initially anhydrous cement, $H(t)$, whose derivative provides rate curves as in Fig. 1:

$$H(t) = H_{0,c} \frac{MM_c}{MV_{gel}} X(t) V_{w,0} = \quad \rightarrow \quad \alpha(t) = \frac{H(t)}{H_{0,c}} = \frac{MM_c}{MV_{gel}} \cdot X(t) \frac{wc}{\rho_w} \tag{7}$$

$\alpha(t)$ is the degree of hydration, i.e., the fraction of solid C_3S that has been dissolved and converted to C–S–H at time t. $H_{0,c}$ is the heat released per unit mass of dissolving cement, viz., the reaction enthalpy: –517 J/g of C_3S for the reaction

in Eq. 2 (Costoya 2008). MM_c is the molar mass of cement, and MV_{gel} is the molar volume of the C–S–H gel. The latter depends on the gel porosity and thus on the morphology of the gel itself. Similarly to Eq. 6, also Eq. 7 neglects the change of volume of anhydrous cement dissolving during early hydration. Equation 7 also exploits the fact that each mole of precipitated C–S–H requires a mole of dissolved C_3S: see Eqs. 1 and 2. $H(t)$ is directly proportional to the degree of hydration $\alpha(t)$, which is the mass ratio between the cement that has dissolved at time t and the initial mass of anhydrous cement.

Equations 5, 6, and 7 show that the BNG model includes physical variables whose values are assigned during the preparation of the cement mix: wc, S_c, ρ_w, $H_{0,c}$, and MM_c. On the other hand, the model also relies on three parameters that are not assigned and that describe the structure and precipitation kinetics of the C–S–H gel: G, N_S, and MV_{gel}. These three parameters are typically unknown and difficult to measure; hence they are used to fit experimental data of cement hydration, e.g., those in Fig. 1. Therefore the BNG model is typically used to infer information about the structure and precipitation kinetics of the C–S–H gel (Thomas et al. 2011). On the other hand, the model per se cannot predict changes in hydration rate, precipitation kinetics, or C–S–H gel morphology that may be caused by changes in the chemical composition of the cement. Despite these limitations, the BNG model provides a useful framework to discuss some of the mechanisms that may control the early kinetics of cement hydration.

4.3   BNG Results and the Issue with the Water-to-Cement Ratio

Figure 2 shows hydration rate curves predicted by BNG simulations of a C_3S paste, where all the C_3S particles are assumed to be spherical and with same initial diameter of $7\,\mu$m. The results in Fig. 2 are obtained using a C–S–H gel growth rate $G = 0.036\,\mu$mh^{-1}, a density of C–S–H nuclei on the C_3S surface $N_S = 4\,\mu$m^{-2}, and a molar volume of the gel (including solid and gel pores) MV_{gel} $= 110\,$cm^{3}mol^{-1}: all these values are in the range of previous literature results for C_3S hydration modeled by BNG (Thomas 2007; Scherer et al. 2012; Masoero et al. 2014c).

All the curves in Fig. 2 display a strong initial acceleration, in agreement with the experimental results in Fig. 1a. In the model, the acceleration stems from the increasing surface of the hemispherical C–S–H gel domains growing from the randomly distributed C_3S surfaces. The acceleration is independent of the water-to-cement mass ratio w/c. All the curves in Fig. 2 display a peak rate at *ca.* 10 h. The peak is caused by both tangential impingement of C–S–H gel domains growing from the same C_3S surface and bulk impingement of C–S–H gel growing from different surfaces. In particular, the effect of tangential impingement is indicated by the fact that simulations with very high w/c, and therefore with negligible bulk impingement, still display a peak. On the other hand, the effect of bulk impingement is indicated by the fact that increasing the w/c leads to a large rate at its peak value, and it also shifts the hydration peak to a later time.
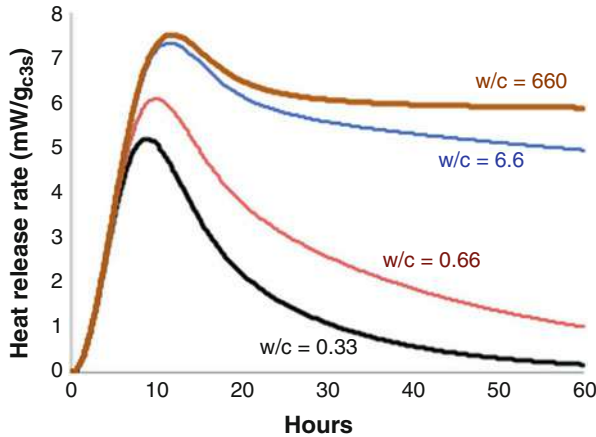
Fig. 2 BNG-simulated rates for C_3S pastes with different water to cement mass ratios (w/c). The specific weight and molar mass of C_3S are $\rho_c = 3150 \, \text{kg/m}^3$ and $MM_c = 228.4 \, \text{g/mol}$

The curve with w/c = 0.33 in Fig. 2 well resembles the experimental results in Fig. 1a. However, the experiments in Fig. 1a show that the w/c ratio does not affect the early hydration rate, and a similar weak effect of w/c has also been shown in experiments at much higher w/c levels than those in Fig. 1a (Garrault et al. 2006). By contrast, the BNG results in Fig. 2 predict a strong impact of w/c, both on the peak rate at *ca.* 10 h and on the post-peak deceleration. This suggests that deceleration in the experiments is not controlled by bulk impingement of C-S-H gel, which is the main post-peak deceleration mechanism in the BNG model.

One possible way to reconcile the BNG model with the weak effect of w/c on early hydration is to assume that (1) the C–S–H gel grows initially with much smaller density (viz., higher MV_{gel}) compared to the C–S–H gel in mature pastes and (2) that the density of early-age C–S–H gel is a decreasing function of the w/c. Recent analyses of experimental results from ^{1}H NMR relaxometry support these hypotheses (Königsberger et al. 2016). However, (i) it is still unclear whether the NMR results are quantitatively in agreement with hydration data; (ii) the lack of w/c dependence in samples hydrating at very high w/c, e.g., 50 or more (Garrault et al. 2006), would imply extremely and unlikely low C–S–H densities; and (iii) the possible physical origin of a w/c-dependent C–S–H gel density is not understood.

Another possible hypothesis to remove w/c dependence from the results of BNG simulations is to assume that early C–S–H precipitation occurs only within a "reaction zone," which is a limited distance from the surface of the cement grains (Masoero et al. 2014c). BNG simulations have shown that the size of this reaction zone would be between 0.4 and 1 μm, which is in the same range as the size of C–S–H gel domains observed experimentally after early hydration (Masoero et al. 2014c; Garrault et al. 2006). However, a mechanistic explanation for such a reaction zone has not been provided yet. The next section shows that the evolving

concentrations of calcium and silicon ions in the cement solution may provide an explanation for the deceleration and thus a mechanistic explanation of the reaction zone hypothesis.

4.4   Chemical Drive for Deceleration in the BNG Model

The BNG model can predict early deceleration without invoking bulk impingement of C–S–H domains, if one considers that the linear growth rate G depends on the evolving chemical composition of the cement solution. Recently, Bullard et al. (2015) implemented this concept into the BNG model by taking:

$$G = G_0 \left[\beta_{CSH}(t) - 1\right] \qquad (8)$$

G_0 is a constant and β_{CSH} is the supersaturation of the cement solution with respect to C–S–H precipitation (Eq. 1). The linear relationship between G and β in Eq. 8 is consistent with linear transition state theory (Lasaga 2014).

β_{CSH} is a function of the concentrations of ions in solution; hence G changes in time following the evolution of these concentrations. Using cellular automaton (CA) simulations, Bullard et al. (2015) modeled the dissolution, diffusion, and precipitation reactions taking place during the early hydration of a high-w/c C$_3$S paste, obtaining the $\beta_{CSH}(t)$ curves in Fig. 3a. The figure highlights the effect of varying a so-called opacity parameter in the CA simulations. This opacity is an empirical parameter associated with the domains of precipitated C–S–H gel in the CA simulations and is related to the inverse of the C–S–H gel permeability. A high opacity means that the precipitated C–S–H gel creates a diffusion barrier between the dissolving C$_3$S grains and the bulk cement solution, leading to lower concentrations of ions in solution and thus a lower β_{CSH} (and G) during the early hydration, as shown in Fig. 3a.

Combining the $\beta_{CSH}(t)$ curves in Fig. 3a with the β-dependent expression of G in Eq. 8, the BNG model leads to the hydration rate curves in Fig. 3b. The curve assuming a constant G, independent of β_{CSH}, is qualitatively similar to the high-w/c curves in Fig. 2, which captures qualitatively well the experiments (e.g., in Fig. 1a) during the acceleration stage but then significantly differs from the experiments at the rate peak and beyond. The curves in Fig. 3b assuming β-dependent G show an initial acceleration that is similar to the case with constant G, except for an initial spike at time $t \approx 0$, which is due to β_{CSH} in Fig. 3a starting very high and then decaying exponentially fast during the first minutes of hydration. Acceleration up to the main hydration peak is still due to the hemispherical growth of C–S–H gel domains. The solution supersaturation becomes important at the hydration peak: Fig. 3a shows that the maximum rate at the peak is indeed an increasing function of β_{CSH}, which is smaller when the opacity is assumed to be high. Past the peak, the deceleration displayed by the β-dependent curves in Fig. 3b is caused by the continuous decrease of β_{CSH} during early hydration. In this way, the BNG model with β-dependent linear growth rate G is effectively uncoupling deceleration from bulk impingement of C–S–H gel. This allows for BNG simulations of pastes with high w/c to still display deceleration, in line with the experiments.
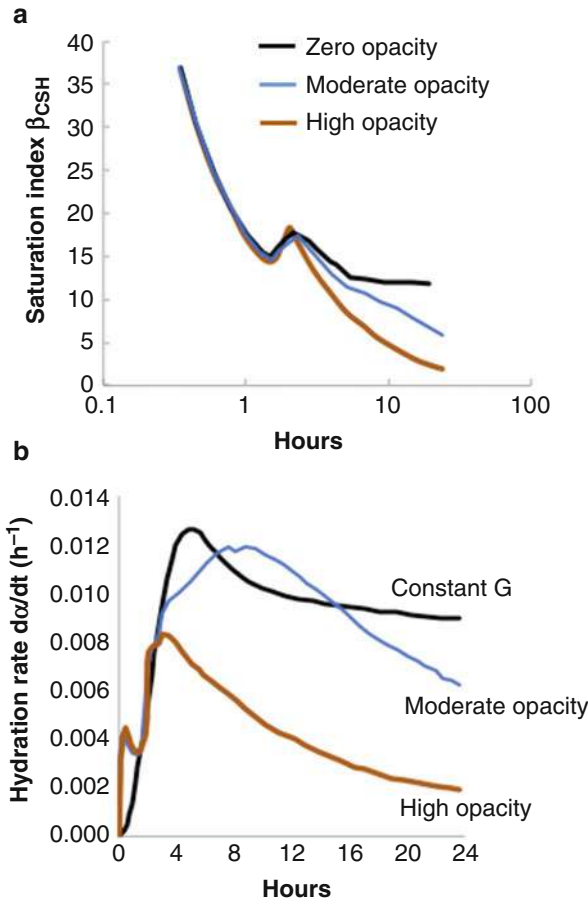
Fig. 3 (**a**) Temporal evolution of C_3S solution supersaturation with respect to C–S–H precipitation: results from cellular automata simulations assuming different levels of C–S–H "opacity" (Bullard et al. 2015). This opacity is an empirical parameter quantifying the difficulty for ions to diffuse through domains of C–S–H gel: see the main text for further discussion. (**b**) BNG simulation results for the hydration of a C_3S paste with high w/c: effect of the relationship between G and β_{CSH} from panel (**a**). (All data in this figure are adapted from Bullard et al. 2015)

5   Nanoparticle Simulations at the Nano-to-Micro Mesoscale

The previous section has shown that if a BNG model accounts for the dependence of the precipitation kinetics on the supersaturation of the solution, the resulting calculations capture both morphology-driven acceleration and w/c-independent deceleration during cement early hydration. However BNG per se can only be used to fit experimental results and not to predict the hydration of new cements. The cause of this limitation is that BNG models rely on four parameters to describe the precipitation kinetics and morphology of the C–S–H gel. These parameters are the linear growth rate of the gel (G), the number of C–S–H nuclei per unit surface

of C_3S (N_S), the molar volume of the gel (MV_{gel}), and the "opacity" of the gel with regard to ion diffusion. Quantifying these parameters as functions of cement chemical composition and curing conditions is difficult. A possible way to address this task is to couple BNG with other models describing the chemo-mechanics of C–S–H precipitation at smaller length scales. Such models must therefore address the mesoscale between the nanometer of chemical reactions and the micrometer above which the gel can be considered as a homogeneous phase in BNG. This is the mesoscale at which the porous morphology of the C–S–H gel develops. In the past decade, nanoparticle simulations have emerged as a possible framework for mesoscale models of C–S–H precipitation (Del Gado et al. 2014; Ioannidou et al. 2016).

5.1 Monte Carlo Insertion and Aggregation of Nanoparticles

Nanoparticle simulations of C–S–H gel precipitation are based on two processes: (i) inserting new particles into a simulation box, to model the precipitation of solid C–S–H nano-units, and (ii) rearranging the particles to minimize the free energy of the system between subsequent particle insertions (Masoero et al. 2012). Both these processes depend on energy contributions coming from interaction potentials, which generate interaction forces between the particles.

5.1.1 Interaction Potentials for C–S–H

The most common interaction potentials in nanoparticle simulations of C–S–H gel precipitation are pairwise, spherical, single-well, or well-shoulder potentials. The main energy well, defining the equilibrium distance between two particles, is typically described by a Lennard-Jones term or by a more general Mie term (Masoero et al. 2012; Ioannidou et al. 2014) (viz., a Lennard-Jones-like potential with exponents that may differ from the usual 12–6 (Mie 1903)). The depth of the energy well ε_0, viz., the interaction strength, is usually particle size dependent (Masoero et al. 2014a). Concerning the magnitude of ε_0 for C–S–H, three options have been proposed thus far (the values below are for particles with a 10 nm diameter):

- Weak interactions, mostly electrostatic, experimentally measured by atomic force microscopy indentation (Plassard et al. 2005). These led to well-shoulder potentials with ε_0 in the 200–500 $k_B T$ range, depending on the concentration of calcium ions in solution (Ioannidou et al. 2014). Interactions of this type imply that C–S–H nanoparticles are not covalently bonded.
- Strong interactions, with ε_0 being the energy to separate two covalently bonded C–S–H particles. This leads to $\varepsilon_0 \approx 19{,}000$ $k_B T$, which includes the strain energy inside the particles that are being separated (Masoero et al. 2014b). These interactions capture the experimentally measured mechanical properties of the C–S–H gel, but they overestimate the changes of interaction energy during precipitation.

- Intermediate interactions, with $\varepsilon_0 \approx 1{,}100 \ k_B T$ linearly related to the surface energy of the C–S–H. The rationale is that two particles at their equilibrium distance have no interface between them (Shvab et al. 2017). In this way, the C–S–H gel is effectively considered as a porous continuum, albeit modeled as an agglomerate of nanoparticles, as opposite to the case of weak electrostatic interactions above.

The weak and intermediate interactions above have ε_0 in the same order of magnitude. To date, they are the most appropriate ones to describe interactions during cement early hydration. However, these interactions underestimate the mechanical properties of the C–S–H gel by at least one order of magnitude (Ioannidou et al. 2016) when compared to experimental measurements from nanoindentation (e.g., Constantinides and Ulm 2007). Future improved formulations of interaction potentials should combine a weak/intermediate ε_0 during precipitation, with strong interactions when the system is mechanically deformed. This would require history-dependent interactions, because two particles at a given distance from each other may represent either incomplete contact during precipitation (weak interaction) or two coalesced particles under strain (strong interaction). History-dependent interactions are yet to be addressed in nanoparticle simulations of C–S–H precipitation.

Another challenge is orientation-dependent interactions. At the molecular level, C–S–H is an anisotropic, layered material (Taylor 1997). This anisotropy contributes to the variety of morphologies that the C–S–H gel can display at the micrometer level, e.g., granular, fibrillar, or foil-like, depending on the ionic strength of the cement solution (Richardson 2008; Tajuelo Rodriguez et al. 2015). Orientation-dependent Gay-Berne interactions have been proposed to obtain complex morphologies in particle simulations, but their use is still limited (Yu and Lau 2015).

Finally, one should note that all the interactions in this section have $\varepsilon_0 \gg k_B T$: this will become a problem for the Monte Carlo particle insertion procedure discussed hereafter.

5.1.2   Monte Carlo Particle Insertions and Deletions

Monte Carlo simulations of C–S–H precipitation define probabilities to decide whether to insert a new particle with assigned diameter at a certain random location in the simulation box (P^{in}) or to delete a randomly selected particle (P^{del}) (Frenkel and Smit 2002; Masoero et al. 2012). Typical formulations for these probabilities are:

$$P^{in} = \min\left[1 \ , \ \Psi \exp\left(\frac{\mu - \Delta U}{k_B T}\right)\right] = \min\left[1 \ , \ \Psi \beta_{CSH} \exp\left(-\frac{\Delta U}{k_B T}\right)\right] \quad (9)$$

$$P^{del} = \min\left[1 \ , \ \Psi \exp\left(-\frac{\Delta U}{k_B T}\right)\right] \quad (10)$$

ΔU is the change in total interaction energy caused by the insertion or deletion event. μ is the chemical potential of the solution with regard to C–S–H precipitation,

which is related to the supersaturation β_{CSH} (Lasaga 2014). $\Psi < 1$ is an arbitrary positive constant controlling the quality of the sampling of possible locations for particle insertion and of possible particle deletions. Small Ψ imply that the Monte Carlo algorithm will attempt many possible insertions and deletions before accepting an event: this leads to time-consuming simulations resulting however into more stable configurations (viz., configurations with lower total interaction energy) compared to simulations with large Ψ. Vice versa, large Ψ lead to faster simulations but also favor metastable configurations.

Both cases of small and large Ψ are problematic. Starting with the case of large Ψ, e.g., $\Psi = 1$, the energy scale $\varepsilon_0 \gg k_B T$ discussed in the previous section implies that particle insertion at any location (not overlapping with other existing particles) has $P^{in} = 1$, because $\beta_{CSH} > 1$ during early hydration (Fig. 3a), $\Delta U < 0$ in Lennard-Jones like potentials, and ΔU is scaled by a large ε_0. Vice versa, deleting an existing particle implies $\Delta U > 0$, and this, scaled by a large ε_0, gives $P^{del} \approx 0$. In this case, the simulation reduces to random space filling (Masoero et al. 2014a). This becomes an issue when one tries to capture a BNG process using nanoparticle simulations. To understand why, consider an empty simulation box containing only a surface made of particles representing C_3S, and assume that the interaction potential between a C_3S particle and the C–S–H particles to be inserted is similar to the interaction potential between two C–S–H particles. In principle, new C–S–H particles should have a higher probability to form near the C_3S surface, but in practice any location for insertion has $P^{in} = 1$ due to the large Ψ. Therefore, one can only capture mechanisms of homogeneous nucleation using Monte Carlo simulations of particle insertion and deletion with large Ψ.

At the other end, simulations with a very small $\Psi \ll 1$ and with monodisperse C–S–H particles (viz., diameter fixed to a single possible value, typically around 5–10 nm Allen et al. 2007) would indeed predict a BNG-like mechanism in the scenario described in the previous paragraph. In this case, what BNG models call "growth" is actually the product of subsequent aggregations of new C–S–H particles forming nearby the previously inserted ones. On the other hand, simulation with $\Psi \ll 1$ are problematic when the C–S–H particles are allowed to have diameters within a certain range, viz., if size polydispersity is considered. In this case, the proportionality between ε_0 and the particle size (e.g., with the particle surface, as discussed in Sect. 5.1.1) implies that large particles have higher P^{in} as long as there is space to insert term with $\Delta U < 0$, i.e., not overlapping with other particles. In the extreme case of $\Psi \rightarrow 0$ and infinite particle size polydispersity (i.e., newly inserted particles can have diameter ranging from 0 to the size of the simulation box), the first particle to be inserted would be the largest possible one, thus as large as the simulation box. This is indeed the correct answer for a Monte Carlo algorithm whose target is to find the lowest-energy state. This is also realistic for C–S–H, in the sense that in the limit of time tending to infinity, a fully crystallized single domain of C–S–H is the most thermodynamically stable state. On the other hand, this thermodynamic limit bears only limited relevance to the process of C–S–H precipitation and morphology development during cement early hydration (Del Gado et al. 2015).

In conclusion to this section, Monte Carlo simulations of particle insertion/deletion using the probabilities in Eqs. 9 and 10 lead to unrealistic or irrelevant mesoscale mechanisms of C–S–H precipitation, in the limits of $\Psi \to 0$ and $\Psi \to 1$. More realistic configurations, able to capture various experimentally observed structural and mechanical features of the C–S–H gel (Ioannidou et al. 2016), can be enforced by using intermediate values of Ψ, or restricting the range of possible particle sizes, or alternating steps of particle insertion/deletion with multiple steps of particles displacement (not discussed here but explored in the literature, e.g., in Ioannidou et al. 2014). In these cases, however, one should always remember that the sequence of Monte Carlo moves going from an initially empty simulation box to a final C–S–H gel morphology bears no physical information about the actual kinetics of the precipitation process. The only rigorous results of these Monte Carlo simulations are the final morphologies and configurations that they produce, not their temporal evolution during hydration.

5.2   Chemical Kinetics and Kinetic Monte Carlo

The limitations of Monte Carlo simulations with regard to describing realistic kinetics of precipitation can be overcome using kinetic Monte Carlo (KMC) (Voter 2007). In KMC simulations, the probability of inserting a particle is (only insertions are discussed in this section, but the same concepts apply to particle deletions):

$$P^{in} = \frac{R^{in}}{Q} \tag{11}$$

where R^{in} is the rate of the average insertion event at the selected location and Q is a normalizing constant. R^{in} can be expressed starting from fundamental concepts of chemical kinetics, as shown later in this section. Q plays a similar role as the inverse of Ψ in Eqs. 9 and 10, but differently from Ψ, it is not arbitrary and is defined as the sum of the rates of all the possible events. Furthermore, Q is related to the average waiting time for an event; hence KMC generates a proper temporal evolution of the system, differently from the previously discussed Monte Carlo simulations, which did not have a timescale.

Akin to P^{in} in Eq. 9, the position-dependent insertion rate R^{in} depends on both the chemical drive for precipitation, β_{CSH}, and on the interaction free energy ΔU. However, differently from P^{in}, the dependence on ΔU is not simply $\exp[-\Delta U)/(k_B T)]$, but a more complex expression that accounts for the size of the particle to be inserted. A physically sound expression of this dependence of R^{in} on ΔU avoids the issue of large particles appearing before small ones in the simulations. A recently proposed derivation of R^{in} by Shvab et al. (2017) is presented hereafter.

To express R^{in}, one must consider that a particle insertion is actually a coarse-grained representation of n chemical reactions taking place at the molecular scale. For example, forming a C–S–H nanoparticle with a diameter of 10 nm implies

ca. 1,000 chemical reactions of C–S–H precipitations as per Eq. 1. The rate of an individual chemical reaction, r_i, is the building block of R^{in}. Following transition state theory, one can write:

$$r_i = \frac{k_B T}{h} \frac{1}{\gamma^*} \beta_{CSH} \ \exp\left[-\frac{\Delta G^* + \chi(\gamma \Delta \Omega_i + \Delta U_i)}{k_B T} \right] \tag{12}$$

h is the Planck constant. γ^* and ΔG^* are the standard state activity coefficient and activation energy of the transition state associated with the chemical reaction. γ is the solid-solution interfacial energy. $\Delta \Omega_i$ and ΔU_i are the changes in particle surface area and total interaction potential caused by the ith chemical reaction during particle formation. χ is a number between 0 and 1, quantifying the contribution of $\Delta \Omega_i$ and ΔU_i to the precipitation rate (see Shvab et al. 2017 for more details).

To go from r_i in Eq. 12 to the coarse-grained R^{in}, one must describe the spatially resolved mechanism by which the molecules assemble to form a particle. Various classical and nonclassical mechanisms are employed today across the field of chemical kinetics (Van Driessche et al. 2017). For C–S–H, Shvab et al. (2017) have recently obtained an expression for R^{in} under the assumptions that one-molecular C–S–H nuclei are always present in the cement solution with fixed concentration C_{cn} and that their growth into particles follows a classical mechanism of layer-by-layer radial growth. For interactions with $\varepsilon_0 \sim \Omega$, where Ω is the particle surface area, this leads to:

$$R^{in} = k r_0^* \beta_{CSH} \ a^2 \left\{ \sum_{i=1}^{n} \exp\left[\chi \left(\frac{\gamma_i + \frac{\Delta U^{in}}{\Omega}}{k_B T} \right) \frac{8a^2 i}{(2i-1)^2} \right] \right\}^{-1} \tag{13}$$

k is a constant that depends: on the algorithm to sample the possible positions for particle insertion, on C_{cn}, and on the solid volume fraction in the surrounding of the location chosen to attempt the insertion. $r_0^* = \frac{k_B T}{h} \frac{1}{\gamma^*} \exp\left(-\frac{\Delta G^*}{k_B T}\right)$ is the reaction rate in standard state. a is the characteristic linear size of a molecule, *ca.* 0.67 nm for C–S–H. n is the number of reactions taking place in series: for the classical layer-by-layer growth mechanisms, n is equal to the particle radius divided by a. Similar to P^{in} Eq. 9, also R^{in} in Eq. 13 depends linearly on β_{CSH} and exponentially on ΔU. However, ΔU in R^{in} is divided by Ω, and this ratio is independent of the particle size (because Eq. 13 is based on the assumption that $\Delta U \sim \Omega$). Furthermore, R^{in} is inversely related to n, which is a measure of the particle size; hence large particles will have smaller R^{in} than small particles, as is realistic to expect.

Figure 4 shows results from KMC simulations of C–S–H precipitation, which employ R^{in} from Eq. 13 and a similarly defined rate for particle deletion (see Shvab et al. 2017). All the parameters in R^{in} have been taken from the literature on C–S–H, except for two of them: r_0^*, which is assumed to be constant and calibrated to capture the correct timescale, and the interaction strength (75% of ε_0) between C–S–H particles and the particles forming the C_3S surface. Depending on the simulation, β_{CSH} was assumed to either be constant or to change in time as shown
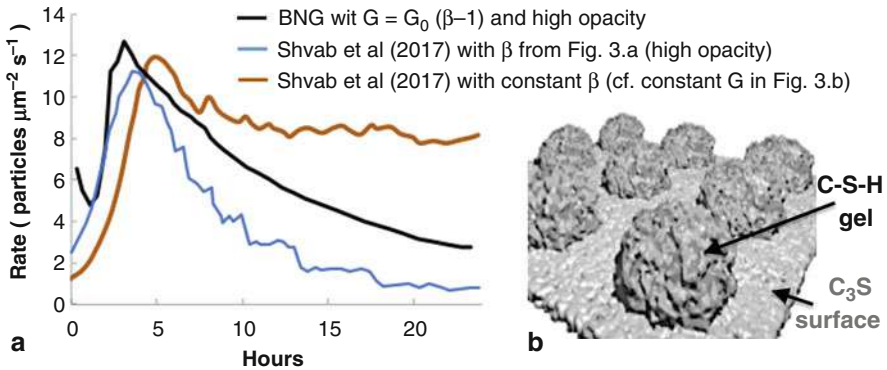
Fig. 4 (**a**) Hydration rate curves from BNG and from KMC simulations: effect of β (**b**) Snapshot from a KMC simulation showing a BNG mechanism emerging from particle insertion and aggregation. (Data adapted from Bullard et al. 2015 and from Shvab et al. 2017)

in Fig. 2a. The curves in Fig. 4a are consistent with the BNG results in Fig. 2b, in that the acceleration stage is weakly affected by β_{CSH} and is instead governed by the morphological evolution of the hemispherical domains of C–S–H gel (see Fig. 4b) and that the deceleration stage is controlled by the decrease of $\beta_{CSH}(t)$ during hydration.

5.3 Insights into BNG Parameters from Nanoparticle Simulations

Nanoparticle KMC simulations provide a more fundamental insight into the four free parameters of BNG (see introduction to Sect. 5), also linking them to the chemistry of the cement paste.

- The linear growth rate G can be read directly from a nanoparticle simulation by looking at the time derivative of the radius of the hemispherical domains in Fig. 4b. G depends on all the quantities contributing to R^{in} in Eq. 13: this also entails the linear dependence on β_{CSH} already included in recent BNG simulations. Furthermore, Eq. 13 shows that R^{in} and thus G are functions of other chemistry-dependent quantities such as the standard state rate r_0^*, the solid-solution interfacial energy γ, and the interaction potentials. Finally, since the hemispherical gel domains in Fig. 4b include pores between the particles, G also depends on the gel morphology (e.g., its porosity) which depends in turn on the interaction potentials and again on the material's chemistry.
- The number of nuclei N_S per unit surface of C$_3$S depends on the interaction strength between C–S–H and C$_3$S and thus also on their interfacial energy. However, N_S depends also on the morphology and roughness of the C$_3$S surface, as well as on the presence of crystallographic or chemical defects that may locally

    increase the surface energy or the strain energy of the C_3S. KMC nanoparticle simulations can be used to explore the impact of these factors on N_S.

- The molar volume of the C–S–H gel MV_{gel} depends on the porosity and thus on the morphology of the agglomerates in Fig. 4b. There is significant scope for nanoparticle simulations to relate the emergence of complex morphologies to the functional form and chemistry dependence of the interactions between particles.
- The "opacity" of the C–S–H gel is also related to the pore structure and morphology of the C–S–H particle agglomerates in Fig. 4b. Coupling nanoparticle simulations with simulations of ion diffusion through the C–S–H gel, e.g., Lattice-Boltzmann simulations (Zalzale 2014; Patel 2016), would help with rigorously defining and quantifying the opacity parameter.

6   Conclusion

Boundary nucleation and growth (BNG) models predict an acceleration of the rate of C–S–H formation that is consistent with the acceleration of cement early hydration rate measured by isothermal calorimetry. However, BNG models also predict that the water-to-cement ratio of a cement paste should have a significant impact on the early hydration rate: this prediction disagrees with the experimental evidence. This issue can be overcome by considering a linear growth rate in BNG that depends linearly on the supersaturation of the cement solution with respect to C–S–H precipitation.

    BNG relies on four parameters describing the sub-micrometer structure and precipitation kinetics of the C–S–H. These parameters are usually calibrated from experimental results. New cement formulations require new calibrations; hence the predictive ability of BNG is limited. An alternative approach to inform BNG is offered by nanoparticle simulations of C–S–H nucleation and growth by aggregation. When used in the Monte Carlo framework, nanoparticle simulations produce configurations that capture a number of structural features and mechanical properties of the mesoporous C–S–H gel phase. However, Monte Carlo simulations lead to unphysical kinetics, also due to the large energy scale of the mechanical interactions between C–S–H nanoparticles.

    To overcome the limitations of Monte Carlo simulations, new simulations of nanoparticle insertion, deletion, and aggregation have been recently proposed in the framework of kinetic Monte Carlo (KMC). The key advantages of KMC simulations are to (i) reformulate the probabilities of particle insertion and deletion in terms of rates, which introduce an explicit dependence on particles size that removes the issues raised by the large energy scale in Monte Carlo simulations; (ii) provide a rigorous definition of time; (iii) link precipitation rates to solution chemistry using coarse-grained transition state theory; and (iv) predict mechanisms and overall precipitation kinetics that are consistent with BNG simulations. As a result, KMC simulations of C–S–H precipitation enable a more fundamental approach to the calibration of the parameters in the BNG model, linking them to the chemistry of the cement solution and thus expanding the predictive ability of the BGN model.

Overall, this manuscript shows that top-down BNG modeling can be combined with bottom-up simulations to gain a better understanding of the mechanisms of C–S–H precipitation and to clarify how these depend on the chemical composition and on the curing conditions of the cement paste. A similar approach can also benefit research on a wider class of mesostructured materials, whose formation occurs by gelation and precipitation from solution. KMC simulations however still need to overcome several barriers before being fully deployed to predict the kinetics of mesostructured materials formation. To this end, and focusing on simulations of cement hydration, future efforts should address (i) more accurate interaction potentials, which will probably have to consider anisotropy and history dependence; (ii) more chemical complexity and multiple hydration products beyond C–S-H; (iii) describing more in detail the mechanism and kinetics of C–S–H precipitation at the molecular scale; (iv) coupling particle simulations with ion diffusion and local saturation in the solution; (v) allowing for particles to grow and dissolve partially, and anisotropically, to capture complex morphologies that have been experimentally observed, e.g., fibrils and foils (Tajuelo Rodriguez et al. 2015); and (vi) describing more in detail the morphology and chemical structure of the C_3S surface, as well as its interactions with the C–S–H. Driven by a collective effort in the materials modeling and simulation community, nanoparticle simulations are displaying an increasing promise to support the formulation of future more sustainable cements, by providing improved and more fundamental predictions of cement hydration, properties, and durability.

References

Allen AJ, Thomas JJ, Jennings HM (2007) Composition and density of nanoscale calcium-silicate-hydrate in cement. Nat Mater 6:311–316

Avrami M (1939) Kinetics of phase change. I. General theory. J Chem Phys 7(12):1103–1112

Avrami M (1940) Kinetics of phase change. II. Transformation-time relations for random distribution of nuclei. J Chem Phys 8(2):212–224

Bažant Z, Donmez A, Masoero E, Aghdam SR (2015) Interaction of concrete creep, shrinkage and swelling with water, hydration, and damage: nano-macro-chemo. In: Hellmich C, Pichler B, Kollegger J (eds) Proceedings of the 10th international conference on mechanics and physics of creep, shrinkage, and durability of concrete and concrete structures, Vienna. ASCE Library, pp 1–12

Bullard JW, Jennings HM, Livingston RA, Nonat A, Scherer GW, Schweitzer JS, Scrivener KL, Thomas JJ (2011) Mechanisms of cement hydration. Cem Concr Res 41:1208–1223

Bullard JW, Scherer GW, Thomas JJ (2015) Time dependent driving forces and the kinetics of tricalcium silicate hydration. Cem Concr Res 74:26–34

Cahn JW (1956) The kinetics of grain boundary nucleated reactions. Acta Metall 4(5):449–459

Constantinides G, Ulm FJ (2007) The nanogranular nature of C–S–H. J Mech Phys Solids 55: 64–90

Costoya MM (2008) Effect of particle size distribution on the hydration kinetics and microstructural development of tricalcium silicate. PhD thesis, EPFL, Lausanne

Del Gado E, Ioannidou K, Masoero E, Baronnet A, Pellenq RM, Ulm FJ, Yip S (2014) A soft matter in construction–statistical physics approach to formation and mechanics of C–S–H gels in cement. Eur Phys J Spec Top 223(11):2285–2295

Del Gado E, Ioannidou K, Masoero E, Pellenq RJM, Ulm FJ, Yip S (2015) The meso-scale texture of cement hydrate gels: out-of-equilibrium evolution and thermodynamic driving. In: Hellmich C, Pichler B, Kollegger J (eds) Proceedings of the 10th international conference on mechanics and physics of creep, shrinkage, and durability of concrete and concrete structures, Vienna. ASCE Library, pp 34–38

European Commission (2011) A roadmap for moving to a competitive low carbon economy in 2050. European Commission, Bruxelles

Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications. Academic Press, New York

Garrault S, Behr T, Nonat A (2006) Formation of the C-S-H layer during early hydration of tricalcium silicate grains with different sizes. J Phys Chem B 110(1):270–275

Grant SA, Boitnott GE, Korhonen CJ, Sletten RS (2006) Effect of temperature on hydration kinetics and polymerization of tricalcium silicate in stirred suspensions of cao-saturated solutions. Cem Concr Res 36(4):671–677

Ioannidou K, Pellenq RJM, Del Gado E (2014) Controlling local packing and growth in calcium–silicate–hydrate gels. Soft Matter 10:1121–1133

Ioannidou K, Krakowiak KJ, Bauchy M, Hoover CG, Masoero E, Yip S, Ulm FJ, Levitz P, Pellenq RJM, Del Gado E (2016) Mesoscale texture of cement hydrates. Proc Natl Acad Sci 113(8):2029–2034

Jennings HM (2008) Refinements to colloid model of CSH in cement: Cm-ii. Cem Concr Res 38(3):275–289

Juenger M, Winnefeld F, Provis JL, Ideker J (2011) Advances in alternative cementitious binders. Cem Concr Res 41(12):1232–1243

Juilland P, Gallucci E, Flatt R, Scrivener K (2010) Dissolution theory applied to the induction period in alite hydration. Cem Concr Res 40(6):831–844

Königsberger M, Hellmich C, Pichler B (2016) Densification of CSH is mainly driven by available precipitation space, as quantified through an analytical cement hydration model based on NMR data. Cem Concr Res 88:170–183

Lasaga AC (2014) Kinetic theory in the earth sciences. Princeton University Press, Princeton

Manzano H, Moeini S, Marinelli F, Van Duin AC, Ulm FJ, Pellenq RJM (2012) Confined water dissociation in microporous defective silicates: mechanism, dipole distribution, and impact on substrate properties. J Am Chem Soc 134(4):2208–2215

Masoero E, Del Gado E, Pellenq RJM, Ulm FJ, Yip S (2012) Nanostructure and nanomechanics of cement: polydisperse colloidal packing. Phys Rev Lett 109(15):155503

Masoero E, Del Gado E, Pellenq RJM, Yip S, Ulm FJ (2014a) Nano-scale mechanics of colloidal C–S–H gels. Soft Matter 10:491–499. https://doi.org/10.1039/C3SM51815A

Masoero E, Jennings H, Ulm F, Del Gado E, Manzano H, Pellenq R, Yip S (2014b) Modelling cement at fundamental scales: from atoms to engineering strength and durability. Comput Model Concr Struct 1:139–148

Masoero E, Thomas JJ, Jennings HM (2014c) A reaction zone hypothesis for the effects of particle size and water-to-cement ratio on the early hydration kinetics of C3S. J Am Ceram Soc 97(3):967–975

Mie G (1903) Zur kinetischen theorie der einatomigen körper. Annalen der Physik 316(8):657–697

Muller AC, Scrivener KL, Gajewicz AM, McDonald PJ (2012) Densification of C–S–H measured by 1h NMR relaxometry. J Phys Chem C 117(1):403–412

Nicoleau L, Nonat A (2016) A new view on the kinetics of tricalcium silicate hydration. Cem Concr Res 86:1–11

Olivier JG, Muntean M (2014) Trends in global CO_2 emissions: 2014 Report. PBL Netherlands environmental assessment agency and institute for environment and sustainability of the European commission's joint research centre, The Hague

Patel R (2016) Lattice Boltzmann method based framework for simulating physico-chemical processes in heterogeneous porous media and its application to cement paste. PhD thesis, Ghent University

Plassard C, Lesniewska E, Pochard I, Nonat A (2005) Nanoscale experimental investigation of particle interactions at the origin of the cohesion of cement. Langmuir 21(16):7263–7270

Richardson I (2008) The calcium silicate hydrates. Cem Concr Res 38(2):137–158

Scherer GW, Zhang J, Thomas JJ (2012) Nucleation and growth models for hydration of cement. Cem Concr Res 42(7):982–993

Scrivener KL, John VM, Gartner EM (2018) Eco-efficient cements: potential, economically viable solutions for a low-CO_2, cement-based materials industry. Cem Concr Res 114:2–26

Shvab I, Brochard L, Manzano H, Masoero E (2017) Precipitation mechanisms of mesoporous nanoparticle aggregates: off-lattice, coarse-grained, kinetic simulations. Cryst Growth Des 17(3):1316–1327

Tajuelo Rodriguez E, Richardson I, Black L, Boehm-Courjault E, Nonat A, Skibsted J (2015) Composition, silicate anion structure and morphology of calcium silicate hydrates (CSH) synthesised by silica-lime reaction and by controlled hydration of tricalcium silicate (C3S). Adv Appl Ceram 114(7):362–371

Taylor HF (1997) Cement chemistry. Thomas Telford, London

Thomas JJ (2007) A new approach to modeling the nucleation and growth kinetics of tricalcium silicate hydration. J Am Ceram Soc 90:3282–3288

Thomas J, Biernacki J, Bullard J, Bishnoi S, Dolado J, Scherer G, Luttge A (2011) Modeling and simulation of cement hydration kinetics and microstructure development. Cem Concr Res 41:1257–1278

Ulm FJ, Coussy O (1996) Strength growth as chemo-plastic hardening in early age concrete. J Eng Mech 122(12):1123–1132

Van Driessche AES, Kellermeier M, Benning LG, Gebauer D (Eds) (2017) New perspectives on mineral nucleation and growth: from solution precursors to solid materials. Springer, Cham

Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Sickafus KE et al (eds) Radiation effects in solids. Springer, Dordrecht, pp 1–23

Yu Z, Lau D (2015) Nano-and mesoscale modeling of cement matrix. Nanoscale Res Lett 10(1):173

Zalzale M (2014) Water dynamics in cement paste: insights from lattice Boltzmann modelling, PhD thesis, EPFL Lausanne, Switzerland

In Situ AFM Investigations and Fracture Mechanics Modeling of Slow Fracture Propagation in Oxide and Polymer Glasses

10

Matteo Ciccotti and Matthieu George

Contents

Abstract

Fracture propagation is inherently a multiscale problem, involving the coupling of many length scales from sample dimension to molecular level. Fracture mechanics provides a valuable link between the macroscopic scale of the structural loading of the samples and the scale of the process zone for brittle materials. Modeling the toughness of materials requires yet an investigation at scales smaller than this process zone, which is nanometric in oxide glasses and micrometric in polymer glasses. We present here the important insights that have

M. Ciccotti (✉)
Laboratoire Sciences et Ingénierie de la Matière Molle (SIMM), ESPCI Paris, PSL University, Sorbonne Université, CNRS, Paris, France
e-mail: matteo.ciccotti@espci.fr

M. George
Laboratoire Charles Coulomb (L2C), Université de Montpellier, CNRS, Montpellier, France
e-mail: matthieu.george@umontpellier.fr

been obtained through an in situ experimental investigation of the strain fields in the micrometric neighborhood of a propagating crack. We show the richness of atomic force microscopy combined with digital image correlation although it limits the observations to the external surface of the sample and to very slow crack propagation (below nm/s). For oxide glasses, this novel technique provided enlightening information on the nanoscale mechanisms of stress corrosion during subcritical crack propagation (Ciccotti, J Phys D Appl Phys 42:214006, 2009; Pallares et al., Corros Rev 33(6):501–514, 2015), including the relevance of crack tip plasticity (Han et al., EPL 89:66003, 2010), stress-induced ion exchange processes (Célarié et al., J Non-Cryst Solids 353:51–68, 2007), and capillary condensation in the crack tip cavity (Grimaldi et al., Phys Rev Lett 100:165505, 2008; Pallares et al., J Am Ceram Soc 94:2613–2618, 2011). An extension of this technique has recently been developed for glassy polymers (George et al., J Mech Phys Solids 112:109–125, 2018), leading to novel insights on the transition between crazing and shear yielding mechanisms and to promising new ways to link the toughness properties to the time-dependent large strain material properties of these nominally brittle materials.

1   Introduction

Fracture propagation is inherently a multiscale problem, involving the coupling of many length scales from sample dimension to molecular level. Fracture mechanics provides a valuable link between the macroscopic scale of the structural loading of the samples and the scale of the process zone for brittle materials such as oxide and polymer glasses. On the one hand, LEFM (see Insert A) allows measuring toughness of a given material on a given standard test sample and uses this information to predict the resistance to fracture propagation of a different structure composed of the same material including a well-defined crack. On the other hand, LEFM does not allow explaining the measured toughness value based on the macroscopic properties of the material, nor its dependence on crack propagation velocity or temperature.

Figure 1 shows that the toughness of very different materials (including the ones investigated in the present manuscript) presents significant variations when crack propagation velocity spans over many different orders of magnitude, going from pm/s to km/s. The overall trend of these curves is surprisingly similar, consisting of a slow velocity regime (frequently called "subcritical") with a $\Gamma(v)$ with positive slope as a function of the velocity v up to a critical velocity v_c. Above this instability point (see Insert A), crack propagation velocity rapidly attains values close to the sound propagation speed in the material. Kinetic energy dominates the fracture energy in this dynamic regime (Freund 1990; Fineberg and Marder 1999). On the other hand, in the slow velocity regime, which is the object of the present manuscript, crack propagation is quasi-static, and the materials toughness is determined by the dissipative processes associated with crack propagation.
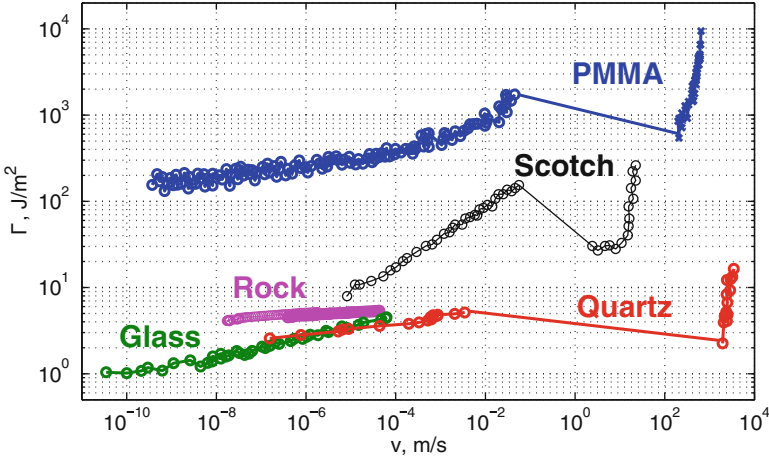
Fig. 1 Fracture energy curves $\Gamma(v)$ for some different materials

The ranking of the materials toughness in this subcritical regime is thus not trivially related to the strength of the specific interatomic bonds.

Insert A: Linear Elastic Fracture Mechanics (LEFM)

LEFM describes the conditions of propagation of a preexisting sharp crack in linear elastic materials. The original argument by Griffith (1920) associated the creation of a new crack to the conversion of mechanical energy (external work W and variations of the elastic energy U_{el}) into a thermodynamic (reversible) energy cost per unit area $\Gamma = 2\gamma$ to create two new free surfaces:

$$G = \frac{\partial W}{\partial A} - \frac{\partial U_{el}}{\partial A} = \Gamma \qquad [\text{J/m}^2]$$

where the *strain energy release rate G* represents the loading condition and structural response of the precracked sample and the *fracture energy Γ* is a material property.

An equivalent description of LEFM can be expressed in terms of a singular stress field $\sigma_{ij}(r, \theta)$ in the neighborhood of the crack tip in the form of an inverse square root dependence of the stress tensor σ_{ij} on the distance r from the crack tip:

$$\sigma_{ij}(r, \theta) = \frac{K_I}{\sqrt{2\pi r}} f_{ij}(\theta)$$
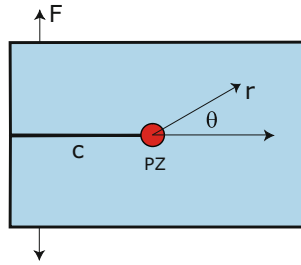
where $f_{ij}(\vartheta)$ are universal oscillating functions of the angle θ.

When limiting to mode I (opening) cracks, the equilibrium/propagation condition can thus be expressed as $K_I \geq K_{Ic}$ [Pa·m$^{1/2}$], where the *stress intensity factor* (SIF) $K_I = \sqrt{E'G}$ is the loading parameter and the *fracture toughness* $K_{Ic} = \sqrt{E'\Gamma}$ is a material property. The effective elastic modulus is $E' = E$ for plane stress and $E' = E/(1 - v^2)$ for plane strain, E and v being the Young modulus and Poisson's ratio of the material.

The elastic crack opening profile $u_y(x)$ of an initially sharp slit can be shown to have locally a parabolic shape:

$$u_y(x) = \frac{K_I}{E'}\sqrt{\frac{8x}{\pi}}$$

The LEFM formalisms can be extended to inelastic materials provided all inelastic deformations are limited to small region around the crack tip, called the "process zone" as shown by Irwin and Orowan (Orowan 1955; Irwin 1957). The irreversible fracture energy $\Gamma(v, T) \gg w$ includes energy dissipation and can become several orders of magnitude larger than 2γ. Moreover, it can depend on crack propagation velocity v, on temperature and environmental conditions (especially relative humidity). The stability of steady-state crack propagation requires both the structural stability of the sample $dG(c)/dc \geq 0$ and intrinsic stability of the material $d\Gamma(v)/dv \leq 0$.

For hard and stiff elastoplastic materials, the size of the plastic process zone can be estimated as a function of their yield stress σ_Y through Dugdale model:

$$\ell_D = \chi \frac{K_{Ic}^2}{\sigma_y^2}$$

where χ ranges between $1/2\pi \simeq 0.16$ and $\pi/8 \simeq 0.39$ depending on the choice of a circular (McClintock and Irwin 1964) or elongated (Dugdale 1960) process zone.

Typical dissipation mechanisms involve plastic or viscoelastic deformation of the bulk material in the neighborhood of the crack tip which is affected by a propagating stress singularity. Moreover, the strong deformation of the material in this process zone can induce the activation of some mesoscale mechanisms, such as microcracks, cavitation, crazing/fibrillation, shear banding/localization, rupture and/or sliding of interfaces between different phases of the material (Lawn 1993; Williams 1984; Jones 2015), or molecular processes such as stress-induced diffusion of ions or molecules, stress-activated chemical reactions, or phase changes (Lawn 1993; Ciccotti 2009; Hutter 2013).

The scientific field of modeling materials toughness is thus a strongly multi-disciplinary domain, which implies handling mechanics at the macroscopic scale, in situ investigations of micromechanisms at the mesoscale, and physicochemical modeling at the atomistic scales. The focus of the present manuscript is in the development of novel AFM-based techniques to provide in situ investigation of the mesoscale mechanisms in brittle materials such as oxide and polymer glasses, where the size of the process zone is, respectively, nanometric and micrometric (see Insert A). On the one hand, an appropriate modeling of these mesoscale observations requires to accurately investigate and model the mechanical aspects at the macroscopic scale, which provides the pertinent loading and boundary conditions for the process zone scale. On the other hand, the clear observation and modeling of the mesoscale mechanisms provide the pertinent conditions that affect the molecular processes acting at an even smaller scale. The mesoscale is frequently the scale where most of the energy dissipation takes place, but the amount of this energy dissipation is conditioned by the phenomena acting in a molecular neighborhood of the crack tip. The mesoscale is thus an essential bridge between the mechanical and molecular scales and a key factor to model materials toughness.

The key for a proper communication between the three scales lays into the achievement of a well-conditioned steady-state propagating fracture, where the processes at all scales are conveniently synchronized. This allows the energy flow to be coherently defined from the macroscopic scale of the loading stage down to the ultimate molecular failure scale and thus to attribute to each scale the proper contribution in the emergence of the fracture toughness of the material.

The novel technique presented here exploits the combination of an AFM (Bruker ICON, Santa Barbara, USA) with a very stable and compact in situ loading configuration (Fig. 2) based on the double cleavage drilled compression (DCDC) sample (Janssen 1974) and a very stiff and compact loading stage (Deben Microtest, Woolpit, UK). The macroscopic nature of the sample and its excellent stability allow obtaining a well-defined mode I loading and a very straight crack path. This allows for a very stable steady-state fracture propagation in materials where the fracture toughness K_c can be expressed as an increasing function of the crack propagation velocity v (see Insert A).

The straight AFM topographical imaging (see Insert B) of the crack tip neighborhood allows to identify the pertinent mechanisms acting at the process zone scale. Furthermore, the continuum imaging of the same region of the sample surface while it is progressively crossed by the slowly propagating crack allows extracting the
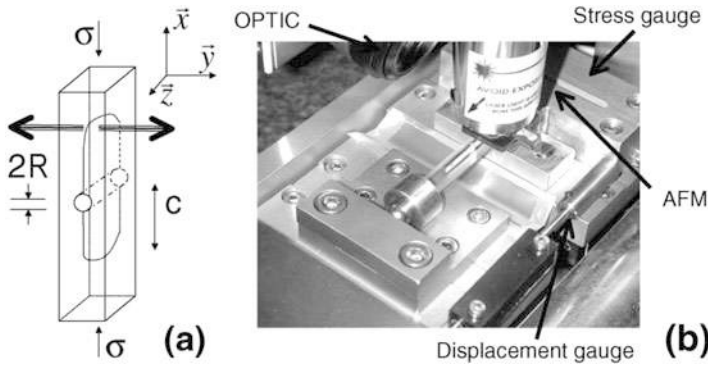
Fig. 2 (**a**) Sketch of the DCDC specimen. (**b**) Experimental setup for in-situ AFM imaging of the slow crack propagation in a DCDC specimen

crack tip displacement field and in turn determines the surface strain and strain rate fields. The variations of the direct topographical signal provide the out-of-plane displacements associated with crack propagation, while the access to the in-plane displacement field can be obtained through digital image correlation (see Insert D).

The in situ AFM technique allows accessing the process zone with unprecedented spatial resolution and real-time monitoring. However, experimental access to such a high spatial resolution entails two main constraints arising from the intrinsic nature of scanning probe microscopies. The first is that observations are limited to the free surface of the sample, which is subjected to a different mechanical condition than the bulk. The second is that the maximum imaging rate is presently limited to one image per minute, which limits the crack propagation velocity to the slow range between pm/s and nm/s. While the use of these techniques is limited to a range of temperature around ambient conditions, the composition of the atmosphere in a glove box containing the setup can be set with great precision to a given mixture of N_2, H_2O, and CO_2, which is a very important asset to study stress corrosion or more general environmental effects on crack propagation.

Figure 3 represents a typical image that was obtained in the process zone of a crack propagating in a glassy thermoplastic polymer (PMMA) where one can appreciate the richness of information on the detailed mechanisms at action. Before we can attempt any interpretation it is wise to fully examine the multiscale mechanical condition corresponding to the observed phenomenon.

The manuscript will be organized as follows. Chapter 2 presents the developments of the multiscale modeling of the DCDC crack propagation technique. Chapter 3 presents the well-established advances in modeling the stress-corrosion crack propagation in oxide glasses made possible by the in situ AFM technique. Chapter 4 presents recent extensions of the technique to investigate slow crack propagation in glassy polymers and very promising new observations. Chapter 5 drives a critical conclusion about the technique and its perspectives toward linking the worlds of nanoscale mechanics and numerical modeling. Four Inserts are distributed along the text to consistently provide with required base knowledge in fracture mechanics, atomic force microscopy, capillary condensation, and digital image correlation.

Fig. 3 3D representation of the topography of the crack tip region at the free surface of a PMMA sample measured by AFM
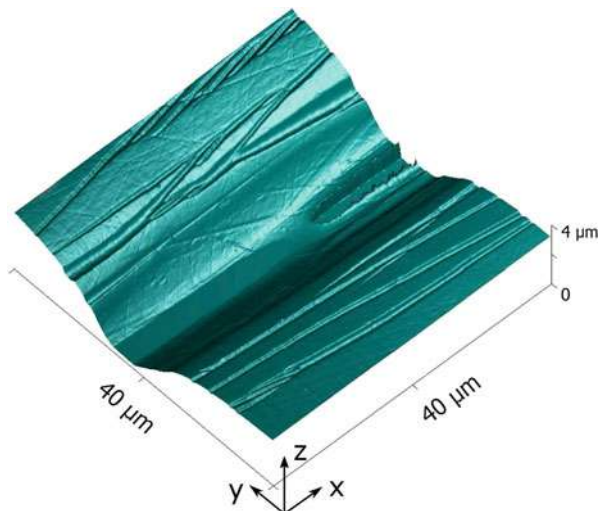
Insert B: Atomic Force Microscopy (AFM)

Atomic force microscopy is widely used to non-destructively characterize the surface structure of materials at the nanometric scale. In imaging mode, the AFM sharp probing tip, mounted on a flexible cantilever, is scanned over the surface of the material, and the interactions between the tip and the surface, measured through the deflection of the cantilever, are translated into a three-dimensional image of the surface topology.
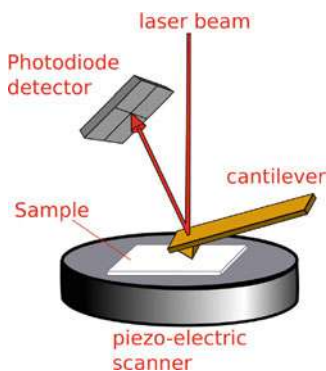
figure from Roduit (2010)

Topographical images can be obtained either in contact (static) mode or in dynamic mode (also called *tapping*), in which the cantilever is submitted

to a sinusoidal excitation. A feedback loop acting on the vertical position of the cantilever is used to keep the amplitude of the deflection (or of the oscillations) constant during the scan, in order to get the surface topography by recording the vertical position or *height signal*. A second kind of image, called the *error signal image*, represents the difference between the measured deflection (or oscillation amplitude) and its prescribed setpoint value. This image is very precious to highlight the contours of the structures in the imaged region. The vertical resolution of the AFM topographical images can be as small as 0.1 nm and is essentially limited by vibrational and electronic noise, whereas lateral resolution is essentially hampered by the finite radius of the scanning tip which induces a smearing of the measurement along the surface plane and is thus generally at best of the order of 10 nanometers. At the nanometric scale of the AFM images, it is important to keep in mind that differential thermal dilations between the sample and the instrument during a scan are generally not negligible, inducing image distortions (called *thermal drifts*) that must be minimized and taken into account.

When operating in dynamic mode, the phase delay between the sinusoidal excitation and the response of the cantilever can also be acquired during the scan. This provides a third kind of image, called *phase image*, which can contain useful information on the nature of the scanned surface. The phase signal is indeed related to the energy dissipated per cycle by the surface-tip interactions (Cleveland et al. 1998) and is generally used to reveal changes of the viscoelastic properties of the sample materials and changes of wetting properties of the sample surface or to detect the presence of nanoscale liquid film condensed on the solid substrate.

2   Multiscale Modeling of the DCDC Sample

The DCDC sample was initially developed by Janssen (1974) to study slow fracture in silicate glasses, which possess very high values for both Young modulus $E \approx 70$ GPa and yield strength $\sigma_y \approx 10$ GPa. The sample (Fig. 2a) consists of a prism of dimensions $2L \times 2w \times 2t$ (corresponding to x, y, z directions) with a cylindrical cross hole of radius R drilled through the specimen thickness $2t$. The sample is loaded with a compressive force F and thus a compressive stress $\sigma = F/4wt$ applied to the two opposite faces. This induces a tensile stress at the two poles of the central hole that are aligned with the sample loading direction. During the test in stiff and hard materials like glass, two symmetric cracks of length c are spontaneously nucleated at the crown of the central hole. The two cracks then propagate in opposite directions along the midplane of the sample (direction x in the $x - z$ plane), driven by the mode I opening induced at the crack tip.

In order to fit the DCDC sample into a compact loading setup to be set into an AFM apparatus, the glass sample dimensions are fixed to $40 \times 4 \times 4 \, \text{mm}^3$, with a hole radius $R = 500 \, \mu\text{m}$. A precision of $10 \, \mu\text{m}$ is required on all dimensions, and the surfaces should be polished to a RMS roughness of 0.5 nm (for an area of $1 \times 1 \, \mu\text{m}^2$) to insure proper AFM imaging.

In these very brittle materials, the macroscopic sample deformation is essentially linear elastic. The stress intensity factor has been derived by several authors via 2D finite element simulations (Janssen 1974; Michalske and Freiman 1983; He et al. 1995; Pallares et al. 2009). When the crack tip is at a sufficient distance ($>w$) from the central hole and the sample extremities, the mode I stress intensity factor is a decreasing function of the crack length c. The most refined analytical approximation for this relation is (Pallares et al. 2015):

$$\frac{\sigma \sqrt{\pi R}}{K_I} = \left[d_0 + d_1 \frac{w}{R} + d_2 \left(\frac{w}{R} \right)^2 \right] + \left[d_3 + d_4 \frac{w}{R} + d_5 \left(\frac{w}{R} \right)^2 \right] \frac{c}{R} \qquad (1)$$

where $d_0 = 0.3156$, $d_1 = 0.7350$, $d_2 = 0.0346$, $d_3 = -0.4093$, $d_4 = 0.3794$, and $d_5 = -0.0257$. Since $dK_I/dc < 0$, crack propagation is very stable at both imposed load and imposed displacement. Moreover, the directional stability of the straight crack propagation is excellent due to the strongly negative values of the T stress induced at the crack tip by the applied compression (Fett et al. 2005).

When using the DCDC sample to study subcritical propagation with a constant applied displacement, the crack progressively slows down as a consequence of the increase of the crack length. This allows to progressively reach very slow velocities without any activation of the loading stage. At the very low crack propagation velocities required by AFM observations (below 10 nm/s), the variations of the crack length c, and consequently of K_I, are very weak. A steady-state propagation in very pure asymptotic mode I loading ($K_{II}/K_I \leq 5\%$) can thus be observed over conveniently long in situ investigations: a few days or weeks depending on the crack propagation velocity. The details of the setup and techniques can be found in Celarie (2004) and Grimaldi et al. (2008).

We address here a first major issue related to the necessity of establishing the proper mechanical loading in the neighborhood of the micrometric region observed by AFM where the crack front intersects the external surface of the DCDC sample (cf. red circle in Fig. 4). While the 2D modeling of the DCDC sample assumes the invariance of stress and displacement fields along the z axis, the boundary effects on a finite thickness sample generally imply the crack front to assume a curved shape, which is parabolic for the DCDC sample (cf. top right image in Fig. 4). The 2D symmetry of the crack tip field is thus broken in the proximity of the external surface, which is intersected by the crack front with a non-normal angle $\alpha \sim 10°$. In the presence of such a tilted configuration, the singularity of the crack tip stress field takes a radically different mathematical form, called "corner point singularity," which should be expressed in terms of spherical coordinates around the intersection point (Benthem 1977; Bazant and Estenssoro 1979; Dimitrov et al.
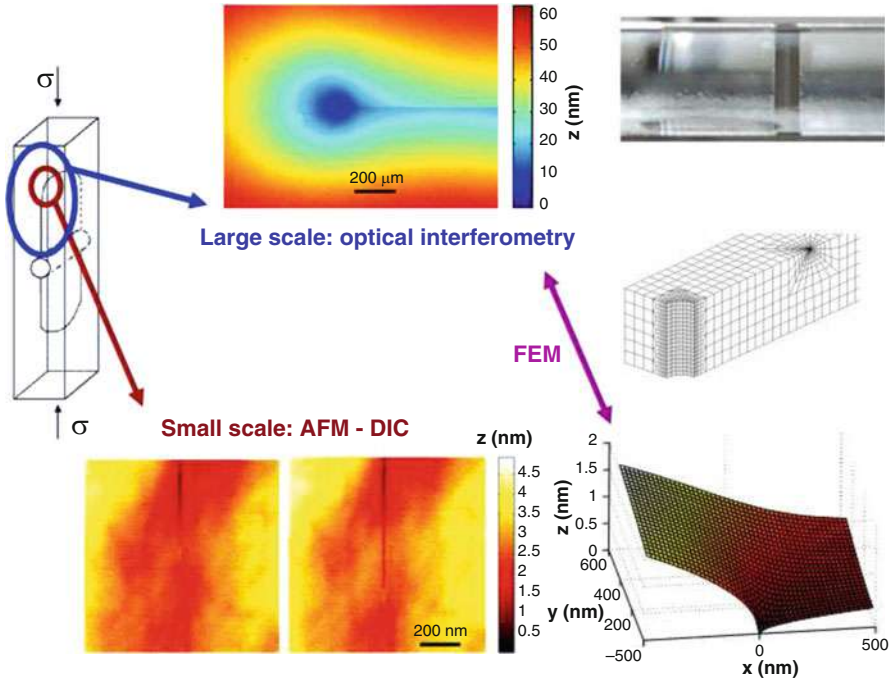
Fig. 4 Multiscale combined experimental and numerical analysis of the DCDC sample. The large-scale measurements by optical interferometry are illustrated on the top. The two small-scale AFM images in the bottom represent two successive steps of steady-state crack propagation that lead to the estimation of the local out-of-plane displacement field represented on their right. The finite-element mesh on the right is used to bridge the two scales

2006). The practical consequence is that the crack propagation at the surface may be different than in the bulk. However, if we insure a well-conditioned steady-state crack propagation, the crack front assumes a constant shape, and the intersection angle α can be shown to be a function of the material's Poisson ratio (Pallares et al. 2015). Under these specific conditions, the stress intensity factor takes an essentially constant value along the crack front and up to the external surface. Both the loading conditions and propagation kinetics become thus consistent between the surface and the bulk of the sample.

This scenario was validated as illustrated in Fig. 4 by combining (1) a 3D finite element modeling of the linear elastic deformation of a DCDC sample taking into account a curved crack front shape, (2) the experimental measurement by optical reflection interferometry of the elastic crack opening profile, and (3) the experimental measurement by optical profilometry of the out-of-plane displacement field of the external surface of the sample, which is the same observable measured in the in situ AFM investigation but at larger macroscopic scale.

The outcome of the excellent agreement between all these measured and calculated displacement fields is that:

(1) the crack opening profile in the midplane of the sample is well described by the Williams expansion (Williams 1957; Pallares et al. 2009):

$$u_y(X) = \frac{K_I}{E'} \sqrt{\frac{8}{\pi}} \left(X^{1/2} + \alpha_3 X^{3/2} + \alpha_5 X^{5/2} \right) \qquad (2)$$

where K_I is provided by Eq. (1), $\alpha_3 = 1.319/w$ and $\alpha_5 = 0.515/w^2$.

(2) the out-of-plane displacement field u_z at the external surface can be expressed as (Han et al. 2010; Pallares et al. 2009):

$$u_z(r, \theta) = \frac{K_I(1 + \nu)}{\sqrt{2\pi}E} \sqrt{r} \left(A + B \cos \frac{\theta}{2} + C \cos \theta \right) \qquad (3)$$

where K_I is again provided by Eq. (1) and the angular dependence is approximated by a truncated Fourier expansion (the coefficients are $A = 0.28$, $B = 0.67$, $C = -0.72$ for our sample configuration and mechanical properties). We remark that the fact that the $\sqrt{r}$ singularity for u_z is the same as for the in-plane crack opening profile is not a trivial result, and it is not predicted by 2D solutions neither in plane stress nor in plane strain.

3   Stress Corrosion in Oxide Glasses

Stress corrosion is a hydrolysis chemical reaction happening at the crack tip, which would normally be very slow on glass at ambient temperature, but is significantly accelerated due to local stress concentration (cf. Ciccotti 2009 for a recent review). Most ordinary oxide glasses are silicate glasses, where about 80 wt.% is constituted by a densely cross-linked network of silica tetrahedra, SiO_4, bonded by a common oxygen atom, thus creating $Si - O - Si$ (siloxane) bridges that locally form some rings of radii down to 0.5 nm. The basic corrosion reaction is a three-step reaction involving water adsorption on the $Si - O$ bonds, then the hydrolysis reaction itself (exchange of electron and proton), and finally the separation of the bond in two separate silanol groups ($Si - OH$), which correspond to one step advance of the fracture in the main silicate network of the glass (Fig. 5 Left). The vision of stress corrosion as a sequence of individual bond breakings by stress (and temperature)-enhanced hydrolysis was established in the 1970s (Wiederhorn 1967; Wiederhorn and Bolz 1970; Michalske and Bunker 1984). The kinetics of stress corrosion crack propagation can be represented by a three-region scheme on a graph describing the logarithm of crack velocity v (spanning from cm/s to pm/s) against the stress intensity factor (SIF) K_I (Fig. 5 Right), when K_I is smaller than the toughness K_{Ic} that is of the order of 1 MPa·m$^{1/2}$ for most glasses.

Region I is the proper stress-corrosion regime, where the stress enhances exponentially crack propagation velocity according to accelerated hydrolysis reactions. Region II is a plateau where crack velocity is limited by the migration of water molecules toward the very confined crack tip. Region III is a very steep region
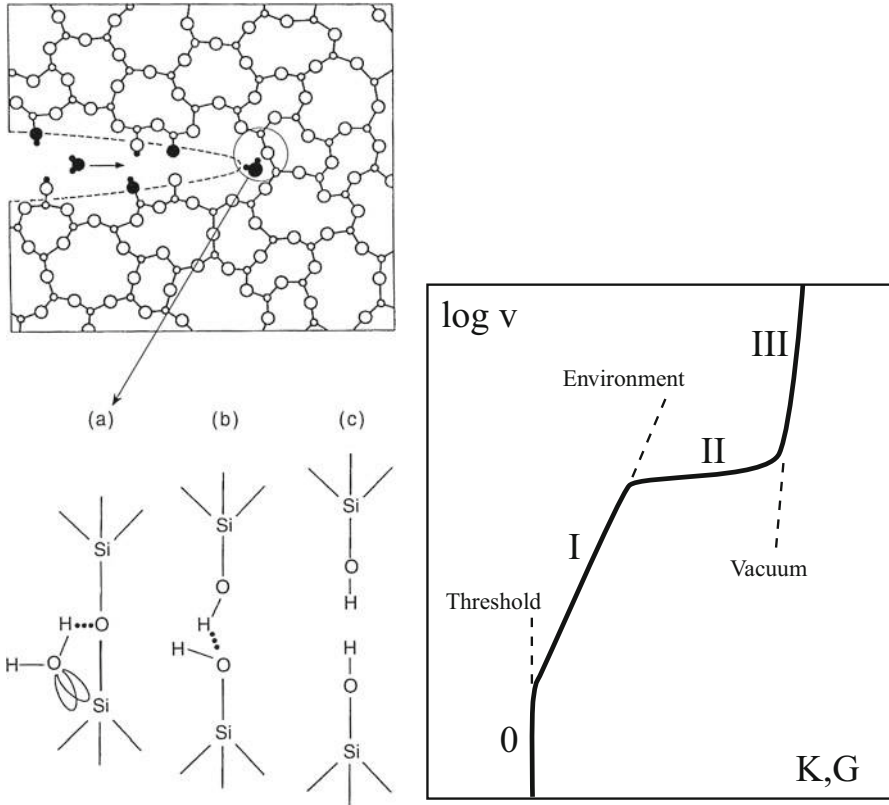
Fig. 5 Left: basic mechanisms of the stress-corrosion reactions (from Michalske and Freiman 1983). Right: Schematic $v(K)$ diagram for subcritical crack propagation in oxide glasses

where the SIF is so high that the crack can propagate without any contribution from water hydrolysis. In some glasses such as alkali-silicate glasses, an additional region 0 corresponds to a threshold for crack propagation. The phenomenology of the stress-corrosion regime (region I) consists of an exponential dependence of the crack velocity on the stress intensity factor, an almost proportional dependence on humidity, and an increase of crack velocity with temperature. This was experimentally established by Sheldon Wiederhorn (Wiederhorn 1967; Wiederhorn and Bolz 1970) who also proposed the following model equation:

$$v = A \left(\frac{p_{H_2O}}{p_0} \right)^m \exp\left(-\frac{\Delta E_a - bK_I}{RT} \right) \tag{4}$$

which is based on the chemical reaction rate theory for the hydrolysis of siloxane bonds by water molecules. The dependency on humidity p_{H_2O}/p_0, the main reactant, is a multiplicative power law with an exponent m close to 1, which is

the order of the reaction. The dependency on both temperature T and K_I appears through an Arrhenius term, where R is the gas constant and the activation barrier ΔE_a can be reduced by the presence of local tensile stress σ_{tip} at the crack tip of radius (we can write $\sigma_{tip} V_a = b K_I$, where V_a is an activation volume of molecular dimensions).

Phenomenological equations such as Wiederhorn's (Eq. 4) can describe the dependence of crack velocity on stress and on environmental parameters for most typical glasses and are compatible with a thorough consistent modeling based on the sharp-crack atomic-bonding paradigm (Lawn 1993). Yet the detailed nature of the stress-corrosion mechanisms that occur at the crack tip has been debated for decades (Marsh 1964; Maugis 1985; Gehrke et al. 1991; Tomozawa 1996), and a general disagreement can be found on the relevance of several accessory phenomena that may participate in the stress-corrosion mechanisms at different stages of the process. Stress corrosion can involve a complex interplay between the diffusion of reactive molecules (mainly water) into the crack cavity and into the glass network, the corrosion (or dissolution) of the network itself, and the migration of weakly bonded alkali ions under chemical or stress gradient (Gehrke et al. 1991; Bunker 1994). We note that in pure silica glass, the absence of ion migration can lead to significantly different local corrosion mechanisms. All three phenomena are typically very slow under ambient conditions in the unstressed material, but they can significantly accelerate in the highly stressed neighborhood of the crack tip, depending especially on the nature of the environment and of its confinement in the crack tip cavity.

3.1 Crack Tip Environmental Condition

In the classical picture of stress corrosion in moist air, the local environmental condition at the crack tip is assumed to be dry, and water molecules have to diffuse in the air phase along the sharp crack cavity before reaching the most strained bonds at the crack tip (Wiederhorn 1967). However, several speculations or indirect observations (Crichton et al. 1999; Lawn 1993) have suggested the presence of liquid phase due to capillary condensation inside the very sharp cracks in hydrophilic materials such as glass (see Insert C). According to theoretical estimations, the Kelvin radius for capillary condensation of water at ambient conditions is subnanometric (see Insert C) and comparable with the radius of stretched crack tips in glass (Wiederhorn 1967). If present, this capillary-condensed phase should be very small and difficult to detect with optical methods.

The phase-imaging technique associated with the in situ AFM observations (see Insert B) provided the first clear observations of this wet phase filling the crack cavity over a length of about 1 micron (Fig. 6). This allowed to establish that the crack tips in silicate glasses in moist atmosphere are invariably filled with a stable liquid phase, mainly composed by water. Since macroscopic failure properties of materials are essentially governed by mechanisms taking place in the vicinity of the crack tip, such a liquid condensate has a potential for playing an important role on
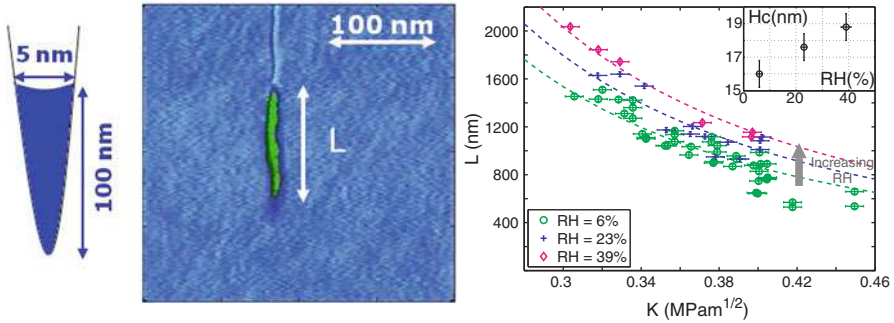
Fig. 6 Analysis of water condensate at the crack tip in silica glass: (left) Sketch of the coupling between the crack opening profile and the water condensate; (center) AFM phase image of the water condensate; (right) length of the condensate L as a function of the stress intensity factor K_I for different values of relative humidity. (Data after Grimaldi et al. 2008)

the propagation of the crack, in spite of its micrometric extension. First, it affects the chemistry allowing for crack propagation at low velocity by determining the local environmental condition for the stress-corrosion reactions at the crack tip. But it can also have a mechanical effect due to the highly negative Laplace pressure in the condensate (see Insert C) that acts on the crack walls and is responsible for the closure of the crack when the stresses become lower than some threshold K_0.

Insert C: Capillary condensation

Capillary condensation consists of the spontaneous condensation of a liquid water into a small gap confined between solid surfaces in the presence of an unsaturated moist atmosphere.
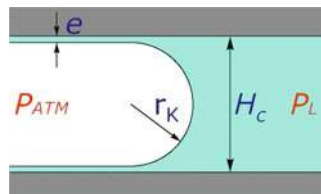
It is an ubiquitous phenomenon at nanoscale, occurring in all confined geometries, divided media, cracks, or contacts between surfaces (Bowden and Tabor 1950). Because of the curvature of the meniscus delimiting the condensation, the pressure in the two phases is not equal due to the *Laplace law* of capillarity:

$$P_L - P_{ATM} = \frac{\gamma_{lv}}{r}$$

where γ_{lv} is the liquid-vapor surface tension and r is the radius of mean curvature of the interface, the pressure being higher on the concave side. This induces a pressure drop in the pressure P_L of the liquid meniscus, which in turn induces a decrease of its chemical potential and thus allows to reach equilibrium with an undersaturated atmosphere when the average curvature radius equals the *Kelvin radius* r_k according to the *Laplace-Kelvin equation*:

$$\frac{\gamma_{lv}}{r_K} = -\rho_l k_B T \ln(RH) \tag{5}$$

where ρ_l is the number of molecules per unit volume in the liquid phase (Thomson 1871) and RH stands for the relative humidity.

In the case of perfect wetting, as for water on silica surfaces, the solid surfaces are also covered by an equilibrium condensed film of nanometric thickness e due to the effect of van der Waals short-range interactions between the surface and the water molecules through the condensed film (which is characterized by the *Hamaker constant A_{slv}*). In equilibrium conditions, the condensed film has the same negative pressure as the capillary-condensed phase, and the equilibrium thickness e of the film is given by:

$$e = \left(-\frac{A_{lsv}}{6\pi \rho_l k_B T \ln(RH)} \right)^{1/3} \tag{6}$$

In order to have capillary condensation, the gap H between two silica surfaces should be inferior to the critical value H_c (cf. Charlaix and Ciccotti 2010 for more details):

$$H_c = 2r_k + 3e \tag{7}$$

Silica glass is particularly suited to investigate the physicochemical properties of the capillary condensate since the stress-corrosion reaction does not involve corrosion products nor ion exchange phenomena. When measuring the length L of the condensate as a function of the stress intensity factor K_I (that controls the crack opening profile as seen in Insert A and Sect. 2), the condensation was shown to be at equilibrium with an atmosphere of moist nitrogen (Fig. 6), and the critical condensation distance H_c (see Insert C) was measured for several values of the relative humidity (Grimaldi et al. 2008).

When progressively unloading a crack, the condensate length L keeps increasing up to a maximum value $L_{\max}$, where the capillary forces overcome the crack opening forces associated to the macroscopical loading K_0. At this point, the crack is macroscopically observed to suddenly recede leading to complete closure. When

combining the measurement of the closure threshold K_0 with the $L - K_I$ curves by AFM phase imaging, the value of $L_{\max}$ can be adjusted as the single fitting parameter of the equation:

$$\frac{L}{L_{\max}} = \left(\frac{K_I}{K_0} - \sqrt{\left(\frac{K_I}{K_0} \right)^2 - 1} \right)^2 \tag{8}$$

and the values of both the Laplace pressure ΔP and the critical condensation distance H_c (see Insert C) can be determined by:

$$\Delta P = K_0 \sqrt{\frac{\pi}{8 L_{\max}}} \qquad H_c = \frac{K_0}{E'} \sqrt{\frac{8 L_{\max}}{\pi}} \tag{9}$$

This provided the first access to the physical properties of the wetting liquid at crack tip during stress-corrosion propagation. Concerning silica glass, these properties were shown to be very sensitive to impurities (Grimaldi et al. 2008) and to reach very stable values close to those of pure water in perfect wetting conditions when performing the experiment under a mixture of pure nitrogen and moisture (Pallares et al. 2011). Since the water condensate is in equilibrium with the moist atmosphere, the chemical activity of liquid water under the equilibrated negative Laplace pressure is the same as that of the neighboring moist air (see Insert C). This can explain why the classical stress-corrosion Eq. (4) well describes the dependency of the crack propagation data on relative humidity (Ciccotti 2009).

When studying more common glass compositions, such as soda-lime glass, the role of the crack tip condensation can be more pernicious due to stress-enhanced leaching that causes alkali ions to diffuse into the liquid condensate. The in situ AFM phase imaging reveals that the condensate length is significantly larger than for the silica glass (Célarié et al. 2007). Moreover, this length increases with time under constant loading and environmental conditions, without reaching a clear equilibrium. This reflects the change in composition of the liquid condensate due to the progressive enrichment in alkali ions that can be monitored by the AFM technique.

This unprecedented access to the composition of the nanoscale liquid condensate opens the way to more accurate investigations of two important factors affecting stress-corrosion kinetics: (1) the first major point is that this local environmental condition directly affects the rate of the corrosion reactions; (2) the second major point is that it allows for an estimation of the amount of alkali leached out of the bulk glass, which is responsible for a compressive stress buildup at the neighborhood of the crack tip (Wiederhorn et al. 2011), which also affects the stress-corrosion rate according to Eq. (4), and which was shown to cause the occurrence of a stress-propagation threshold in alkali-modified glasses (Fett et al. 2005).

Mastering the concomitant role of the local environmental condition and the compressive stress buildup induced by the interplay of hydration, leaching, and

hydrolysis reveal to be the necessary path toward a better modeling of the unsolved issue of the effect of glass composition on stress-corrosion kinetics in moist air (Wiederhorn et al. 2013).

3.2   Role of Plastic Deformation

In previous sections we used linear elastic fracture mechanics down to the molecular scale to interpret stress-corrosion mechanisms. A major objection to the sharp-crack atomic-bonding paradigm comes from the hypothesis of a significant contribution of plastic deformation at the crack tip in the slow crack propagation mechanisms. Since the early works of Dugdale (Dugdale 1960) on metals and the observations of plastic behavior of glass under compression in indentation and scratch marks (Taylor 1949), several investigations have been carried out to track the evidence of plastic behavior in the strong tensile stress field at crack tips in glasses.

The fracture energies measured by Wiederhorn (1969) for six glasses in an inert environment range between 7 and $10 \, J/m^2$, a figure that is tenfold higher than twice the typical values of the surface tension of glasses $\gamma = 0.5 \, J/m^2$ (Griffith 1920), which suggests a significant contribution of irreversible processes in crack propagation (see Insert A). By considering a typical yield stress of $\sigma_y \sim 10 \, GPa$ for glass, the size of the crack tip plastic region associated to such fracture energies by the Dugdale model would be limited to a few nanometers (see Insert A). In the subcritical stress-corrosion regime, the fracture energies and the estimated sizes of the plastic regions are even lower. However, the extent of penetration of reactive molecules from the environment into the glass network has been much debated (Marsh 1964; Tomozawa 1984), and this can enhance damage and plasticity on larger regions around the crack tip (cf. Ciccotti 2009).

Since the early development of AFM techniques, it has been possible to probe in situ crack propagation or the post-mortem crack surface morphologies to explore these mechanisms at their relevant nanometric scale, leading to remarkable observations (Guilloteau et al. 1996; Celarie et al. 2003; Bonamy et al. 2006) suggesting that crack tip plasticity would affect a region of about 100 nm, which is significantly larger than Dugdale's prediction. These have attracted a lot of attention on the nanomechanics of oxide glasses since they were thought to be connected with interesting out-of-equilibrium issues (Tomozawa 1996). Yet, a coherent interpretation of these AFM nanoscale investigations was only achieved after the development of the thorough multiscale analysis of the mechanics of fracture from the macroscopic scale where the loading is applied to the sample, to the submicrometric scale where the crack tip singularity lives (presented in Sect. 2).

When estimated at the scale of the in situ AFM images of the crack tip region, the displacements provided by the proper local elastic solution are of the order of the nanometer. While the in-plane displacements are below the lateral resolution of the AFM, the out-of-plane displacements are measurable, but they are of the same order
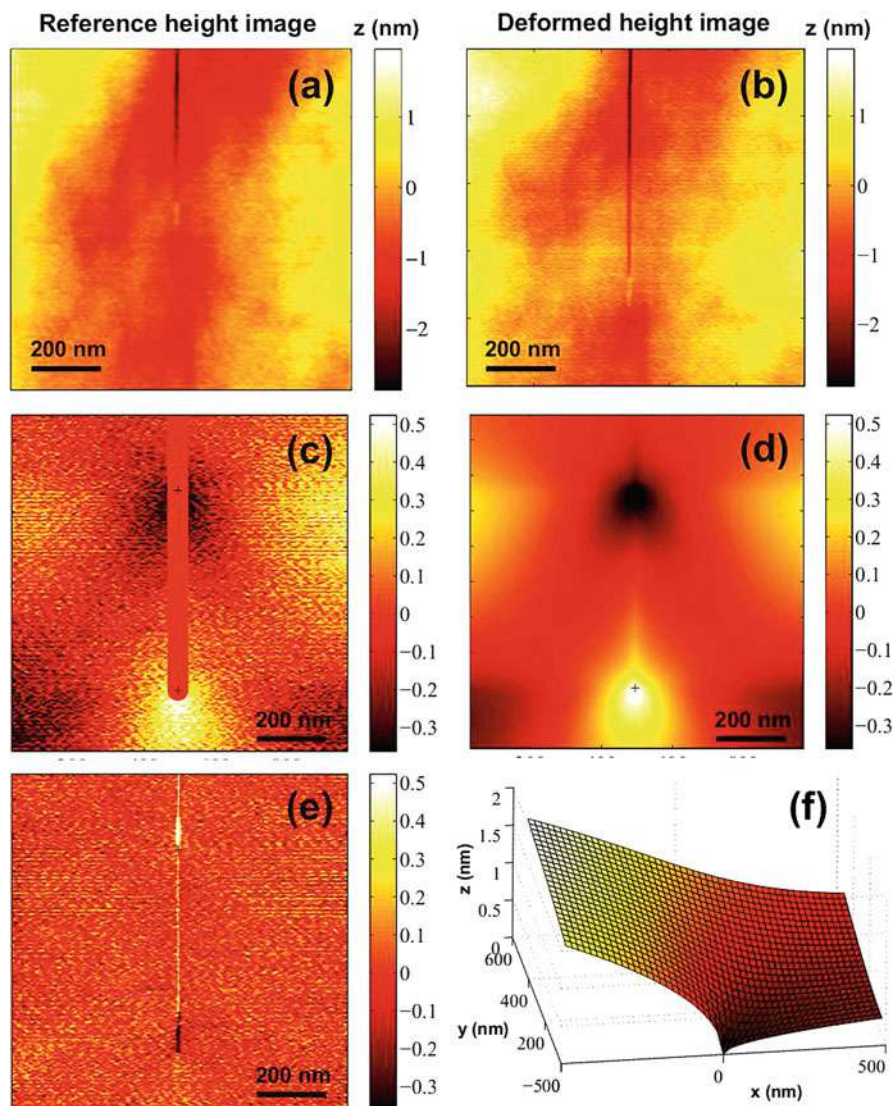
Fig. 7 (**a**) and (**b**) are, respectively, the reference and deformed AFM height images for the IDIC analysis. (**c**) Raw residual field obtained after correction of in-plane displacement. Note that the crack path is masked for the IDIC analysis. (**d**) Fitted displacement field. (**e**) Remaining residual after fit subtraction. (**f**) Out-of-plane displacement associated with the crack opening. A cut through the symmetry plane $y = 0$ is shown, with the crack tip located at the origin. (After Han et al. 2010)

of magnitude as the lowest surface roughness that can be obtained by polishing. In order to achieve an accurate measurement of both these displacement fields, an extension of the digital image correlation (DIC) technique to AFM topographical images was performed by Han et al. (2010) (see Insert D).

The results of this analysis (reported in Fig. 7) provide strong evidence that the proper 3D elastic solutions describe faithfully the mechanics of the DCDC sample from the macroscopic scale, where the load is applied down to the lateral resolution limit of the AFM measurements, represented by the 10-nm size of the probe. The values of the SIF K_I independently obtained by either the macroscopic mechanical analysis of the structural response of the DCDC sample or by the submicroscale measurement of the crack tip asymptotic displacement fields are in excellent agreement (within 5% uncertainty).

A complementary piece of information to the in situ investigations of fracture propagation is the post-mortem AFM analysis of the fracture surfaces. Concerning the quest for traces of plastic deformations, (Guin and Wiederhorn 2004) performed a careful recombination of the nanoscale topography of two opposite fracture surfaces, obtaining a perfect match down to the AFM height resolution of 0.1 nm. Bonamy et al. (2006) have looked for traces of plastic deformations by examining the post-mortem fracture surface profile in terms of crack propagation path. The fracture surfaces of brittle materials are known to possess self-affine statistical scaling properties that can be analyzed by estimating the lateral correlation functions. While the observed cutoffs between different scaling regimes have frequently been interpreted as signatures of changes from brittle to ductile fracture propagation regimes, an accurate metrologic analysis of AFM roughness images showed that for such small roughness amplitudes, the small-scale regime is likely to be caused by intrinsic bias in the AFM imaging process (Lechenault et al. 2010) and suggested that only the large-scale regime contains relevant metrologic information on the fracture surfaces. Pallares et al. (2017) showed that this large-scale regime is likely to be extended down to the relevant scale of material heterogeneities, which are close to molecular dimensions in glasses. Moreover, as shown by Wiederhorn et al. (2007) and Pallares et al. (2017), the amplitude of the surface roughness in the stress-corrosion regime is a decreasing function of the applied stress intensity factor K_I. This result, which has the opposite trend than what is observed during dynamic crack propagation (giving origin to mist, mirror, and hackle progressive roughening during instable crack acceleration), can be understood in terms of a model based on brittle crack propagation by LEFM down to the molecular scales, where the crack tip propagation is deflected by residual stress fields induced by heterogeneities in the local elastic moduli or in the local thermal expansion coefficients (Wiederhorn et al. 2007). This is a further coherent observation with the fact that stress-corrosion fracture is strongly coupled to residual stress fields and not significantly affected by crack tip plasticity. These very subtle issues constitute a playground where MD simulations could bring relevant modeling.

Insert D: Digital Image Correlation (DIC)

Digital image correlation is classically used to extract the in-plane displacement field $u(\mathbf{x})$ in a region of interest Ω between a reference image $f(\mathbf{x})$ and a deformed image $g(\mathbf{x})$, taken on the same surface at different times. $f(\mathbf{x})$ and $g(\mathbf{x})$ usually represent the gray level at the position $\mathbf{x}$ in the image. The *conservation of the optical flow* between the two images can be written as:

$$f(\mathbf{x}) = g(\mathbf{x} + \mathbf{u}(\mathbf{x})) \tag{10}$$

The in-plane displacement field $\mathbf{u}(\mathbf{x})$ can thus be found by minimizing the sum of squared differences:

$$\int_{\Omega} (f(\mathbf{x}) - g(\mathbf{x} + \mathbf{u}(\mathbf{x})))^2 \, d\mathbf{x} \tag{11}$$

To measure a non-integer displacement, an interpolation scheme of the grey level between pixels can be used. With AFM topographic images, DIC can be expanded to deal with displacements in three dimensions since $f(\mathbf{x})$ and $g(\mathbf{x})$ represent the vertical positions corresponding to each pixel. Equation (10) can be generalized as:

$$f(\mathbf{x}) = g(\mathbf{x} + \mathbf{u}(\mathbf{x})) + v(\mathbf{x}) \tag{12}$$

where $v(\mathbf{x})$ denotes the out-of-plane displacement field. In this case, the roughness of the two surfaces is used to obtain the in-plane displacement field, and then the subtraction of the registered images allows obtaining the out-of-plane displacement field by eliminating the contribution of roughness.

Roux and Hild (2006) have developed a more robust method, called *"Integrated" DIC approach* (IDIC), where kinematic or mechanical information are taken into account in the minimization procedure. The displacement field $[\mathbf{u}(\mathbf{x}), v(\mathbf{x})]$ is then sought as a linear combination of a few expected fields.

When treating the in situ AFM images, some of these fields are chosen to represent regular transformations such as uniform extension or shear caused by the AFM drifts, whereas the others represent the linear elastic fracture mechanics of in-plane and out-of-plane singular fields on each image. The steady-state crack propagation condition can be used to enforce the identity of the crack tip fields on the two images after translation along the crack path direction, thus reducing the number of unknowns and providing robust values for the stress intensity factor K_I.

4   Toughness of Glassy Polymers

4.1   About Fracture in Glassy Polymers

Glassy polymers are widespread in technical and structural applications since the 1930s due to their excellent mechanical properties, combined with low mass density and ease of implementation. Although their elastic moduli and yield strength are lower than other structural materials such as glasses or metals, glassy polymers possess a relatively high fracture energy. This results from a combination of plastic deformation and the occurrence of several localization mechanisms such as crazing and shear banding (Donald and Kramer 1982). Extensive studies have been devoted to the modeling of the relationship between the mechanical properties and the material composition and microstructure. Their main aim was combining contributions from both macroscopic constitutive laws, mesoscale damage mechanisms, and molecular issues (Williams 1984; Brown 1991; Réthoré and Estevez 2013; Halary et al. 2011).

   Although glassy polymers (i.e., with a glass transition temperature well above ambient temperature) can belong to different families with different macromolecular structures such as thermoplastics, thermosets, and semicrystalline polymers, their macroscopic mechanical properties are qualitatively quite similar up to yield condition. These materials have a typical elastic modulus of $E \approx 1\,\text{GPa}$, a yield strength of about $\sigma_y \approx 100\,\text{MPa}$, and a fracture toughness of about $K_c \approx 1\,\text{MPa}\sqrt{\text{m}}$. The size of the inelastic process zone at the crack tip is thus of the order of magnitude of the Dugdale length $\ell_D \sim 15 \div 40\,\mu\text{m}$ (see Insert A).

   Understanding and engineering the toughness of these polymers rely on the ability of describing and modeling the bulk deformation and energy dissipation mechanisms at the process zone scale. It is thus pertinent to use the same DCDC sample geometry and loading setup to perform in situ AFM investigations while dealing with the macroscopic scale of the sample by using the tools of linear elastic fracture mechanics. Although the materials of this class share similar properties on scales larger than ℓ_D, the nature of the inelastic processes occurring in the process zone can be very different depending on the nature and structure of the macromolecular network, which results in a very different large strain behavior. The kinetics of crack propagation can also be very variable, changing from very stable steady-state crack propagation over a wide range of velocities in thermoplastics to jerky stick-slip fracture propagation in thermosets.

   On the opposite of oxide glasses, the major dissipation mechanism in glassy polymers is provided by plastic deformation and acts on a more mesoscopic scale that is $1000\times$ larger. As a consequence, fracture propagation appears to be significantly less sensitive to the degree of cleanness and purity of the atmosphere, and most of current measurements are conducted under current atmosphere with controlled temperature.

The first applications of the in situ AFM technique have been focused on two prototypes of thermoplastic and thermoset glassy polymers, namely, PMMA and DGEBA-IPD epoxy resin. The fracture properties of these well-known glassy polymers were extensively investigated with optical and post-mortem investigations in the 1970s, which resulted in the proposition of several microscale models for their toughness (Kinloch and Williams 1980; Williams 1984). The implementation of in situ AFM techniques allows the visualization at the pertinent scales and is thus a very promising tool for stating which are the pertinent mechanisms at action for different polymers under well-defined crack propagation conditions.

For thermoplastics (such as PMMA) slow fracture propagation is a smooth process up to an instability leading to dynamic crack propagation. The fracture properties are described by a curve reporting the toughness $K_c(v, T)$ or the fracture energy $\Gamma(v, T)$ as a function of the steady-state crack propagation velocity v and temperature T (cf. Fig. 1). The quasi-static slow regime and dynamic regime are associated to the two positive slope branches of the $\Gamma(v)$ curve, which are separated by a hypothetic negative slope region where steady-state crack propagation is unstable and cannot be measured. The slow quasi-static regime extends over several decades of propagation velocities, and the $\Gamma(v)$ curve is dominated by viscoplastic dissipation in a crack tip process zone of the size of tens of micrometers. For PMMA at ambient conditions, the process zone has long been known to be constituted of crazes, i.e., of nanometric fibrils being nucleated and drawn across the crack lips thus providing extensive energy dissipation and crack propagation resistance (Doll 1983; Kramer 1983).

For thermosets (such as epoxy resins), fracture propagation is affected by stick-slip dynamics (cf. Fig. 8a). A static crack will not propagate up to an initiation SIF, named $K_{\text{start}}(\dot{\delta}, T)$, that is, a decreasing function of the macroscopic loading rate $\dot{\delta}$
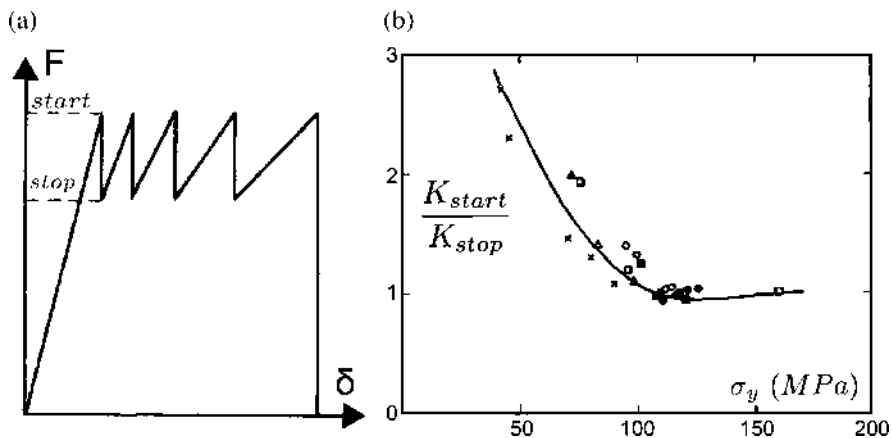
Fig. 8 Characteristics of the stick-slip behavior in thermosets. (**a**) The sample loading curve $F(\delta)$ presents a saw tooth behavior. (**b**) The stick-slip amplitude $K_{\text{Start}}/K_{\text{Stop}}$ diminishes with the yield stress σ_y or equivalently when increasing the loading rate $\dot{\delta}$ or decreasing temperature T

applied to the sample or of decreasing temperature T. Crack initiation is followed by dynamic propagation ($v \sim 100\,\text{m/s}$) that causes the sudden drop of the load observed in Fig. 8a. Eventually, the crack will stop for a lower value of K_I, called K_{stop}, which is quite insensitive to the rate of loading and temperature. As shown by Kinloch and Williams (1980), the coupled dependency of K_{start} on the loading rate and temperature can be expressed as a single master curve in terms of the yield stress σ_y taken at a reference strain rate and temperature (cf. Fig. 8b). As proposed by Phillips et al. (1978), based on poorly resolved optical observations, the stick-slip mechanism can be interpreted as the consequence of plastic blunting of the crack tip during the stick phase, followed by the nucleation of a sharp fast brittle fracture inside the stretched static process zone.

4.2   DCDC Techniques for Glassy Polymers

The application of the DCDC configuration to glassy polymers implies solving several technical issues since these materials are less stiff ($E \approx 1\,\text{GPa}$) and less hard ($\sigma_y \approx 100\,\text{MPa}$) than silicate glasses. Moreover, they present a slow viscoelastic relaxation at macroscopic scale. Since these materials are not brittle enough to provide a spontaneous crack initiation from the central hole, the crack initiation has to be implemented with the help of instrumented blades. Sample dimensions have to be optimized (George et al. 2018) in order to minimize buckling and yielding during crack propagation while preserving the small sample dimensions required by in situ AFM investigations (for PMMA we used $2L = 45\,\text{mm}$, $2w = 7.2\,\text{mm}$, $2t = 4.95\,\text{mm}$, and $R = 1.2\,\text{mm}$; for epoxy resin $2L = 40\,\text{mm}$, $2w = 8\,\text{mm}$, $2t = 4\,\text{mm}$, and $R = 1.33\,\text{mm}$). Under these conditions, the stable propagation of very straight cracks with slow velocities in the desired range for AFM investigation can be achieved on both our test polymers, presenting steady-state conditions over several days. We remark that for epoxy, this slow fracture regime was never observed before.

The estimation of the stress intensity factor acting at the crack tip is made considerably more difficult by the occurrence of stress relaxation, larger deformations, and plastic yield due to the shear stress concentration in the region of the central hole. A main point is that during slow crack propagation ($v < 10\,\text{nm/s}$ for AFM investigations), bulk stress relaxation in the DCDC sample is the dominating cause of the reduction of K_I during a test rather than crack propagation, which is very limited in extent. Although the linear elastic modeling of the whole sample (cf. Eq. (1)) is not any more adequate to provide K_I, an experimental measurement of the crack opening profile can be used to estimate K_I (cf. Eq. (2)) at the elastic scale surrounding the process zone (far from the central hole where bulk plastic yielding can occur). Since the accuracy of this technique is limited to 10%, it can only be used to set the average value of K_I during the test. However, since the crack length is not changing significantly during the test, we can use the recording of the measured applied force F to obtain the time evolution of K_I to be combined with the measurement of the crack tip velocity v performed by optical or atomic
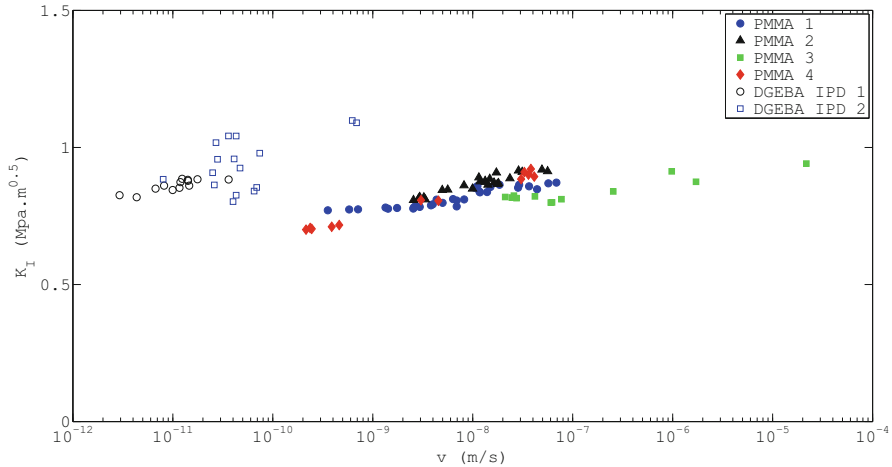
Fig. 9 $K_{Ic}(v)$ data for PMMA (four samples) and DGEBA-IPD epoxy resin (two samples) obtained by the present technique

force microscopy (cf. George et al. 2018 for more details). In order to provide a better control and resolution on the measurement of K_I, a full multiscale analysis of the DCDC sample under the proper elastoplastic constitutive laws is under way combining both experimental techniques and finite element modeling. However, the present level of control is sufficient to soundly interpret the in situ AFM observations that were recently obtained for the two polymers (George et al. 2018; Nziakou et al. 2019).

4.3 Slow Crack Propagation in Glassy Polymers

During our experimental campaign, we could obtain a successful crack initiation and well-conditioned steady-state crack propagation over different samples of PMMA and epoxy resin. With the estimation of K_I explained in the last section, $K_{Ic}(v)$ curves presented in Fig. 9 were obtained. For PMMA the $K_{Ic}(v)$ are in excellent agreement with data from the literature Doll (1983) within the expected 10% accuracy. For brittle epoxy resins, these curves are the first experimental evidence of the existence of a steady-state slow crack propagation regime. Propagation velocities for the epoxy resin were found to be below 10^{-9} m/s, which explains why they could not be detected by optical methods, while they are comfortably in the range of in situ AFM measurements.

The following section illustrates how in situ AFM imaging during steady-state crack propagation provides an unprecedented way to investigate the process zone at the crack tip in two polymers.

4.3.1 In-Situ AFM Observation in PMMA Samples

In the case of PMMA, optical microscopy revealed that steady-state crack propagation is accompanied by a whole range of localized yield and damage processes at the free surface of the sample, extending over a large neighborhood of the crack tip. In situ AFM measurements acquired in the crack tip region allow nevertheless to provide a much more accurate observation and quantification of this complex process zone (cf. Fig. 3). On the one hand, AFM images reveal an extended "necking" of the free surface of the sample (a few micrometer deep), inside which a sharp crack tip is propagating. On the other hand, a network of secondary minor surface grooves of very shallow depth and small width are superimposed to the main necking region. Note that the very fine lines that can be distinguished on the AFM images correspond to surface scratches due to polishing of the sample surface prior to the mechanical tests. The surface necking region at the free surface of the sample extends about 300 μm ahead of the observed crack tip, which is significantly larger than the typical 30–40 μm size generally observed for the craze region in the middle of the sample Doll (1983). However, an enhancement and potential modification of the mechanisms of crack plasticity at the free surface should be expected Kramer (1983). In order to proceed to a more systematic characterization of this extended process zone region over its whole size, we positioned the AFM scanning head in an initially unyielded region 300 μm ahead of a crack tip propagating at a few nm/s. A series of 150 AFM images of $40 \times 40\,\mu m^2$ size (the largest scan size providing sufficient spatial resolution) was then acquired every 15 min over a period of about 36 h without moving the scanning head position. The imaging ended after the crack tip had passed through the imaging region and traveled 300 μm ahead of it. The extreme stability of the steady-state propagation of the crack in our experimental setup allowed to maintain a well-conditioned crack propagation over the 36 h of measurements (average velocity $\langle v \rangle = 5 \pm 3$ nm/s). The full movie obtained from these successive acquisitions as well as the full-field reconstruction can be found in George et al. (2018).

For glassy polymers in situ topographic AFM images directly provide the vertical component u_z of the crack tip displacement field, which is much larger than surface roughness. The in-plane components (u_x, u_y) can be obtained by analyzing the progressive distortion of the dense array of the initially straight and crossed polishing lines, which can easily be identified on the AFM error images (see Insert B) as illustrated in Fig. 10.

The qualitative observation of the distorted polishing lines allows appreciating the smoothly changing nature of the strain field in the central necking region, opposed to the sharp transitions to a large uniform stretch of about 450% inside the secondary grooves (as estimated from the sharp changes of direction of the polishing lines).

In order to provide a straightforward measurement of the 2D projected strain tensor averaged over a gage length of about 2 μm in the main necking region, the change of length and orientation of the base vectors of a series of triplets (three
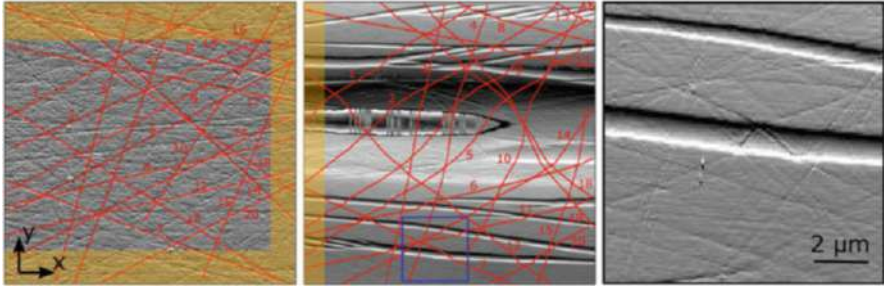
Fig. 10 Measurement of the in-plane strain field of the sample surface by tracking the changes of shape of the polishing line network. A limited series of remarkable lines has been highlighted in red and numbered in order to allow for an easier interpretation of the complex deformation field. The blue square box is zoomed in the right picture to highlight the sharp deflections of the polishing lines on the secondary grooves
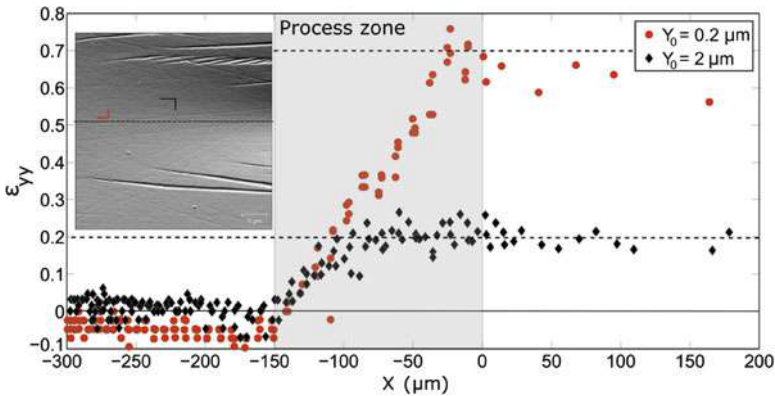
Fig. 11 In-plane nominal strain field $\varepsilon_{yy}(X)$ in the direction normal to the crack as a function of the distance X from the crack tip and for two values of the initial offset Y_0 from the central axis of the main necking region

intersection points between polishing lines) are monitored through the whole image series as a function of the position center of the gage region. Figure 11 presents the evolution of the normal strain component $\varepsilon_{yy}(X)$ for two such triplets centered at different distances from the crack axis ($Y_0 = 0.2$ and $2\,\mu$m). For both curves, the nominal strain presents a significant rise starting at about the same distance of $150\,\mu$m from the crack tip and then follows an approximately linear trend with distance. While for $Y_0 = 2\,\mu$m, the strain saturates at 20% at a $100\,\mu$m distance from the crack tip, for the triplet along the crack axis ($Y_0 = 0.2$), the plastic strain keeps rising until the crack tip reaches a large figure of about $\varepsilon_{yy} = 70\%$. In both cases, the plastic strain remains frozen behind the crack tip. The longitudinal strain component $\varepsilon_{xx}(X)$ did not present any significant trend out of the 5% scatter that

constitutes the detection limit. This manual image correlation technique is thus very efficient for identifying the plastic process zone and evaluating the plastic strain field inside it.

This very rich surface fracture pattern is quite different from the well-known crazing behavior that occurs at the crack front in the bulk of the PMMA samples where plane strain conditions are prevailing Doll (1983); Kramer (1983). The analysis of the surface displacement field associated with crack propagation, together with post-mortem investigation by scanning electron microscopy (SEM), leads to the interpretation of these observations as a change of fracture mechanism from crazing in the bulk to shear yielding in a surface layer, where the prevailing plane-stress conditions involve a local reduction of stress triaxiality. The thickness of the layer affected by this change of mechanism is about $10\,\mu m$, which is the same order of magnitude as the Dugdale length in plane strain as argued by Williams (1984).

The maximum plastic stretch of 70% reached close to the crack tip is comparable to the maximum plastic extension obtainable in uniaxial tensile tests on PMMA close to its glass transition temperature (Gsell and Jonas 1979). This observation is a nice proof that although PMMA is macroscopically brittle in tension at ambient temperature, it can be plastically cold drawn at scales smaller or comparable with the Dugdale length. The limiting extension is provided by the entanglement network, which still resists plastic flow before the material is broken.

The complex pattern of secondary grooves that develops along the main necking region was attributed to secondary surface crazing by carefully analyzing the real-time development of the surface strain fields (George et al. 2018). These secondary surface crazes are shown to be formed in the first half of the main necking region, where only elastic tensile deformations are prevailing with a principal direction orthogonal to the fracture plane. The secondary crazes are shown to be nucleated on these polishing lines (cf. movie in the SI of George et al. (2018)) that are more closely orthogonal to the surface tensile stress and which act as surface defects for craze nucleation.

4.3.2   In-Situ AFM Observation in DGEBA-IPD Epoxy Resin

In the case of epoxy resin, the neighborhood area of the crack tip presents a very different organization as can be seen in Fig. 12. The absence of both the very extended surface necking region and of the secondary grooves is to be noticed. The surface displacement field rather appears as a very smooth shear yielding field with a clear strain concentration toward the crack tip. This process zone mechanism is essentially the same as the one expected for the crack propagation in the bulk, even if probably enhanced in size due to the surface loss of triaxiality (Williams 1984). The absence of crazing in epoxies is generally attributed to the high cross-linking density (Halary et al. 2011).

By applying the same triplet analysis described for PMMA along the crack propagation axis, we can clearly identify a plastic zone starting at $7\,\mu m$ from the crack tip, where the plastic strain increases almost linearly up to a maximum of about 30%. These measurements are compatible, respectively, with Dugdale
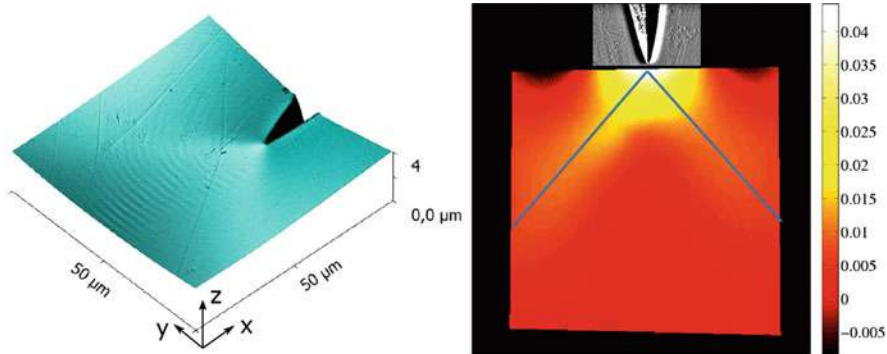
Fig. 12 (Left) 3D representation of the topography of the crack tip region $(50 \times 50\,\mu\text{m}^2)$ measured by AFM at the free surface of an epoxy resin sample. (Right) Strain field ε_{yy} between two $50\,\mu\text{m}$ size AFM images separated by crack propagation of $1\,\mu\text{m}$ as obtained by DIC (after Nziakou et al. 2019)

estimation of the process zone size and with the maximum strain allowed by the finite stretchability of the tightly cross-linked polymer network. In the presence of such a smooth strain field, it is possible to compute the full strain field between successive steps of crack propagation by digital image correlation (see Insert D) as represented in Fig. 12(right). This can be interpreted either as the gradient of ε_{yy} along the direction of crack propagation (if divided over a gage length of $\Delta X = 1\,\mu\text{m}$, corresponding to the distance traveled by the crack between the two images) or as the strain rate (if divided over the time lag of $\Delta t = 1.5$ days).

This analysis reveals that under steady-state propagation condition, the crack tip strain field presents two highly active regions forming an angle of about 38° to the crack propagation axis. We can remark that this picture is quite different from the crack tip fields obtained for the monotonic loading of a non-propagating crack. This information is very precious when modeling the difference between the fracture energy in steady-state and stick-slip crack propagation.

4.4 Link Between Steady-State and Stick-Slip Crack Propagation

The in situ AFM investigation with nanometric resolution has shown for the first time that for brittle thermosets like epoxy resin, a steady-state slow crack propagation behavior can be observed pretty in the same way as in thermoplastics. We also provided the measurement of the $K_{Ic}(v)$ curve over the accessible velocity range $[10^{-12}\text{–}10^{-9}]$ m/s in a well-conditioned steady-state regime. This phenomenon was never detected before because the maximum velocity v_c attainable in this slow velocity regime is in the range of nm/s, which is comparable with the growth rate of human nails. While this low velocity was very difficult to measure by optical techniques available in the 1970s, our AFM in situ investigation can accurately measure propagation velocities down to pm/s.

Fig. 13 Classical representation of stick-slip dynamics as a cycle (in red) on the $K_{Ic}(v)$ curve of steady-state fracture propagation
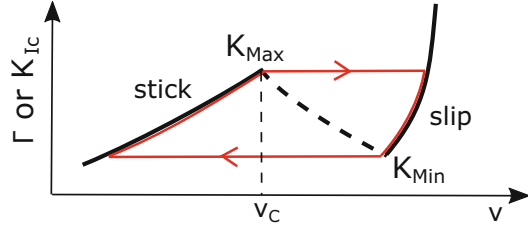
An important complementary observation by Takahashi and Arakawa (1984) has allowed to characterize the dynamic crack propagation in epoxy resins as a $K_{Ic}(v)$ curve with v systematically larger than 100 m/s, which is difficult to be measured during stick-slip dynamics, even with a fast camera. On the other hand, Hattali et al. (2012) have recently developed an advanced loading fixture allowing to obtain very high loading rates, and they have shown that thermoplastics like PMMA can also present stick-slip dynamics under these rapid loading conditions.

The combination of all these observations suggests that the steady-state crack propagation curve $K_{Ic}(v, T)$ and the characteristic stick-slip curve $K_{start}/K_{stop}(\dot{\delta}, T)$ are two complementary facets of a unified crack propagation physics for all glassy polymers, in an analog way to what is observed in the peeling of adhesive tapes (Maugis 1985).

However, the link between these two curves is not trivial. The classical picture to link the $K_{Ic}(v)$ with stick-slip dynamics is to identify K_{start} with the peak value K_{max} at the end of the slow quasi-static branch of the $K_{Ic}(v)$ curve and K_{stop} with the minimum value K_{min} of the dynamic branch (cf. Fig. 13), where the change of slope in the $K_{Ic}(v)$ curve induces a crack propagation instability (Maugis 1985). However, this picture cannot account for the reduction of K_{start} with the sample loading velocity $\dot{\delta}$ as represented in Fig. 8 for epoxy resins.

The in situ AFM observations on epoxy resin have recently lead us to propose a unified scenario that links the $K_{Ic}(v)$ and $K_{start}/K_{stop}(\dot{\delta})$ curves through the characterization of the viscoplastic behavior of the glassy polymer as well as the measured strain fields associated to crack propagation (Nziakou et al. 2019).

The key point is that steady-state fracture curve $K_{Ic}(v)$ can only be established after the crack has traveled several times the process zone size R_{ZP} at the same constant velocity v (cf. Fig. 14). We can thus define a characteristic minimum time for steady state:

$$t_{PZ} \sim \frac{R_{PZ}}{v} \tag{13}$$

When observing crack propagation over a longer time $t \gg t_{PZ}$, the steady-state crack tip strain field is well established, and it advances at a constant velocity v. The material's elements in the process zone are thus submitted to a plastic strain from an initial yield value ε_y (that we assume to be small) to a maximum plastic strain ε_{max}, which we assume to be related to the maximum strain sustainable by

Fig. 14 Sketch of the
plastically blunted crack tip
with crack tip radius ρ. The
red region represents the
process zone with size R_{ZP}
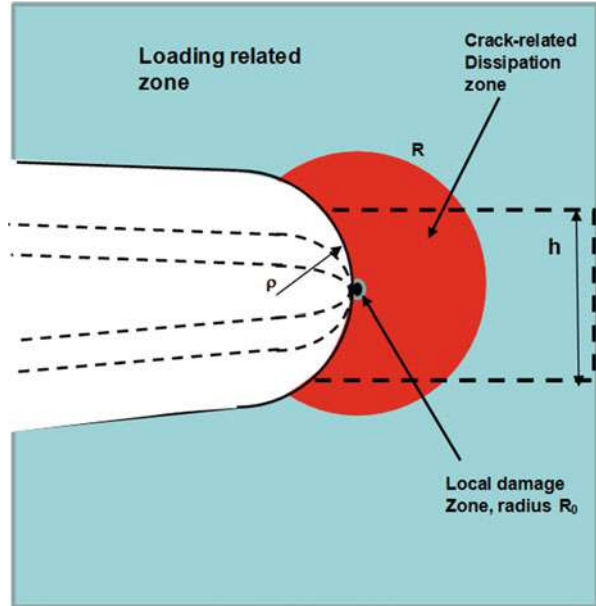
the polymer network in the process zone (and that we can measure by the in situ
AFM measurements). This defines a characteristic strain rate in the process zone
under steady-state crack propagation:

$$\dot{\varepsilon}_{\text{Steady}} \sim \frac{\varepsilon_{\max}}{t_{PZ}} \sim \frac{\varepsilon_{\max} v}{R_{PZ}} \tag{14}$$

The typical value for the volume density of dissipated work in the process zone
is thus:

$$\mathscr{W}_{\text{Steady}}(\dot{\varepsilon}_{\text{Steady}}) \sim \int_0^{\varepsilon_{\max}} \sigma(\varepsilon, \dot{\varepsilon}_{\text{Steady}}) d\varepsilon \tag{15}$$

where the elastoplastic constitutive behavior of the polymer $\sigma(\varepsilon, \dot{\varepsilon}_{\text{Steady}})$ can be
approximated by uniaxial tests performed in the corresponding strain rate range.
We remark that while glassy polymers are brittle under traction at macroscopic
scale, their behavior in the micrometric process zone was shown to experience large
stretches similar to the ones observed in compressive stretches in the same material.
We thus use uniaxial compression tests to estimate the pertinent elastoplastic
behavior by assuming that the strain hardening functions remain the same.

We can thus estimate the steady-state fracture energy $\Gamma(v)$ by multiplying this
typical volume density by the volume of the process zone bR_{ZP}^2 and dividing by
the cross-sectional area bR_{ZP}, where b is the transverse dimension of the sample,
which results in:

$$\Gamma(v) \sim R_{ZP} W_{\text{Steady}}(\dot{\varepsilon}_{\text{Steady}}) \sim R_{ZP} \int_0^{\varepsilon_{\max}} \sigma(\varepsilon, \dot{\varepsilon}_{\text{Steady}}) d\varepsilon \qquad (16)$$

We remark that this rough modeling is only based on measured quantities and it does not possess any adjustable parameter. When substituting typical values of the quantities measured for glassy polymers, we can obtain $\Gamma \sim R_{ZP} \sigma_y \varepsilon_{\max} \sim 10\,\mu\text{m} \cdot 100\,\text{MPa} \cdot 0.3 \sim 300\,\text{J/m}^2$, which provides the good order of magnitude. In principle, R_{ZP} and $\varepsilon_{\max}$ depend on the crack velocity v through the strain rate $\dot{\varepsilon}_{\text{Steady}}$. However, glassy polymers are only weakly viscoplastic, and the process zone size and crack opening displacement in PMMA were observed to be only weakly variable over many decades of velocity (cf. Doll 1983 and Fig. 1). We can thus expect that the weak slope of the $\Gamma(v)$ curves should be related to the weak dependency of the constitutive behavior $\sigma(\varepsilon, \dot{\varepsilon})$ over the strain rate $\dot{\varepsilon}$. This hypothesis has been shown to be consistent with data by Nziakou (2015).

We now turn to explain the link of this modeling with the crack initiation toughness Γ_{Stick} in the stick-slip dynamics, by considering the loading of a sample with a constant velocity $\dot{\delta}$ applied at the loading points. In LEFM, the sample can be characterized by its elastic compliance that depends on the crack length, so that we can write a linear relation $K_I = B(c)\delta$, where $B(c)$ can depend on the structure of the specific sample and can be calculated by FE modeling. Since the crack propagation is only weakly changing during the sample loading, the rate of increase of the stress intensity factor $\dot{K}_I$ is thus proportional to $\dot{\delta}$, and the sample-dependent loading time can be written as:

$$t_{\text{Stick}} = \frac{K_{\text{Start}} - K_{\text{Stop}}}{\dot{K}} \sim \frac{K_{\text{Start}} - K_{\text{Stop}}}{\dot{B}(c) \cdot \dot{\delta}} \qquad (17)$$

Now, if this loading time is shorter than the time t_{PZ} that is necessary for a crack to cross the process zone and reach steady state, we can consider that the strain rate in the process zone is no longer determined by the crack velocity v but rather from the loading rate $\dot{\delta}$ according to:

$$\dot{\varepsilon}_{\text{Stick}} = \frac{\varepsilon_{\text{Start}} - \varepsilon_{\text{Stop}}}{t_{\text{Stick}}} \sim \frac{\varepsilon_{\max}}{K_{\text{Start}} - K_{\text{Stop}}} B(c) \cdot \dot{\delta} \qquad (18)$$

By assuming that the mechanism of plastic blunting and crack reinitiation in the process zone remain similar, so that R_{ZP} and $\varepsilon_{\max}$ assume similar values for the same values of the local strain rate $\dot{\varepsilon}$, we can write a similar prediction for the fracture energy $\Gamma_{\text{Start}}(\dot{\varepsilon})$:

$$\Gamma_{\text{Start}} \sim R_{ZP} \int_0^{\varepsilon_{\max}} \sigma(\varepsilon, \dot{\varepsilon}_{\text{Stick}}) d\varepsilon \qquad (19)$$

where this time the strain rate is determined by the loading rate $\dot{\delta}$ in a sample dependent way as in Eq. (18).

The transition between steady state and stick-slip dynamics is thus related to the value of the loading rate $\dot{\delta}$. If we consider the maximum velocity v_C for which stable steady state can be obtained, this defines a minimum duration of the loading time:

$$t_{\text{Stick}} \sim \frac{K_{\text{Start}} - K_{\text{Stop}}}{B(c) \cdot \dot{\delta}} > t_{PZ} \sim \frac{R_{PZ}}{v_C} \tag{20}$$

And this corresponds to a maximum loading rate $\dot{\delta}_C$ to attain steady-state propagation before initiating stick-slip:

$$\dot{\delta} < \dot{\delta}_C \sim \frac{K_{\text{Start}} - K_{\text{Stop}}}{B(c)} \frac{v_C}{R_{PZ}} \tag{21}$$

This estimation is consistent with the fact that v_C is about 1 cm/s for PMMA, while it is only 1 nm/s for epoxy resins. The critical loading time for PMMA is 3 ms, while it is 8 h for epoxy. Since loading times in a typical fracture test are about minutes, it is sound that we generally obtain steady state in PMMA and stick-slip in epoxy. During our in situ AFM measurements, the loading stage is arrested after a single stick-slip event, and we can thus comfortably explore slow steady-state fracture in epoxy for several days. On the other hand, the measurements of Hattali et al. (2012) have proven that PMMA can undergo stick-slip dynamics under very fast loading, which is also consistent with the experience of brittle behavior of notched PMMA samples under impact loading.

We remark that the conditions given above imply that the strain rates explored in the process zone during stick-slip fracture are larger than the ones explored during the whole steady-state curve $\Gamma(v)$ up to v_C. Since the dependency of K_{Start} on $\dot{\delta}$ is negative, we can relate the transition velocity v_C to the critical velocity for which the steady-state fracture energy, calculated with the steady-state Eq. (16), would change to negative slope.

On the one hand, this is consistent with the classical LEFM modeling relating the stick-slip to a change of slope in the $\Gamma(v)$ curve (Maugis 1985). On the other hand, the negative branch of the $\Gamma(v)$ curve is not defined since steady-state fracture is unstable. The good way to understand the transition is to start from the measured large strain constitutive behavior $\sigma(\varepsilon, \dot{\varepsilon})$ and to determine the change of slope of the energy density $\mathscr{W}(\dot{\varepsilon})$ as a function of strain rate according to Eq. 15.

As a conclusion, the steady-state curve $\Gamma(v)$ and the stick-slip curve $K_{\text{Start}}(\dot{\delta})$ are two different expressions of the same process zone mechanism. Both can be related to the measured constitutive behavior $\sigma(\varepsilon, \dot{\varepsilon})$, but in the first case, the pertinent strain rate of the process zone is determined by the crack propagation velocity v, while for the second it is determined by the loading rate $\dot{\delta}$.

5   Conclusions

The present multiscale investigation offers a consistent mechanical view of the steady-state subcritical fracture propagation in a DCDC specimen for both oxide glasses and glassy polymers. The 3D analysis of the crack tip field is shown to be an unavoidable step to link the macroscopic features of the loading and the nanoscale asymptotic crack tip field as measured at the accessible free surface of the sample through in situ AFM observations. One key of the analysis is the quality of the steady-state crack propagation. By enforcing several hours of equilibrated slow crack growth in a very stable sample such as DCDC, the crack front progressively finds a shape that remains invariant throughout the entire propagation, equilibrated through all scales and under a very pure mode I loading (thanks to the principle of local symmetry). In this frame, observations performed at the nanoscale during slow crack propagation provide rich informations on the local environment of the crack tip, on the mechanical behavior of micrometric neighborhood of the crack tip, and finally on the mechanisms driving its propagation.

Our in situ AFM investigation technique allowed to deepen the understanding of the stress-corrosion mechanisms in oxide glasses. By using AFM phase imaging, we showed that the crack tip cavity is so sharp that it is filled with a spontaneous water condensation from the ambient moist atmosphere in a submicrometric neighborhood of the crack tip (Grimaldi et al. 2008). While this condensate was shown to be close to pure water at condensation equilibrium in pure silica glass (Pallares et al. 2011), complementary observations on alkali-silicate glasses (Célarié et al. 2007) have shown that the chemical composition of this condensate keeps evolving in time due to the leaching of sodium out from the glass. These observations are of the greatest importance to the understanding of the stress-corrosion kinetics under ordinary moist atmosphere since this condensate constitutes the effective local environmental condition at the crack tip region where the stress corrosion occurs. In silicate glasses, we also proved that the use of continuum mechanics and linear elasticity was valid down to 10-nm scales from the crack tip. Full 3D elastic solutions were used here as a first reasonable guess to establish a connection from the macroscopic scale down to the nanometric neighborhood of the corner point singularity. When these corner point solutions are tested against the in situ AFM submicrometric images by digital image correlation, the residual fields present extremely low values and especially no observable trends out of a 10-nm region from the crack tip, thus providing a striking evidence of the appropriateness of these elastic solutions to describe the crack tip fields down to the extreme lateral resolution of AFM imaging. This set of measurements constitutes a clear and sound solution to the long debate on the length scales of the inelastic processes acting at the crack tip of glasses during stress-corrosion slow crack propagation (Wiederhorn et al. 2013). We can conclude our discussion on oxide glasses by stating that the combination of the investigations of crack tip stresses along with the understanding of the local environmental condition is likely to offer an undisputable key to a thorough modeling of the role of glass

composition on the measured phenomenological parameters of the stress-corrosion curves (Wiederhorn et al. 2013). Another major perspective comes from the recent outstanding developments in the performances of molecular dynamics toward the simulation of stress-corrosion reactions on space and time scales that could soon approach from below the limits of the measurable range accessible by our in situ AFM experimental investigations (DuCormack and Cormack 2005; Kermode et al. 2008; Mischler et al. 2005), thereby opening a formidable opportunity for a novel experimental/modeling dialog.

Concerning glassy polymers, the first important required achievement is the ability to properly initiate and propagate a very well-conditioned mode I fracture and to ensure a stable steady-state condition over several days in order to allow for an extensive campaign of in situ AFM observations. The second achieved step is to obtain a valuable estimation of the crack driving conditions (in terms of a stress intensity factor K_I) even when the DCDC sample undergoes some bulk plastic yielding in the central hole region due to low strength/toughness ratio. This is assessed by the agreement of the measured $K_{Ic}(v)$ crack propagation curves with the ones reported in the literature. These two achievements are a necessary condition to link the microscopic scale observations of the crack tip mechanisms with the macroscopic scale of the sample loading. The acquisition of long AFM image series during several days of slow steady-state crack propagation allows for a rich and quantitative real-time evaluation of the damage mechanisms and the related strain fields at the free surface of the sample in a region that is progressively crossed by the crack tip process zone. While the out-of-plane displacement field is directly provided by the AFM topographic images, the in-plane displacement field can be obtained by tracking the relative motion of remarkable surface points or by applying full digital image correlation when possible. Our results obtained on typical thermoplastic and thermoset polymers confirm the general applicability of the technique to different glassy polymers and allow to appreciate a very different organization of the plastic strain fields in the process zone.

The AFM observations presented for PMMA reveal a very rich and unexpected surface fracture pattern, which is quite different from the well-known crazing behavior that occurs at the crack front in the bulk of the PMMA samples where plane strain conditions are prevailing (Doll 1983; Kramer 1983). The analysis of the surface displacement field associated with crack propagation leads to the interpretation of these observations as a change of fracture mechanism from crazing in the bulk to shear yielding in a surface layer, where the prevailing plane-stress conditions involve a local reduction of stress triaxiality. While changes in the process zone size are well known to occur in plane-stress surface layers (Williams 1984), our in situ AFM investigation was the first to provide a clear observation of a global change of mechanisms at the surface of a thick sample. The surface shear necking region presents a tenfold increased length with respect to the craze in the bulk of the sample. The plastic strain field is shown to affect only half of this length, and it presents a very smooth evolution up to 70% plastic extension close to the crack tip. This is comparable to the maximum plastic extension obtainable in uniaxial tensile tests on PMMA close to its glass transition temperature (Gsell and Jonas

1979). This observation is a nice proof that although PMMA is macroscopically brittle in tension at ambient temperature, it can be plastically cold drawn at scales smaller or comparable with the Dugdale length. The limiting extension is provided by the entanglement network, which still resists plastic flow before the material is broken. The complex pattern of secondary grooves that develops along the main necking region can be attributed to secondary surface crazing by carefully analyzing the real-time development of the surface strain fields.

The in situ AFM observations also revealed for the first time the existence of a steady-state slow crack propagation regime for an epoxy resin at velocity below 1 nm/s. Local measurements of the strain field showed that epoxy resin can also be plastically cold drawn, at scales smaller or comparable with the Dugdale length, up to an extent comparable to the maximal reachable plastic stretch allowed by the crosslink network. These results are the first steps for the development of a model unifying steady-state crack propagation curve $K_{Ic}(v, T)$ and the characteristic stick-slip curve $K_{\text{start}}/K_{\text{stop}}(\dot{\delta}, T)$ as two complementary facets of the same crack propagation physics for all glassy polymers.

This novel technique is thus very promising to achieve important insights on the effect of the different macromolecular structures of glassy polymers on their fracture toughness. It opens also the way toward well-conditioned nanoscale experimental measurements that can be confronted to molecular modeling. A very promising direction is the investigation of the effect of confinement on crack propagation in composite materials when the characteristic size of the structure is comparable to the size of the process zone.

References

Bazant ZP, Estenssoro LF (1979) Surface singularity and crack propagation. Int J Solids Struct 15:405–426

Benthem JP (1977) State of stress at the vertex of a quarter-infinite crack in a half-space. Int J Solids Struct 13:479–492

Bonamy D, Ponson L, Prades S, Bouchaud E, Guillot C (2006) Scaling exponents for fracture surfaces in homogeneous glass and glassy ceramics. Phys Rev Lett 97:135504

Bowden FP, Tabor D (1950) Friction and lubrication in solids. Clarendon Press, Oxford, U.K

Brown HR (1991) A molecular interpretation of the toughness of glassy polymers. Macromolecules 24:2752–2756

Bunker BC (1994) Molecular mechanisms for corrosion of silica and silicate glasses. J Non-Cryst Solids 179:300–308

Célarié F, Prades S, Bonamy D, Ferrero L, Bouchaud E, Guillot C, Marliére C (2003) Glass breaks like metal, but at the nanometer scale. Phys Rev Lett 90:075504

Célarié F (2004) Dynamique de fissuration a basse vitesse des matériaux vitreux. PhD thesis, Université Montpellier 2

Célarié F, Ciccotti M, Marlière C (2007) Stress-enhanced ion diffusion at the vicinity of a crack tip as evidenced by atomic force microscopy in silicate glasses. J Non-Cryst Solids 353:51–68

Charlaix E, Ciccotti M (2010) Capillary condensation in confined media. In: Sattler K (ed) Handbook of nanophysics: principles and methods. CRC Press, Boca Raton, p 12–1

Ciccotti M (2009) Stress-corrosion mechanisms in silicate glasses. J Phys D Appl Phys 42:214006

Cleveland JP, Anczykowski B, Schmid AE, Elings VB (1998) Energy dissipation in tapping-mode atomic force microscopy. Appl Phys Lett 72:2613–2615

Crichton SN, Tomozawa M, Hayden JS, Suratwala TI, Campbell JH (1999) Subcritical crack growth in a phosphate laser glass. J Am Ceram Soc 82:3097–104

Dimitrov A, Buchholz FG, Schnack E (2006) 3D-corner effects in crack propagation. Comput Model Eng Sci 12:1–25

Doll W (1983) Optical interference measurements and fracture mechanics analysis of crack tip craze zones. Adv Polym Sci 52/53:105–168

Donald AM, Kramer EJ (1982) The competition between shear deformation and crazing in glassy polymers. J Mater Sci 17:1871–1879

Du J, Cormack AN (2005) Molecular dynamics simulation of the structure and hydroxylation of silica glass surfaces. J Am Ceram Soc 88:2532–2539

Dugdale DS (1960) Yielding of steel sheets containing slits. J Mech Phys Solids 8:100–104

Fett T, Guin JP, Wiederhorn SM (2005) Interpretation of effects at the static fatigue limit of soda-lime-silicate glass. Eng Fract Mech 72:2774–2791

Fineberg J, Marder M (1999) Instability in dynamic fracture. Elsevier, Phys Rep 313:1–108

Freund LB (1990) Dynamic fracture mechanics. Cambridge University, Cambridge; Rate J Mat Sci 14:583–591

Gehrke E, Ullner C, Mahnert M (1991) Fatigue limit and crack arrest in alkali containing silicate glasses. J Mater Sci 26:5445–5455

George M, Nziakou Y, Goerke S, Genix AC, Bresson B, Roux S, Delacroix H, Halary JL, Ciccotti M (2018) In situ AFM investigation of slow crack propagation mechanisms in a glassy polymer. J Mech Phys Solids 112:109–125

Griffith AA (1920) The phenomena of rupture and flow in solids. Phil Trans R Soc Lond A 221:163–198

Grimaldi A, George M, Pallares G, Marlière C, Ciccotti M (2008) The crack tip: a nanolab for studying confined liquids. Phys Rev Lett 100:165505

G'Sell C, Jonas JJ (1979) Determination of the plastic behaviour of solid polymers at constant true strain. J Mater Sci 14:583–591

Guilloteau E, Charrue H, Creuzet F (1996) The direct observation of the core region of a propagating fracture crack in glass. Europhys Lett 34:549–553

Guin JP, Wiederhorn SM (2004) Fracture of silicate glasses: ductile or brittle? Phys Rev Lett 92:215502

Han K, Ciccotti M, Roux S (2010) Measuring nanoscale stress intensity factors with an atomic force microscope. EPL 89:66003

Halary JL, Lauprétre F, Monnerie L (2011) Polymer materials. Wiley, Haboken

Hattali ML, Barés, Ponson L, Bonamy D (2012) Low velocity surface fracture patterns in brittle material: a newly evidenced mechanical instability. Math Sci Forum 706–709:920–924

He MY, Turner MR, Evans AG (1995) Analysis of the double cleavage drilled compression specimen for interface fracture energy measurements over a wide range of mode mixities. Acta Metall Mater 43:3453–3458

Hutter K (2013) Deformation and failure in metallic materials. Springer, Berlin

Irwin GR (1957) Analysis of stresses and strains near the end of a crack traversing a plate. J Appl Mech 24:361–364

Janssen C (1974) Specimen for fracture mechanics studies on glass. In: Proceedings of the 10th International Congress on Glass, Kyoto, pp 10.23–10.30

Jones RM (2015) Mechanics Of composite materials, 2nd edn. CRC Press, Philadelphia

Kermode JR, Albaret T, Sherman D, Bernstein N, Gumbsch P, Payne MC, Csanyi G, De Vita A (2008) Low-speed fracture instabilities in a brittle crystal. Nature 455:1224–1228

Kinloch AJ, Williams JG (1980) Crack blunting mechanisms in polymers. J Mat Sci 15: 897–996

Kramer EJ (1983) Microscopic and molecular fundamentals of crazing. Adv Polym Sci 52/53:1–56

Lawn BR (1993) Fracture of brittle solids, 2nd edn. Cambridge University, Cambridge

Lechenault F, Pallares G, George M, Rountree C, Bouchaud E, Ciccotti M (2010) Effects of finite probe size on self-affine roughness measurements. Phys Rev Lett 104:025502

Marsh DM (1964) Plastic flow and fracture of glass. Proc R Soc London A 282:33–43

Maugis D (1985) Review: subcritical crack growth, surface energy, fracture toughness, stick-slip and embrittlement. J Mater Sci 20:3041–3073

McClintock FA, Irwin GR (1964) Plasticity aspects of fracture mechanics. In: Fracture toughness testing and its applications. ASTM STP 381; Philadelphia, pp 84–113

Michalske TA, Freiman SW (1983) A molecular mechanism for stress corrosion in vitreous silica. J Am Ceram Soc 66:284–288

Michalske TA, Bunker BC (1984) Slow fracture mode based on strained silicate structures. J Appl Phys 56:2686–2693

Mischler C, Horbach J, Kob W, Binder K (2005) Water adsorption on amorphous silica surfaces: a Car-Parrinello simulation study. J Phys Cond Matt 17:4005–4013

Nziakou Y (2015) Analyse multi-échelle des mécanismes d'endommagement des matériaux composites à morphologie complexe destinés à l'aéronautique. PhD Thesis, Université Pierre et Marie Curie

Nziakou Y, George M, Fisher G, Bresson B, Roux S, Halary JL, Ciccotti M (2019) Bridging steady-state and stick-slip fracture propagation in glassy polymers. preprint Soft Matter (submitted)

Orowan E (1955) Energy criteria of fracture. Weld J Res Suppl 34:S157–S160

Pallares G, Ponson L, Grimaldi A, George M, Prevot G, Ciccotti M (2009) Crack opening profile in DCDC specimen. Int J Fract 156:11–20

Pallares G, Grimaldi A, George M, Ponson L, Ciccotti M (2011) Quantitative analysis of crack closure driven by Laplace pressure in silica glass. J Am Ceram Soc 94:2613–2618

Pallares G, George M, Ponson L, Chapuliot S, Roux S, Ciccotti M (2015) Multiscale investigation of stress-corrosion crack propagation mechanisms in oxide glasses. Corros Rev 33(6):501–514. Freund Publishing House Ltd.

Pallares G, Lechenault F, George M, Bouchaud E, Ottina C, Rountree CL, Ciccotti M (2017) Roughness of oxide glass subcritical fracture surfaces J Am Ceram Soc 101:1279–1288

Phillips DC, Scott JM, Jones M (1978) Crack propagation in an amine-cured epoxide resin. J Mat Sci 13:311–322

Réthoré J, Estevez R (2013) Identification of a cohesive zone model from digital images at the micron-scale. J Mech Phys Solids 61:1407–1420

Roduit C, AFM figures 2010, Creative Commons Attribution

Roux S, Hild F (2006) Stress intensity factor measurements from digital image correlation: post-processing and integrated approaches. Int J Fract 140:141–157

Taylor EW (1949) Plastic Deformation of Optical Glass. Nature 163:323

Takahashi K, Arakawa K (1984) Dependence of crack acceleration on the dynamic stress-intensity factor in polymers. Exp Mech 27:195–199

Thomson W (1871) On the equilibrium of vapour at a curved surface of liquid. Philos Mag 42:448–452

Tomozawa M (1984) Effect of stress on water diffusion in silica glass. J Am Ceram Soc 67: 151–154

Tomozawa M (1996) Fracture of glasses. Ann Rev Mater Sci 26:43–74

Wiederhorn SM (1967) Influence of water vapor on crack propagation in soda-lime glass. J Am Ceram Soc 50:407–414

Wiederhorn SM (1969) Fracture surface energy of glass. J Am Ceram Soc 52:99–105

Wiederhorn SM, Bolz LH (1970) Stress-corrosion and static fatigue of glass. J Am Ceram Soc 53:543–548

Wiederhorn SM, Lopez-Cepero J, Wallace J, Guin JP, Fett T (2007) Roughness of glass surfaces formed by sub-critical crack growth. J Non-Cryst Solids 353:1582

Wiederhorn SM, Fett T, Rizzi G, Fünfschilling S, Hoffmann MJ, Guin JP (2011) Effect of water penetration on the strength and toughness of silica glass. J Am Ceram Soc 94:S196–S203

Wiederhorn SM, Fett T, Guin JP, Ciccotti M (2013) Griffith cracks at the nanoscale. Int J Appl Glass Science 4:76–86

Williams JG (1984) Fracture Mechanics of Polymers. Ellis Horwood, Chichester

Williams ML (1957) On the stress distribution at the base of a stationary crack. ASME J Appl Mech 24:109–114

Long Time-Scale Atomistic Modeling and Simulation of Deformation and Flow in Solids

11

Yue Fan and Penghui Cao

Contents

Abstract

Atom-based modeling and simulation are essential for the understanding and development of structural materials such as crystalline and amorphous metals. Classical molecular dynamics simulation enables the following of atomic-level structural evolution to elucidate the atomic processes underlying many macroscopic behaviors; however its predictive power is constrained by an intrinsic time-scale limitation. Here, we describe an alternative approach based on potential energy landscape modeling and transition state theory to probe the microscopic mechanisms controlling deformation and plastic flow observed in experiments. We survey several examples of slow deformation in crystals and metallic glasses to illustrate the computational algorithms used to perform the

Y. Fan

Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA
e-mail: fanyue@umich.edu

P. Cao (✉)

Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA
e-mail: caoph@uci.edu

simulations and to reveal the underpinning elementary plastic processes that operate in crystalline and amorphous materials. We first show the evolution of dislocations and their interactions with obstacles over a wide range of strain rates and temperatures and discuss how they lead to macroscopic behaviors such as flow stress upturn and dislocation channeling. Then we turn to amorphous plasticity where discrete stress relaxation (avalanche) processes arise in serrated flow and creep in metallic glasses. A nonlinear interplay between nonaffine atomic displacement and local shear transformation distortion is revealed that provides a molecular explanation of the deformation-rate upturn.

1   Introduction

To meet the demand of advanced materials with unique functionalities, it is becoming increasingly important to study material performance in complex environments at the fundamental level. Computational modeling and simulation, as widely applied techniques nowadays, can provide unprecedented details on materials from atomic level. While simulations have particular advantages over "real" experiments in terms of lower cost, faster progress, and better control of isolated parameters in affecting the modeled system, there are still unsolved problems in terms of the limited spatial and temporal scales in traditional atomistic modeling techniques. The length scale issue is less challenging because even a few thousand atoms, with the use of periodic boundary condition, can reasonably represent many materials' bulk properties. In addition, the length scale limitation can, at least partially, be addressed by parallelizing the problem and exploiting supercomputer systems. However, the time step in the integration over time cannot be parallelized in the computations. This makes it a formidable challenge to reach long timescales with atomistic simulations. For example, as will be discussed below, the widely used molecular dynamics (MD) simulation can hardly go beyond nanoseconds, which is often more than ten orders of magnitude away from the timescales of the phenomena of interest in real experiments. It is therefore highly desirable to develop and employ new modeling algorithms that can address such timescale issues, while still retaining the atomistic details.

The properties of condensed matter systems are primarily governed by the interactions between constituent atoms in the system. The system's potential energy is a function of its atomic configurations. For a many-body interacting system with defects (i.e., topological disorders), there are many local energy minima, known as inherent structures, representing metastable states of the system. The hyperdimensional energy surface in configuration space is called potential energy landscape (PEL), and material microstructural evolutions correspond to the progressive exploration of different local minima in PEL (Suo 2000; Wei and Zhigang 1996). In the scope of this chapter, the microstructural evolution consists of non-equilibrium processes, which are governed by collective, rare events in the PEL. Since the activation barriers for the rare events are very high, according to transition state theory (TST), the system will be trapped in deep energy basins for an amount

of time (proportional to $\exp\left[\Delta E / k_B T\right]$, where ΔE is the activation barrier), which can be well beyond that accessible by MD. To enable long timescale modeling, an alternative approach is to proactively explore the structures of energy landscape, identify the transition pathways and barriers, and incorporate that information into TST calculations. Different atomistic algorithms have been developed, including the activation-relaxation technique (ART) (Barkema and Mousseau 1996; Cances et al. 2009), the dimer method (Henkelman et al. 1999; Haixuan et al. 2012), hyper- (or meta-)dynamics (Haixuan 1997; Laio and Parrinello 2002), temperature-accelerated dynamics (TAD) (Sorensen and Voter 2000; Bai et al. 2010), and the autonomous basin climbing (ABC) method (Kushima et al. 2009a; Fan et al. 2014; Cao et al. 2012; Yip 2016). Each technique has its own advantages and drawbacks (Fan et al. 2014; Fan 2013). In this chapter we focus on the ABC method. To the best of our knowledge, such algorithm has been found useful in both dislocation system (Fan et al. 2012, 2013; Bai and Fan 2018; Wang et al. 2013; Tao et al. 2018a, b) and amorphous system (Kushima et al. 2009a; Cao et al. 2012, 2013, 2017, 2018; Kushima et al. 2009b) as discussed below in Sects. 2 and 3, respectively. The development of original ABC algorithm (Kushima et al. 2009a) was inspired by Laio and Parrinello's idea of escaping from free-energy minima (Laio and Parrinello 2002). To be more specific, a series of penalty function are applied, and total relaxations are performed to allow the system to climb out of the local basin in the PEL. More recently the ABC method has been extended to capture the multiple competitive pathways in the PEL (Fan et al. 2014). In addition, to enhance the computation efficiency, a novel strategy termed self-learning metabasin escape (SLME) was formulated to merge the newly applied penalty function with those already added ones. Given the limited space here, we refer the readers to the original references for implementation details (Cao et al. 2012; Yan et al. 2016).

2   Evolution of Dislocations in Crystalline Materials

The mechanical behavior of structural metals is controlled by evolutions and interactions between microstructural features known as defects in the materials. Among various types of defects affecting the mechanical properties of structural materials, dislocations are one of the most critical microstructures because they are the main carriers of plastic deformation (Hull and Bacon 2011). In particular, when there are obstacles in slip planes, they will resist the glide motion of dislocations and hence significantly change the system's mechanical performance, including the yield strength, work hardening, strain-rate sensitivity, strain localization, elastic-viscoplastic behavior, etc. (Bakó et al. 2006; Bulatov et al. 2006; Brechet and Estrin 1995; Dunlop et al. 2007; Lebyodkin et al. 1995). Therefore, probing dislocation evolutions and their interactions with obstacles at various environments are the keys to understand, predict, and control mechanical properties of materials. This is a subject that has been extensively studied over the past few decades (Fan et al. 2013; Bacon et al. 2006; Domain and Monnet 2005; Dutta et al. 2012; Gordon et al. 2010;

Gussev et al. 2015; Iyer et al. 2015; Kabir et al. 2010; Kioussis and Ghoniem 2010; Norfleet et al. 2008; Onimus et al. 2004; Osetsky et al. 2006a).

Computational modeling and simulation, accompanied by the continuous enhancement of computing power, have become an effective quantitative approach in materials research, constructively complementing theories and experiments (Yip 2003). Although first principle calculations and MD studies have yielded notable advances in the knowledge of dislocation-mediated mechanics from fundamental levels, they face formidable challenges in probing the non-equilibrium phenomena at realistic timescales. To be more specific, typical MD studies on dislocations are performed either at very high strain rates greater than $10^6\,\mathrm{s}^{-1}$ (Bacon et al. 2006, 2009; Terentyev et al. 2010a, 2008), while on the other hand, typical mechanical test experiments are performed under low strain-rate conditions, i.e., slower than $10^0\,\mathrm{s}^{-1}$ (Dunlop et al. 2007; Onimus et al. 2004; Onimus and Béchade 2009). It has been found that intrinsic timescale limitations in conventional modeling not only could induce huge errors in yield strength quantifications (Zhu et al. 2008) but, more importantly, might provide misleading microstructural evolution mechanisms (Fan et al. 2013). Such challenge therefore calls for the development of a reliable predictive model that can capture the fundamental mechanisms at realistic timescale.

In this section we consider two examples concerning the activities of dislocations over a broad range of timescales by using the ABC-based modeling technique introduced above: (i) nonlinear variation of dislocation mobilities and onset of flow stress upturn and (ii) interaction mechanism map between dislocation and obstacle under various thermomechanical conditions.

2.1   Flow Stress of an Individual Dislocation

At low temperature, the deformation of metals is largely governed by the thermal activation of dislocation glide (Hoge and Mukherjee 1977). Experiments on different structures of metals, including Fe, Ta, Cu, Al, and Zn (Hoge and Mukherjee 1977; Ferguson et al. 1967; Kumar et al. 1968; Campbell and Ferguson 1970; Follansbee et al. 1984; Regazzoni et al. 1987), indicate the dislocation flow stress varies with strain rate in an apparently universal manner. The flow stress increases slowly in an Arrhenius manner at low strain rates but turns upward sharply beyond a certain range of strain rate. Although the results for different metals can be quantitatively different, the flow stress "upturn" behavior when the strain rate reaches the range of 10^3–$10^4\,\mathrm{s}^{-1}$ (Regazzoni et al. 1987; Armstrong et al. 2009) appears to have a more fundamental origin. The onset of non-Arrhenius response has elicited the development of several empirical constitutive models (Hoge and Mukherjee 1977; Armstrong et al. 2009; Zerilli and Armstrong 1987; Follansbee and Kocks 1988; Steinberg et al. 1989; Preston et al. 2003; Remington et al. 2006), including the assumption of phonon drag effects to account for the data at high strain rates. All existing models to date use adjustable parameters to connect the flow stress below and above the critical strain rate for "upturn." Therefore, it would be of interest to seek a more fundamental understanding with less postulated mechanisms or adjustable parameters.

From the PEL perspective, a dislocation line is located on the bottom of the so-called Peierls energy valley until a thermal fluctuation enables it to climb over the activation barrier to glide to the next valley. When an external stress is applied, the depth of the valley, i.e., the glide activation barrier $E(\sigma)$, can be decreased, making the dislocation easier to escape from the valley. As a result, the escape rate can be calculated by using transition state theory (TST):

$$k(\sigma) = k_0 e^{-\frac{E(\sigma)}{k_B T}} \tag{1}$$

where k_0 is the attempt frequency. The activation barrier $E(\sigma)$ can be either empirically assigned or directly identified by employing ABC method at various stress levels.

In the elastic deformation regime, the dependence of stress on applied strain rate as a function of time, t, is given by

$$\sigma = G\varepsilon = G\dot{\varepsilon}t \tag{2}$$

where G is the shear modulus. The ε in Eq. (2) represents the elastic strain, because in this case study we focus on the initiation of dislocation flow, which pertains to the transition from elastic deformation regime to plastic deformation regime. In light of Eq. (2), $k(\sigma)$ can be represented as a function of time, $k(t)$.

The residence probability P(t) that the dislocation does not escape to a neighboring potential energy valley during time t (i.e., the system remains in the elastic deformation regime) is defined as Zhu et al. (2008)

$$\frac{dP(t)}{dt} = -k(t)P(t) \tag{3}$$

or

$$P(t) = \frac{1}{C} \exp\left[-\int_0^t k(t')dt'\right] \tag{4}$$

where C is the normalization factor. Accordingly, the first escape probability distribution p(t) is given by

$$p(t) = -\frac{dP(t)}{dt} = \frac{1}{C}k(t)\exp\left[-\int_0^t k(t')dt'\right] \tag{5}$$

with normalization,

$$\int_0^{t_c} p(t)dt = 1 \Rightarrow C = \int_0^{t_c} k(t)\exp\left[-\int_0^t k(t')dt'\right]dt \tag{6}$$

where $t_c = \sigma_c/G\dot{\varepsilon}$ represents the maximum residence time, at a given non-zero strain rate $\dot{\varepsilon}$. The average residence time is therefore given by

$$\bar{t} = \int_0^{t_c} tp(t)dt = \frac{\int_0^{t_c} tk(t)\exp\left[-\int_0^t k(t')dt'\right]dt}{\int_0^{t_c} k(t)\exp\left[-\int_0^t k(t')dt'\right]dt} \tag{7}$$

In the limit of vanishing $\dot{\varepsilon}$, $k(t)$ is a constant, k, and $t_c \to \infty$, Eq. (7) gives the average time as $\bar{t} = 1/k$, which follows the Arrhenius law. However, for the general condition of non-zero strain rate, the result of Eq. (7) will deviate from the Arrhenius behavior.

We would like to stress here that the derivations of Eqs. (1), (2), (3), (4), (5), (6), and (7) represent a general formalism that is applicable to a wide range of systems where the reaction rate is time-dependent. Relevant examples in the present context would include dislocation nucleation under constant strain-rate loading, glass transition at different cooling rates, as well as dislocation flow. In this case study, we are particularly interested in predicting the variation of flow stress with temperature and strain rate. Since the dislocation will start to glide beyond the residence time $\bar{t}$, we obtain the flow stress by combining Eqs. (2) and (7):

$$\bar{\sigma}_{flow} = G\dot{\varepsilon}\bar{t} = \frac{\int_0^{\sigma_c} \sigma k(\sigma)\exp\left[-\frac{1}{G\dot{\varepsilon}}\int_0^{\sigma} k(\sigma')d\sigma'\right]d\sigma}{\int_0^{\sigma_c} k(\sigma)\exp\left[-\frac{1}{G\dot{\varepsilon}}\int_0^{\sigma} k(\sigma')d\sigma'\right]d\sigma} \tag{8}$$

In summary, once the dislocation migration barrier profile $E(\sigma)$ is known, the flow stress of the corresponding slip system at arbitrary thermomechanical conditions can be calculated according to Eq. (8). There is then only one parameter in Eq. (8), the attempt frequency k_0, which can be taken to be of the order of $10^{12\sim13}\,\mathrm{s}^{-1}$. Below we will apply the constitutive equations just derived to a specific example of the flow stress in bcc Fe.

The deformation of bcc metals at low temperature is known to be controlled by the motion of <111> screw dislocations (Domain and Monnet 2005), the flow mechanism being 3D kink nucleation and propagation (Gordon et al. 2010; Rodney and Proville 2009). For the purpose of testing Eq. (8), we examine a short dislocation of length 5b, b being the Burger's vector, which should glide without kink nucleation. Figure 1b above shows the glide barrier for such a dislocation in bcc Fe with an embedded-atom method-type potential developed by Mendelev et al. (2003). Clearly the barrier shows a monotonically decreasing behavior well described by a (p, q) expression $E(\sigma) = E_0\left[1 - (\sigma/\sigma_c)^p\right]^q$ (Rodney and Proville 2009; Kocks et al. 1975), where p = 0.63 and q = 1.41. For the convenience of benchmark discussion below, we also include a dashed line in Fig. 1b denoting p = q = 1, which is the assumption of a linear behavior with a constant activation volume.

Figure 2a shows the thermal behavior of flow stress for strain rates varying over ten orders of magnitude, from $10^7\,\mathrm{s}^{-1}$ down to $10^{-3}\,\mathrm{s}^{-1}$. In the low temperature limit, absence of thermal activation, all flow stresses approach the Peierls stress, which is 1400 MPa for the given potential. As temperature increases, all the
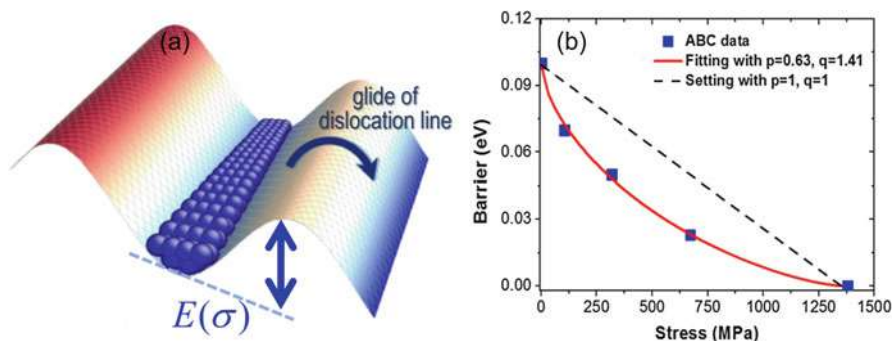
Fig. 1 (**a**) The glide of dislocation corresponds to the hopping between Peierls energy valleys. (**b**) Activation barrier for dislocation glide usually decreases as a function of applied shear stress. Blue squares are the calculated data points by ABC method, and corresponding solid curve is a fit to the realistic (p,q) expression. Dashed line represents a hypothetical linearly decreasing profile

flow stresses monotonically decrease. At a fixed temperature, higher strain-rate loading results in higher flow stress response, suggesting that any quantitative comparison between experimental data and MD simulations must take into account the difference in the strain rate.

The symbols in Fig. 2a represent the corresponding MD results at strain rate of $10^6\,\text{s}^{-1}$ and $10^7\,\text{s}^{-1}$. They are in reasonable agreement with the predictions of Eq. (8) using $k_0 = 1.2*10^{12}\,\text{s}^{-1}$, which matches the Debye frequency satisfactorily. This constitutes a self-consistent test of Eq. (8) with $E(\sigma)$ taken from Fig. 1b (the solid curve) in the range of strain rates where MD is valid. One can see an increasingly sharp drop of flow stress as the strain rate decreases to the range accessible to conventional experiments. The sharp drop has been known as a significant feature of the thermal activation process; this behavior is not well captured by MD simulations at its characteristic strain rates (Domain and Monnet 2005; Zhu et al. 2008). Figure 2a shows that this behavior is at least qualitatively accounted for by the present model.

The variation of flow stress with strain rate is of fundamental interest in experimental studies of crystal plasticity. Figure 2b shows the predicted behavior based on Eq. (8). Under the limit of infinitely high strain rate, the flow stress approaches the Peierls stress. On the other hand, the flow stress is negatively sensitive to the temperature. In the high temperature limit, the flow stress approaches zero regardless of the strain rate. At low $\dot{\varepsilon}$ the flow stress increases only moderately, but as $\dot{\varepsilon}$ increases, above $100\,\text{s}^{-1}$ at 50 K, and $10^4\,\text{s}^{-1}$ at 100 K, it begins to increase much more strongly. This upturn behavior can be analyzed in terms of two factors, stress-dependent activation volume and strain rate-induced non-Arrhenius behavior. Because of the nonlinear stress dependence of the activation barrier (Fig. 1b), the activation volume is very small at high stresses. Such small activation volume leads to a high sensitivity of the flow stress dependence on strain rate (Zhu et al. 2008). In addition, as derived in Eq. (7), there is a non-Arrhenius behavior due to the

Fig. 2 (**a**) Flow stress of the <111> screw dislocation in bcc Fe under different strain-rate and temperature conditions. The solid lines show the calculated results by Eq. (8), and the data points are MD results at high strain rates. (**b**) Predicted flow stress variation at different strain rates. Solid and dashed curves are calculated based on realistic (p,q) barrier profile and hypothetical linear profile shown in Fig. 1b, respectively
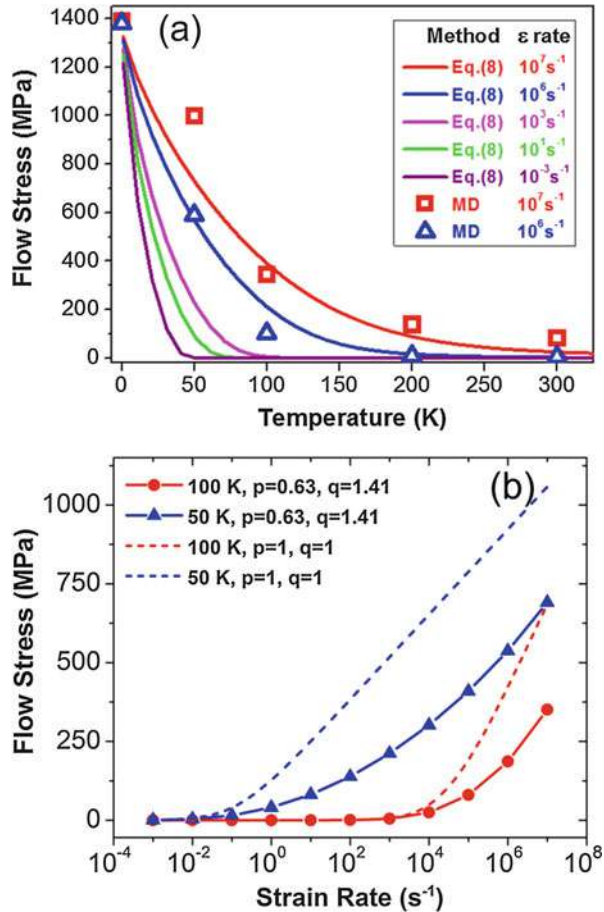
strain-rate loading which also contributes to the upturn in Fig. 2b. To decouple the two contributions, we remove the nonlinearity of $E(\sigma)$ by setting p and q equal to unity (dashed line in Fig. 1b). Now the only nonlinear factor comes from the strain rate-induced non-Arrhenius behavior in Eq. (8). As shown in Fig. 2b, under this condition, the flow stress upturn remains, but the stress is now higher beyond the crossover strain rate. Since the assumption of $p = q = 1$ results in a higher effective barrier and correspondingly a longer residence time, it follows that the flow stress response is higher as well. Our analysis therefore shows the onset of flow stress upturn is to be attributed mainly to the non-Arrhenius behavior induced by strain rate, as described by Eqs. (7) and (8).

To compare the predicted upturn behavior quantitatively with experimental data, we adapt the energy profile $E(\sigma)$ for a longer screw dislocation system in bcc Fe

Fig. 3 Variation of reduced flow stress with strain rate at 300 K. (The experimental data on copper and ion are adapted from Armstrong et al. (2009) and references therein)
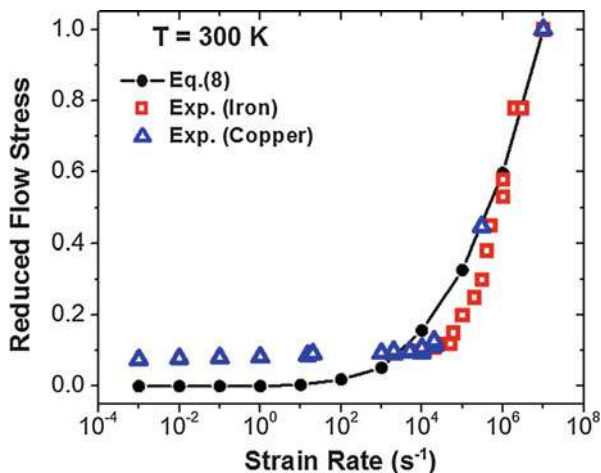
calculated by Gordon et al. (2010) and use it as the input into Eq. (8). Figure 3 shows the variation of flow stress and strain rate at 300 K, as observed experimentally and predicted by our model. Since the flow stress magnitude is significantly influenced by the defect microstructures in the experimental specimens (Armstrong et al. 2009), the quantitative comparison can only be meaningful after an appropriate normalization. Therefore, in Fig. 3, we show the reduced flow stress, defined as the ratio of flow stress to its value at the highest strain rate $10^7 \, \text{s}^{-1}$, as a function of strain rate. It is seen that both the experiments and our calculation results show a significant flow stress upturn with the critical strain rate in the range of 10^4–$10^5 \, \text{s}^{-1}$. The extent of the agreement suggests Eq. (8) and $E(\sigma)$ have essentially captured the mechanism for the flow stress upturn behavior. On the other hand, it is known that the flow stress magnitude depends on the local defect microstructure in the material. Experimental specimens have a complex defect microstructure leading to appreciably higher flow stresses (due to, e.g., dislocation-obstacle interactions to be discussed in the next section) seen in the experiments compared to the results in Fig. 2b. It is therefore intriguing that the reduced flow stress predicted by our model in Fig. 3 is also quantitatively consistent with experiments from different materials (a ductile one, copper, and, a brittle one, iron). This may be attributed to the fact that the energy barrier for dislocation to climb/glide over the defects/obstacles in the material bears a similar stress-activated behavior as the simple dislocation glide represented by the expression $E(\sigma) = E_0 \left[1 - (\sigma/\sigma_c)^p \right]^q$ that was described above (Remington et al. 2006).

In sum, such example shows that, without invoking any specific mechanism or fitting parameters, the PEL-based approach can naturally capture the nonlinear transition of the flow stress from a classical Arrhenius behavior to a fast upturn over a broad range of strain rates.
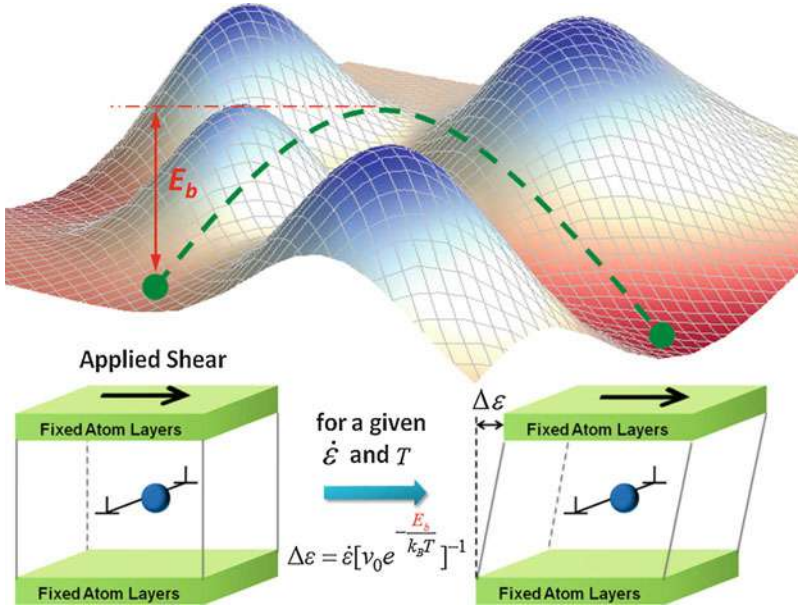
Fig. 4 Schematic ABC-TST modeling on dislocation-obstacle interaction at prescribed thermo-mechanical conditions. The detailed iterative procedures are described in the main text

2.2   Dislocation-Obstacle Interaction Mechanism Map

Under complex environments (e.g., irradiation, severe plastic deformation, etc.), a host of non-equilibrium obstacles/precipitates would be present in the system. It therefore becomes particularly important to probe the dislocation-obstacle interaction mechanisms at various conditions since they will determine many important mechanical properties of structural materials, such as hardening, creep, embrittlement, crack nucleation, etc. (Bakó et al. 2006; Bulatov et al. 2006; Jumonji et al. 1996; Van der Giessen et al. 2001; Shin et al. 2005).

The first key challenge that needs to be addressed is to investigate the dislocation-obstacle interactions at much longer timescale, well beyond the range accessible to MD. As introduced above, the ABC method is a robust algorithm that can drive the system out from a trapped deep energy basin and thus probe the microstructural evolution at long timescales. Therefore, by iteratively combining ABC and TST into a dynamic scenario, one could then study the dislocation-obstacle interactions at prescribed strain rate and temperature, as illustrated in Fig. 4.

The consecutive iterations can be described as:

1. Under a given strain condition (including the initial state with no strain), employ ABC to explore the PEL structure, and thus obtain the dominant reaction pathway at this moment.

2. Apply nudged elastic band (NEB) method to accurately calculate the energy barrier E_b between the current and next minimum states in PEL that are identified in step (i).

3. Determine the thermal activation time defined as $\Delta t = [\nu_0 \cdot \exp(-E_b/k_B T)]^{-1}$, where ν_0 is the attempt frequency (of Debye order 10^{13} s^{-1}). For a specified strain rate $\dot{\varepsilon}$, the corresponding strain increment, $\Delta \varepsilon = \dot{\varepsilon} \cdot \Delta t = \dot{\varepsilon} \ [\nu_0 \cdot \exp(-E_b/k_B T)]^{-1}$, is applied to the system. Go back to step (i) under the new strain condition.

Through this algorithm, the coupling between thermal activation and strain rate is treated in the transition state theory framework. To be more specific, because ABC is employed iteratively on-the-fly, the environmental effects on PEL will be naturally captured. Since the atomic arrangements near the defects (e.g., dislocations and obstacles) are off perfect lattice sites and highly disordered, the underlying PEL structures could be rather complex, and each local minimum state in PEL might connect with multiple different transition pathways, corresponding to different evolution mechanisms. If these transition pathways show different sensitivities to surrounding environments (e.g., strain rate, temperature, stress state, etc.), then a crossover between their relative importance at different conditions might occur, which would then lead to different deformation mechanisms and mechanical behavior.

Figure 5 shows an example on the interaction mechanisms between a $< 11\bar{2}0 >$ $\{10\bar{1}0\}$ dislocation and a self-interstitial atom (SIA) cluster in HCP Zr over a wide range of thermomechanical environments (Fan et al. 2013). Starting from the initial state in Fig. 5a, the PEL-based iterative ABC-TST procedure enables one to uncover the interaction mechanisms over 15 orders of magnitude in timescales, from 10^7 s^{-1} down to 10^{-7} s^{-1}, far beyond the MD's range but appropriate for direct comparison with experiments. It is demonstrated in Fig. 5b that even the same unit process could lead to remarkably different mechanisms. Specifically, in the low temperature and high strain-rate regime, the dislocation tends to simply pass by the obstacle, and the system undergoes a stable homogeneous deformation with normal work hardening, while in the higher temperature and lower strain-rate regime, the obstacle is absorbed and swept by the dislocation. The removal of obstacle will result in a defect-free region along the slip plane known as dislocation channel, which is associated with strain localization and heterogeneous deformation, and could further induce mechanical instability.

The reason why the same initial state could give rise to significantly different mechanical behavior can be explained by the environmental effects on the system's underlying PEL structure revealed by the ABC technique. As shown in Fig. 5c, for such unit process in HCP Zr, the initial state connects with two transition pathways that are associated with different microstructural evolution modes. At the beginning, path I (obstacle removal) has a lower activation energy than path II (simple bypass) does. Therefore, if the surrounding environments do not change significantly, say when the driven strain rate is relatively low, then the system would naturally have a higher chance to go through the path I and give rise to the dislocation
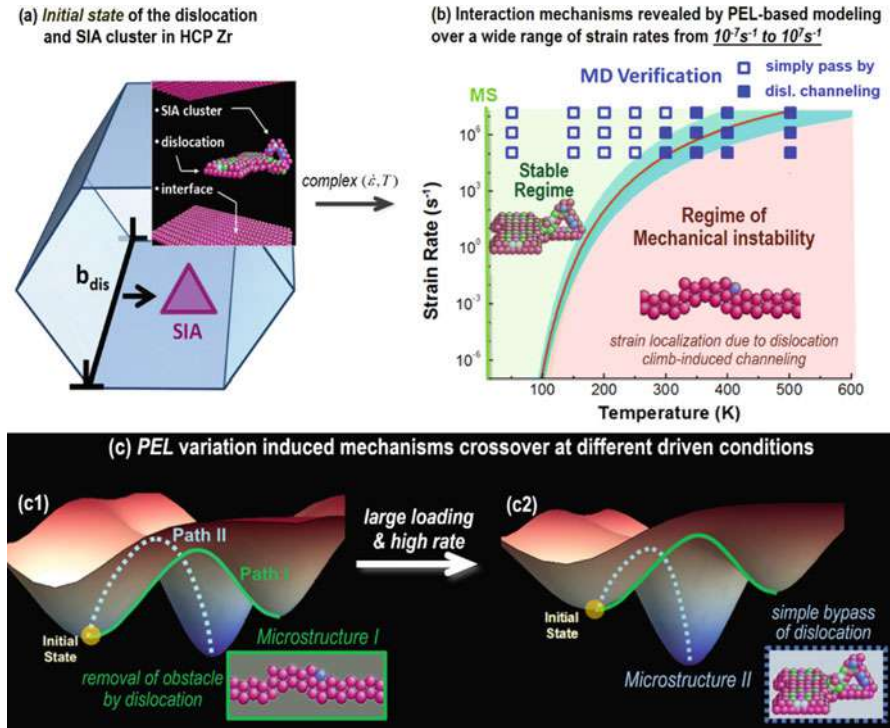
Fig. 5 (**a**) Initial state of the unit process of dislocation-SIA cluster interaction in the prismatic slip system in HCP Zr. (**b**) The same initial state could result in remarkably different evolution mechanisms when the timescale is spanning more than 15 orders of magnitude. In the low temperature and high strain-rate regime, the system undergoes a stable homogeneous deformation process because the dislocation simply bypasses the SIA cluster. However, in the higher temperature and lower strain-rate regime, the SIA cluster is absorbed and removed by the gliding dislocation, which then leads to dislocation channeling and mechanical instabilities. It can be seen that the ABC-based modeling technique allows one to probe the microstructural evolution mechanisms from atomic level at realistic timescale far beyond the traditional MD's range (region of blue squares). (**c**) The crossover of two different mechanisms originates from the variation of system's underlying PEL under different environments. As shown in (c1), the initial state connects with two transition pathways that are associated with different microstructural evolution modes. At the beginning, path I (obstacle removal) has a lower activation energy than path II (simple bypass) does. Therefore, if the driven strain rate is relatively low, i.e., when the surrounding environments do not change appreciably, then the system would naturally have a higher chance to go through the path I, which leads to obstacle removal. On the other hand, under very high strain-rate scenario, since path II is more sensitive to external loading and its activation barrier under large strain becomes lower than path I does, as seen in (c2), then path II becomes more favorable and a bypass mechanism occurs

channel formations. This is indeed what has been observed by TEM experiment (Onimus et al. 2004) at the condition of 600 K and 10^{-4} s^{-1}. However, as shown in Fig. 5c1–c2, the responses of two pathways to external stimuli are different. Path II is more sensitive to the applied shear loading, and its activation barrier becomes

lower at large strain condition. Therefore, under very high strain-rate scenario, path II becomes favorable very quickly, and a different deformation mechanism would occur.

To validate the diagram just predicted, independent MD simulations have been employed. Although MD studies are limited to the very high strain-rate regime, the results at different temperatures (blue squares in Fig. 5b) show a clear mechanism crossover between dislocation bypass and channeling.

We would like to note that the approach in constructing the mechanism map is based on a general thermal activation theory under prescribed strain rate (Fan et al. 2012, 2013), and is not limited to the defects discussed here in the Zr system. For instance, previous MD simulations in bcc Fe report different interaction mechanisms between a dislocation and a SIA loop as a function of temperature (Terentyev et al. 2010a; Bacon et al. 2009). Specifically, at high temperature, the SIA loop is absorbed by the dislocation, while at low temperature, both dislocation and SIA loop are completely recovered. Despite both being at high strain rate, these two different outcomes in bcc Fe can also be understood as a competition of strain rate and temperature, as explained by the mechanism map in Fig. 5b. In other words, the present framework is quite general and can be applied to study different materials (e.g., glass materials, colloidal system) and responses to different conditions (e.g., annealing rate, irradiation rate).

3   Plasticity in Amorphous Solids

Glassy materials deform and flow in response to external mechanical agitation and internal thermal activation. Due to the lack of well-defined structural defects such as dislocations in their crystalline counterpart, studying the mechanistic details, by which the constituent atoms rearrange themselves under stress and temperature environment, is still not well understood. In particular, the stress and strain-rate effects on plastic deformation and its underlying microscopic processes are in need of further studies.

A number of theoretical models have been proposed to describe the mechanisms of atomistic deformation and flow responsible for plastic deformation in amorphous solids. Spaepen considered the distinction between homogeneous and inhomogeneous flows in metallic glasses (Spaepen 1977) and introduced the concept of the local free volume as an order parameter. In this view local strain production and dissipation are assumed to be associated with individual atomic jumps. Argon proposed a plastic deformation model of metallic glasses based on the notion of local shear transformations driven by stress and in the presence of thermal fluctuations (Argon 1979). The atoms participating in such processes essentially undergo an inelastic shear deformation. Twenty years later, Falk and Langer introduced the term shear transformation zone (STZ) in interpreting simulation results of viscoplastic deformation of amorphous solids (Falk and Langer 1998), and the term STZ has become widely adopted in studies of amorphous materials.

On the other hand, significant attempts to understand amorphous plasticity have been carried out using molecular dynamics (MD) simulations. For instance, MD results revealed that structural disorder-induced softening precedes thermal softening which acts as the origin of the shear banding (Cao et al. 2009). The bottleneck well-known in MD is that simulations of deformation often entails extreme conditions of strain rate which are higher by several orders of magnitude than those studied experimentally (Yamakov et al. 2002; Sentjabrskaja et al. 2015). Because slow thermal activations and relaxations could be suppressed due to the short timescale in MD simulations, direct comparison with experiments cannot be straightforwardly performed.

In this section, we discuss two studies of amorphous plastic deformation under low mechanical stress or slow strain rate using the ABC-based PEL sampling technique: (1) constant stress-driven amorphous creep and (2) strain rate-dependent slip avalanches in amorphous plasticity.

3.1   Creep: Deformation and Flow at Constant Mechanical Stress

Creep is a slow deformation phenomenon occurring in materials typically subjected to constant mechanical stress and temperature. Creep deformation has been studied in various amorphous materials such as metallic glasses and colloidal systems (Krisponeit et al. 2014; Huang et al. 2009; Siebenbürger et al. 2012; Schuh et al. 2007; Lu et al. 2003). To identify the elementary processes of deformation and flow in amorphous creep, we implement a meta-dynamic formulation (Cao et al. 2017) that allows transition state trajectories to be generated on timescale of seconds. In the simulation, we apply a prescribed uniaxial tensile stress to the system and execute the following steps:

1. Perform energy minimization on the relaxed system to bring it to the nearest local energy minimum.
2. Apply the autonomous basin climbing (ABC) algorithm (Kushima et al. 2009a; Cao et al. 2012) to obtain the transition state pathway and determine the neighboring local energy minimum state.
3. Compare the internal stress of the new state with the prescribed tensile stress. If the two stresses deviate by more than 1%, perform step (d); otherwise go back to step (b).
4. Perform cell relaxation in the presence of the external stress. The atoms are rescaled to new positions whenever the size of simulation cell is changed, and the final configuration converges to a new local minimum.

The system evolution is then determined by examining the newly sampled configurations of the local energy minima. The activation time of each evolution step can be estimated through transition state theory, $\Delta t_i = \left[v_0 \exp(-\frac{\Delta E_i}{k_B T}) \right]^{-1}$, with attempt frequency v_0 typically taken to be 10^{12} s^{-1}, ΔE_i is the energy barrier of activation path i.
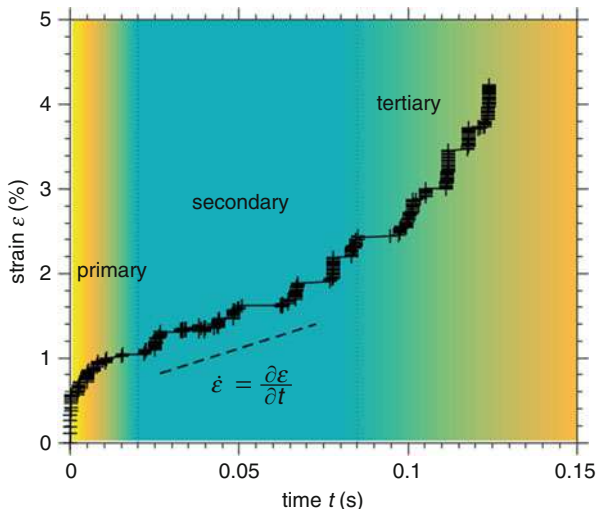
Fig. 6 Simulated creep curve: time evolution of system strain ε at stress 198 MPa and temperature $0.68T_{\mathrm{g}}$. The creep rate $\dot{\varepsilon}$ of the secondary stage is essentially constant and at a minimum. The background colors are drawn schematically to denote the creep rate

Figure 6 shows the simulation result of time development of the system strain. The creep curve shows the classical behavior of three stages of strain evolution. The initial period of strain buildup, consisting of a steep increase followed by a gradual approach to saturation, is known as primary or transient creep. The secondary stage of steady-state creep is a period of linear strain increase in time. The extent of this stage depends on the combination of applied stress and system temperature. For relatively low stresses, the secondary stage may have a considerable extent before the onset of tertiary creep, where strain rate increases without apparent limit. In the case of Fig. 6, the onset of structural instability at a strain level of ∼2.5% is readily observed.

In the present discussion, our interest lies in the mechanisms sustaining the steady-state creep (secondary stage). By varying the value of applied stress, one can determine the creep rate, and its variation with stress is shown in Fig. 7a. We can see a bimodal behavior in the monotonic variation of creep rate $\dot{\varepsilon}$ with stress σ. At low stress, $\dot{\varepsilon}$ is characterized by a creep-rate slope of $n \sim 1.5$, where n is known as the stress exponent. At high stress n increases to ∼5. Notice the stress value for the change in index n. The transition in creep rate behavior is well-known in experiments on different materials (see Fig. 7b for comparison with experiments) (Boyle and Spence 2013; Nieh and Wadsworth 2006; Klueh 2005). It is generally understood to signify thermal activation processes at low stress, changing over to stress activation for reasons that have not been resolved heretofore, particularly regarding the roles of atomic diffusion and shear deformation.

Figure 8 shows the spatial distributions of nonaffine displacement and of deviatoric strain at low and high stresses. At the low stress of 64 MPa, the features of

Fig. 7 (**a**) Stress dependence of creep rate from simulations with two different system sizes of 25,000 and 10,000 atoms. (**b**) Stress variation of creep rate, the two sets of experimental data 1 (crosses) and data 2 (triangles) are adapted from Lu et al. (2003) and Nieh and Wadsworth (2006), respectively. σ_0 and $\dot{\varepsilon}_0$ are threshold values in the creep rate-stress curve
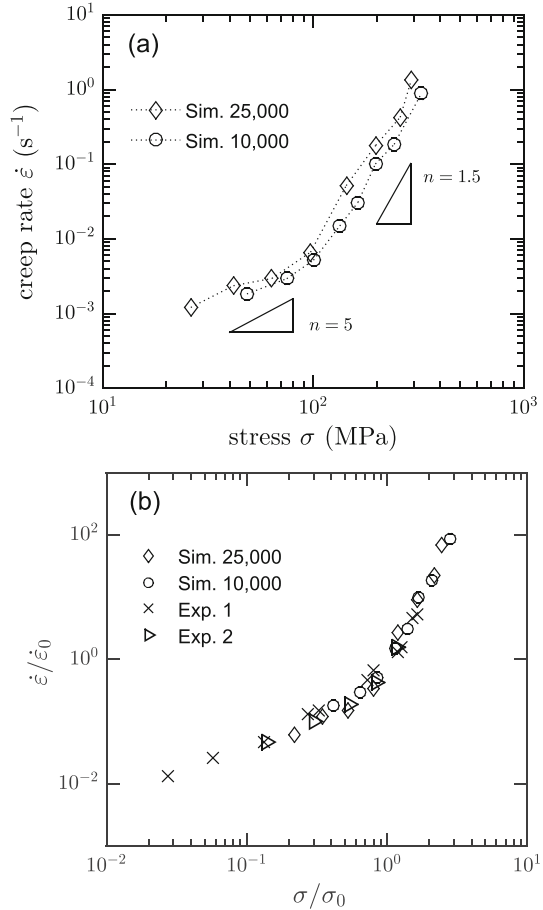
the nonaffine displacement and local strain can be essentially attributed to thermally processes. As shown in Fig. 7c, the thermally activated atomic diffusion produces little local strain burst (local rupture) event. At high stress of 258 MPa indicated by Fig. 8b, the displacement magnitude of the sites of high mobility is about a factor of three larger than the displacements at low stress. Moreover, when looking at Fig. 7d, we see the presence of bi-directional deformation which corresponds to local strain bursts, a striking effect of stress-induced deformation.

We believe the creep rate upturn delineates a regime of nonlinear response in the rheological behavior of amorphous materials, a consequence of the nonlinear coupling between atomic diffusion and shear-induced deformation. More generally, it indicates an interplay between thermal and stress-activated processes, involving loading effects, thermal noise, and stress-induced fluctuations.
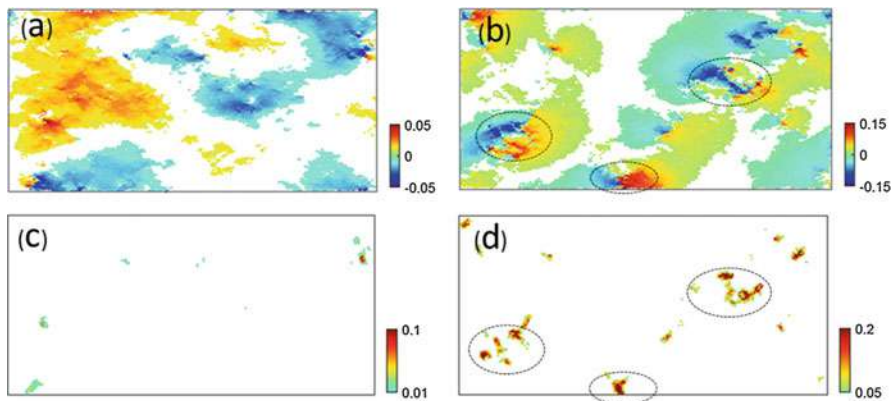
Fig. 8 Spatial distributions of nonaffine displacement (**a**) and (**b**) and of deviatoric strain (**c**) and (**d**). All measures are time incremental, taken over a time interval of 0.01 s. Atoms are colored according to magnitude of their displacements or strains. At each level the left and right panels denote low stress (64 MPa) and high stress (258 MPa), respectively

3.2 Slip Avalanches of Amorphous Plasticity Under Constant Strain Rates

Discrete stress relaxations (slip avalanches), the serrations in stress-strain responses to mechanical deformation are ubiquitous in the dynamics of amorphous plasticity (Schuh et al. 2007; Antonaglia et al. 2014a). A common characteristic of slip avalanches is the strain-rate sensitivity which becomes more pronounced as the rate is lowered (Schuh et al. 2002, 2007; Antonaglia et al. 2014b; Harris et al. 2016). To understand the atomistic mechanisms of stress relaxation at a wide range of strain rate, we implement an atomistic simulation algorithm (Cao et al. 2014a, 2018) that can reach timescales in the range of experiments. To model constant strain-rate simulation, we first select a strain-rate value and use the transition state theory expression $\dot{\varepsilon} = \dot{\gamma}_0 \exp(-E_b/k_B T)$ to determine a value for the activation barrier E_b, where γ_0 is the attempt frequency prefactor and k_B is the Boltzmann constant. After determination of the activation barrier E_b for a strain rate $\dot{\varepsilon}$, we perform the following steps:

1. Perform energy minimization to bring the system to the nearest local energy minimum state.
2. Run the autonomous basin climbing (ABC) algorithm (Cao et al. 2012) to generate a transition state pathway (TSP) trajectory consisting of local energy minima and saddle points, stopping the ABC simulation when an activation energy barrier greater than E_b is found.
3. Examine the local energy minima explored by the ABC algorithm, and perform a standard Monte Carlo (MC) run to pick the most likely state according to

the probability $p = \exp\left(\Delta E_{(i,j)}/k_\text{B}T\right)$, where $\Delta E_{(i,j)}$ is the energy difference between local minima i and j. Save the associated atomic configurations.

4. Apply a small strain increment $\Delta\varepsilon = 10^{-4}$ to the system, perform energy minimization, and repeat steps 1, 2, and 3. Each step thus contributes a strain increment $\Delta\varepsilon$ to the overall imposed strain ε. The associated atomic configurations stored in the process can be used to calculate the corresponding single-atom displacement and the local deviatoric strain. The iteration continues until the strain ε reaches 18%.

5. For quasistatic (QS) simulation, $E_b = 0$, so there is no need to run steps 2 an 3. The simulation thus consists of iterating on steps 1 and 4, performing energy minimization after each strain increment.

In Fig. 9a we show the simulated responses of system-level stress to uniaxial compression at three distinct values of strain rates. The curve QS (quasistatic) (Falk and Maloney 2010) is obtained by potential energy minimization, which is effectively the limit of high strain rate (Fan et al. 2013; Cao et al. 2014b). The other two curves correspond, respectively, to constant strain rates of $2.2 \times 10^7\,\text{s}^{-1}$ and $4.6 \times 10^{-2}\,\text{s}^{-1}$, which are typical of conventional MD simulations and laboratory mechanical testing, respectively.

The overall response at each strain rate shows initially elastic loading up to a yield point, followed by a series of stress drops of varying magnitudes. The effects of strain rate are first a lowering of the entire stress-strain curve as the strain rate decreases from QS to MD to laboratory testing. Secondly, a significant sharpening of the serrations occurs in this progression of strain rates. The variation of the peak (yield) stress with strain rate, known as the stress overshoot, is also evident. In Fig. 9b portions of the flow curve at the three strain rates are shown in greater detail to illustrate the significant sharpening in avalanche response as the strain rate is lowered (Maaß et al. 2011; Antonaglia et al. 2014b). At the lowest strain rate, we see well-resolved intermittent minor relaxations interspersed between large events throughout the entire flow regime. Similar behavior, suggesting a distinction between small and large avalanches, is observed experimentally (Antonaglia et al. 2014a; Wright et al. 2013; Maaß et al. 2011).

In Fig. 9c we show the corresponding deformations at the lowest strain rate associated with the onset of yielding and subsequent flow. In the four panels, one sees the spatial distributions of atomic sites with high deviatoric strain, color coded to indicate the magnitude of D^2_min. The distributions are cumulative strains (with reference to the initial undeformed configuration) during compression at the four numbered stages labeled in Fig. 9a: during elastic loading (point 1), after yielding (2), and during serrated flow (3–4). At 4% strain, we see only a few high-strain sites distributed rather randomly. Immediately after the onset of yielding, at 4.83% strain, a band of high-strain sites spanning the system has clearly formed. Notice also the appearance of a surface step. After the appearance of the shear band, the subsequent plastic flow is essentially dominated by this shear localized region, involving sliding and thickening. For example, with further compression through 10% to 16% strain,
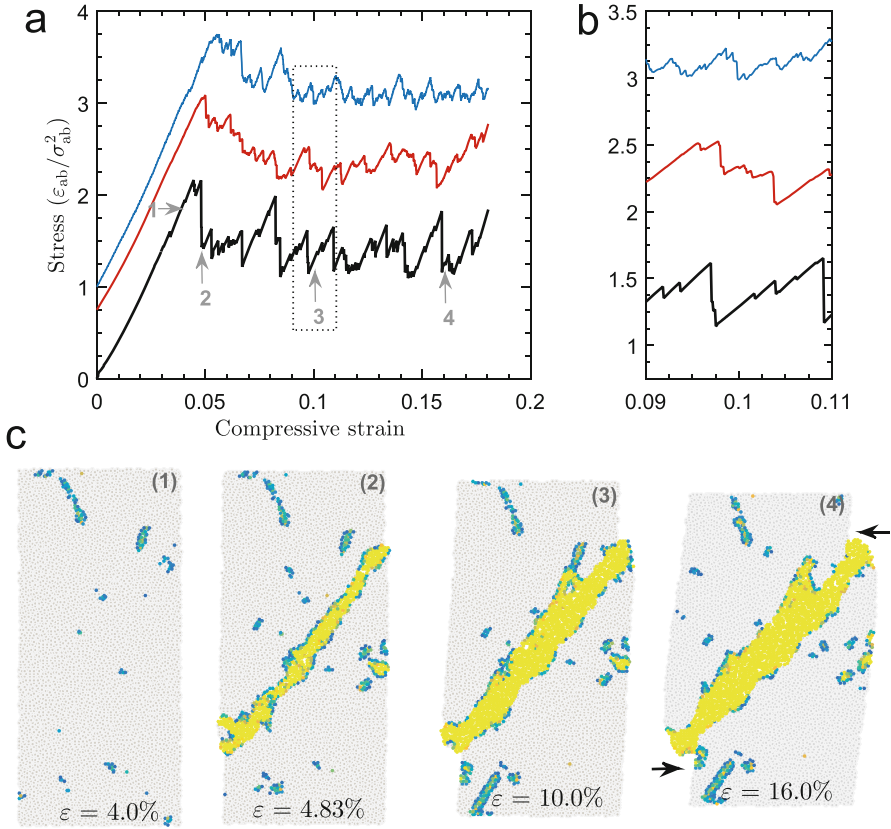
Fig. 9 (**a**) Stress-strain response obtained at three strain rates: QS (top), $2.2 \times 10^7 \, \mathrm{s}^{-1}$ (middle), and $4.6 \times 10^{-2} \, \mathrm{s}^{-1}$ (bottom). The two upper curves have been shifted vertically by 1.0 and 0.75, respectively, to facilitate visual inspection. Arrows indicate the four stages of deformation corresponding to the four panels shown in (**c**). (**b**) Enlarged view of dotted box in (**a**). (**c**) Deviatoric strain distributions in the elastic response regime, after yielding and during the flow stage at a strain rate of $4.6 \times 10^{-2} \, \mathrm{s}^{-1}$. The red spheres indicate atoms subjected to large deviatoric strain, and atoms with strain $D_{\mathrm{min}}^2 < 1$ are colored gray. The arrows point to surface steps at strain of 16.0%. (Images modified from Cao et al. 2018)

the shear band has expanded in thickness as shown in the third and fourth panels of Fig. 9c. Correspondingly, the edge steps have become larger to accommodate the increased strain in the system (Shan et al. 2008).

To probe the nanoscale deformation and flow processes associated with an avalanche, we examine the deviatoric strain D_{min}^2 and the nonaffine atomic displacement. Figure 10 shows the spatial maps of D_{min}^2 for small and large avalanches at the steady-state flow stage of deformation. In Fig. 10a, which pertains to small avalanches, one sees a few local sites of significant strain magnitude. This indicates that during a minor stress drop, there is relatively little activity involving
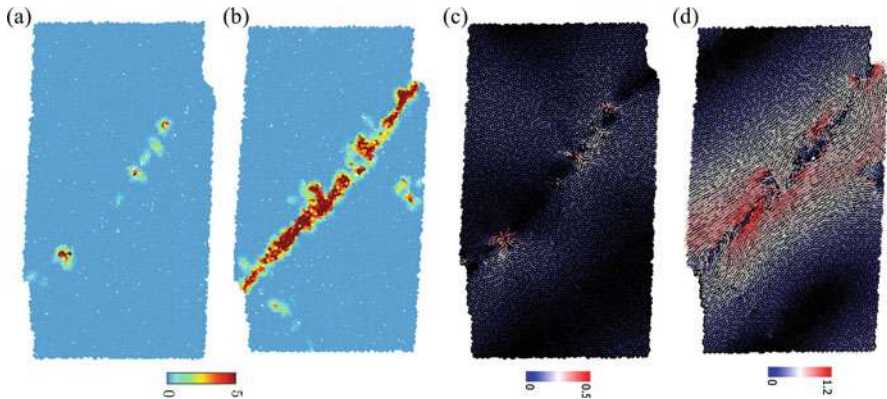
Fig. 10 Visualization of local plastic strains associated with two typical avalanches, small (**a**) and large (**b**). Atoms are color coded by the magnitude of $D^2_{\min}$. The imposed strains in the two cases are 8.3% and 9.7%, respectively. Visualization of nonaffine displacement associated with small (**a**) and large (**b**) avalanches. Arrows are color coded by the nonaffine displacement magnitude. Note that the color scale is different for (**a**) and (**b**). The imposed strains in the two cases are 8.3%

localized shear transformations, and the few activated sites are spread out along the established shear band. In contrast, as seen in Fig. 10b, major stress relaxations involve a large number of activated sites at high strain magnitudes, clustered along a characteristic band-like region. Comparing Fig. 10a, b, one has a clear picture of the different roles that small and large avalanches play in maintaining serrated flow. Through the nonaffine displacements, we can assess atomic motions associated with internal relaxation during system flow. In Fig. 10c–d, we again see the distinction between small and large avalanches now visualized through the non-affine atom displacement. Comparing these results with the distributions of plastic strain production, we have a consistent picture of the molecular processes associated with serrated flow, resolved between small and large avalanches and between the single-particle (diffusion) and collective (cluster deformation) degrees of freedom. The processes are clearly spatially heterogeneous and temporally intermittent. Such details provide meaningful complements to current studies of the mechanisms of amorphous plasticity, based on combining theoretical modeling (Antonaglia et al. 2014a) with high temporal resolution mechanical testing (Wright et al. 2016).

The fact that the system response is different in avalanche sizes implies the existence of two response regimes. The actual evolution of avalanche size suggests the small avalanches can be regarded as precursor events or an incubation stage leading up to the occasional major relaxation event. We see the large avalanches as a form of nonlinear response involving the concerted action of thermally activated atomic diffusion and stress-activated shear of local clusters of atoms. Such responses are characterized by spatial and temporal fluctuations that have been referred to as dynamical heterogeneities. Recognition of this kind of response can stimulate further theoretical work in modeling the weakening mechanism that underlies the onset of avalanches.
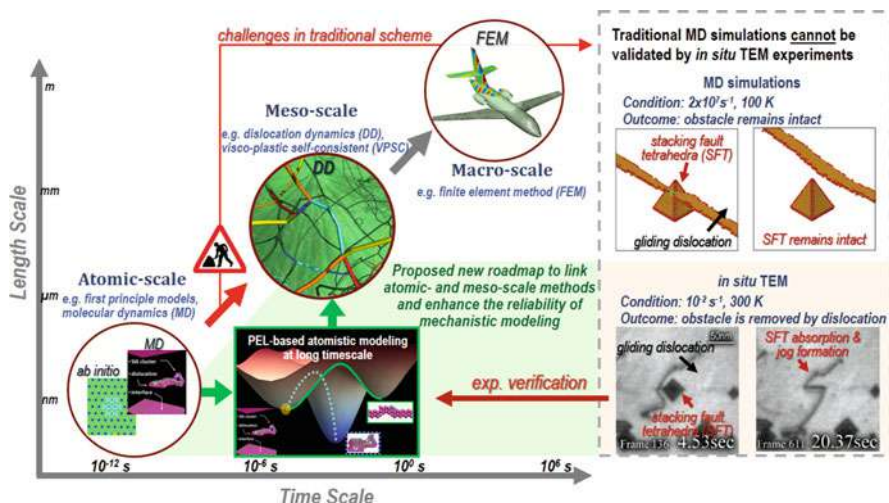
Fig. 11 Typical paradigm of multiscale modeling on dislocation-related problems consists of three different levels. At atomic scale, first principle calculation (e.g., ab initio) and molecular dynamics (MD) are widely used. At mesoscale, dislocation dynamics (DD) and visco-plastic self-consistent (VPSC) methods are most popular. And at macroscale, finite element method (FEM) is the dominant technique. It is expected that the outcomes from finer-scale models are fed into higher-level models as input parameters, which would then enable the inclusion of multi-physical processes at different scales. However, such traditional scheme, particularly the interface from MD to DD/VPSC, could be ill-informed. To be more specific, the right box shows the interaction mechanisms between a gliding dislocation and a stacking fault tetrahedra (SFT, a vacancy-type cluster) in FCC metals probed by MD simulation and by in situ TEM experiment, respectively. In MD study the SFT is found intact after cutting by the dislocation. But in real experiment, where the timescale is ten orders of magnitude longer than in MD simulation, the SFT is apparently absorbed and removed by the dislocation. Such discrepancies demonstrate that, due to its intrinsic timescale limitation, traditional MD simulation cannot capture the accurate and complete knowledge needed for the higher-level models. Given recent progress on the long-timescale atomistic simulation techniques, a new path via the PEL-based modeling has shown great promise in probing materials behavior at realistic timescale while still retaining the atomistic details

4   Summary and Outlook

It is widely appreciated that material microstructural evolution is an immensely complicated multi-physics phenomenon involving spatial and temporal scales ranging over many orders of magnitude. To establish a modeling and simulation framework to address the fundamental molecular mechanisms and to transfer such details to the higher-level models at meso- and macroscales is certainly a grand challenge.

As an example, Fig. 11 represents a well-accepted paradigm for the multiscale modeling of dislocation-related problems. At the macroscopic scale or continuum level, the finite element method (FEM) has been regarded as the most appropriate technique (Needleman and Van der Giessen 2005; Zhong and Zhu 2008). The fidelity of FEM calculation largely depends on whether its mesh can be efficiently

handled and whether there are reliable constitutive descriptions for all the nodes. The constitutive equations for dislocation-mediated deformation are usually provided by mesoscale models, such as dislocation dynamics (DD) and visco-plastic self-consistent (VPSC) method, which have been widely applied (Lebensohn et al. 2007; Turner et al. 1999; Schwarz 1999; Zbib and de la Rubia 2002; Arsenlis et al. 2007; El-Awady et al. 2008). On the other hand, both DD and VPSC models rely on many fitting or empirically assigned parameters, such as the dislocation core structures, the critical resolved shear stress for each slip system, the strain hardening exponent, the dislocation jog/junction formation mechanisms, the dislocation climb rate, etc.. As a result, although these mesoscale models can explain many observed experiments by adjusting various parameters, the predictive power of those models remains to be further improved. To reduce the number of empirical parameters, one still needs to employ more fundamental yet computationally more expensive modeling from the atomic levels, such as first principle calculations and MD method. These atomistic simulations can capture the dislocation core structures at equilibrium environments (Dezerald et al. 2016; Chaari et al. 2014; Proville et al. 2012) and provide useful mechanistic insights into the detailed dislocation-obstacle interaction processes (Terentyev et al. 2008, 2010a, b; Osetsky et al. 2006b).

Therefore, to enable a high-fidelity multiscale modeling, the connection between atomic-scale modeling and mesoscale modeling plays a vital role, because the robustness of multiscale modeling largely relies on whether MD simulations can provide accurate and complete mechanisms or not. Unfortunately, there are circumstances where MD simulations at short timescales cannot represent the realistic mechanism in experiments. An example that supports this is a remarkable controversy between the MD prediction and experimental observation over the dislocation-obstacle interaction mechanism in FCC metals. As shown in the right box of Fig. 11, in situ TEM experiments clearly show that a perfect SFT (a vacancy-type defect cluster named stacking fault tetrahedra) is fully absorbed and removed by a single moving dislocation (Osetsky et al. 2006a, b). However, parallel MD simulations show that the same type of obstacle remains intact after cutting by a moving dislocation (Matsukawa and Zinkle 2004). While the geometries and boundary conditions have been regarded as potential factors that might lead to such discrepancy (Osetsky et al. 2006a), the most probable reason, as discussed by Matsukawa and Zinkle (Matsukawa and Zinkle 2004), is the huge gap in timescale between the in situ TEM ($\sim 10^{-2}\,\mathrm{s}^{-1}$) and MD simulations ($\sim 10^{8}\,\mathrm{s}^{-1}$).

In this chapter we have shown several examples in both crystalline and amorphous materials that the PEL-based novel modeling techniques are capable of uncovering the microstructural evolutions over wide range of timescales and directly compare with experimental observations (Fan et al. 2012, 2013). A particularly notable feature of the PEL framework is that it is fundamentally governed by the interatomic potentials between the constituent particles, and thus no phenomenological assumptions or empirical parameters are needed.

Given these advantages, one could propose a "detour" on the conventional multiscale modeling roadmap. As illustrated by the green arrows in Fig. 11, instead of directly connecting MD with mesoscale methods (Yip and Short 2013), one can first

employ the PEL-based modeling to gain a more comprehensive knowledge at longer timescale and then pass the atomistic insights obtained on-the-fly to the higher-level models. Such a modification may be expected to lead to a more predictive modeling and simulation.

References

Antonaglia J, Wright WJ, Gu X, Byer RR, Hufnagel TC, LeBlanc M, Uhl JT, Dahmen KA (2014a) Bulk metallic glasses deform via slip avalanches. Phys Rev Lett 112(15):155501

Antonaglia J, Xie X, Schwarz G, Wraith M, Qiao J, Zhang Y, Liaw PK, Uhl JT, Dahmen KA (2014b) Tuned critical avalanche scaling in bulk metallic glasses. Sci Rep 4:4382

Argon AS (1979) Plastic deformation in metallic glasses. Acta Metall 27:47–58

Armstrong RW, Arnold W, Zerilli FJ (2009) Dislocation mechanics of copper and iron in high rate deformation tests. J Appl Phys 105(2):023511–023517

Arsenlis A, Cai W, Tang M, Rhee M, Oppelstrup T, Hommes G, Pierce TG, Bulatov VV (2007) Enabling strain hardening simulations with dislocation dynamics. Model Simul Mater Sci Eng 15(6):553

Bacon DJ, Osetsky YN, Rong Z (2006) Computer simulation of reactions between an edge dislocation and glissile self-interstitial clusters in iron. Philos Mag 86(25–26):3921–3936

Bacon DJ, Osetsky YN, Rodney D (2009) Chapter 88 dislocation–obstacle interactions at the atomic level. In: Hirth JP, Kubin LP (eds) Dislocations in solids, vol 15. Elsevier, Burlington, pp 1–90

Bai Z, Fan Y (2018) Abnormal strain rate sensitivity driven by a unit dislocation-obstacle interaction in BCC Fe. Phys Rev Lett 120(12):125504

Bai XM, Voter AF, Hoagland RG, Nastasi M, Uberuaga BP (2010) Efficient annealing of radiation damage near grain boundaries via interstitial emission. Science 327(5973):1631–1634

Bakó B, Groma I, Györgyi G, Zimányi G (2006) Dislocation patterning: the role of climb in mesoscale simulations. Comput Mater Sci 38(1):22–28

Barkema GT, Mousseau N (1996) Event-based relaxation of continuous disordered systems. Phys Rev Lett 77(21):4358–4361

Boyle JT, Spence J (2013) Stress analysis for creep. Elsevier, Kent

Brechet Y, Estrin Y (1995) On the influence of precipitation on the Portevin-Le Chatelier effect. Acta Metall Mater 43(3):955–963

Bulatov VV, Hsiung LL, Tang M, Arsenlis A, Bartelt MC, Cai W, Florando JN, Hiratani M, Rhee M, Hommes G, Pierce TG, de la Rubia TD (2006) Dislocation multi-junctions and strain hardening. Nature 440(7088):1174–1178

Campbell JD, Ferguson WG (1970) The temperature and strain-rate dependence of the shear strength of mild steel. Philos Mag 21(169):63–82

Cances E, Legoll F, Marinica MC, Minoukadeh K, Willaime F (2009) Some improvements of the activation-relaxation technique method for finding transition pathways on potential energy surfaces. J Chem Phys 130(11):114711. https://doi.org/10.1063/1.3088532

Cao AJ, Cheng YQ, Ma E (2009) Structural processes that initiate shear localization in metallic glass 57:5146–5155

Cao P, Li M, Heugle RJ, Park HS, Lin X (2012) A self-learning metabasin escape algorithm and the metabasin correlation length of supercooled liquids. Phys Rev E 86:016710

Cao P, Park HS, Lin X (2013) Strain-rate and temperature-driven transition in the shear transformation zone for two-dimensional amorphous solids. Phys Rev E 88(4):042404

Cao P, Lin X, Park HS (2014a) Surface shear-transformation zones in amorphous solids. Phys Rev E 90(1):012311

Cao P, Lin X, Park HS (2014b) Strain-rate and temperature dependence of yield stress of amorphous solids via self-learning metabasin escape algorithm. J Mech Phys Solids 68: 239–250

Cao P, Short MP, Yip S (2017) Understanding the mechanisms of amorphous creep through molecular simulation. Proc Natl Acad Sci 114(52):13631–13636

Cao P, Dahmen KA, Kushima A, Wright WJ, Park HS, Short MP, Yip S (2018) Nanomechanics of slip avalanches in amorphous plasticity. J Mech Phys Solids 114:158–171

Chaari N, Clouet E, Rodney D (2014) First-principles study of secondary slip in zirconium. Phys Rev Lett 112(7):075504

Dezerald L, Rodney D, Clouet E, Ventelon L, Willaime F (2016) Plastic anisotropy and dislocation trajectory in BCC metals. Nat Commun 7:11695

Domain C, Monnet G (2005) Simulation of screw dislocation motion in iron by molecular dynamics simulations. Phys Rev Lett 95(21):215506

Dunlop JW, Bréchet YJM, Legras L, Estrin Y (2007) Dislocation density-based modelling of plastic deformation of zircaloy-4. Mater Sci Eng A 443(1–2):77–86

Dutta A, Bhattacharya M, Gayathri N, Das GC, Barat P (2012) The mechanism of climb in dislocation–nanovoid interaction. Acta Mater 60(9):3789–3798

El-Awady JA, Bulent Biner S, Ghoniem NM (2008) A self-consistent boundary element, parametric dislocation dynamics formulation of plastic flow in finite volumes. J Mech Phys Solids 56(5):2019–2035

Falk ML, Langer JS (1998) Dynamics of viscoplastic deformation in amorphous solids. Phys Rev E 57(6):7192–7205

Falk M, Maloney C (2010) Simulating the mechanical response of amorphous solids using atomistic methods. Eur Phys J B 75(4):405–413

Fan Y (2013) Atomistic simulation of defect structure evolution and mechanical properties at long time scales, PhD Thesis, Massachusetts Institute of Technology, Cambridge, USA

Fan Y, Osetsky YN, Yip S, Yildiz B (2012) Onset mechanism of strain-rate-induced flow stress upturn. Phys Rev Lett 109(13):135503. https://doi.org/10.1103/PhysRevLett.109.135503

Fan Y, Osetskiy YN, Yip S, Yildiz B (2013) Mapping strain rate dependence of dislocation-defect interactions by atomistic simulations. Proc Natl Acad Sci 110(44):17756–17761

Fan Y, Yip S, Yildiz B (2014) Autonomous basin climbing method with sampling of multiple transition pathways: application to anisotropic diffusion of point defects in HCP Zr. J Phys Condens Matter 26(36):365402

Ferguson WG, Hauser FE, Dorn JE (1967) Dislocation damping in zinc single crystals. Br J Appl Phys 18(4):411

Follansbee PS, Kocks UF (1988) A constitutive description of the deformation of copper based on the use of the mechanical threshold stress as an internal state variable. Acta Metall 36(1): 81–93

Follansbee PS, Regazzoni G, Kocks UF (1984) Mechanical properties of materials at high rates of strain (9th ed.). The Institute of Physics, Inst Phys Conf Ser, p 71

Gordon PA, Neeraj T, Li Y, Li J (2010) Screw dislocation mobility in BCC metals: the role of the compact core on double-kink nucleation. Model Simul Mater Sci Eng 18(8):085008

Gussev MN, Field KG, Busby JT (2015) Deformation localization and dislocation channel dynamics in neutron-irradiated austenitic stainless steels. J Nucl Mater 460:139–152

Haixuan X, Yuri NO, Roger ES (2012) Self-evolving atomistic kinetic Monte Carlo: fundamentals and applications. J Phys Condens Matter 24(37):375402

Harris MB, Watts LS, Homer ER (2016) Competition between shear band nucleation and propagation across rate-dependent flow transitions in a model metallic glass. Acta Mater 111:273–282

Henkelman G, Jonsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111(15):7010–7022. https://doi.org/10.1063/1.480097

Hoge KG, Mukherjee AK (1977) The temperature and strain rate dependence of the flow stress of tantalum. J Mater Sci 12(8):1666–1672

Huang Y, Shen J, Chiu Y, Chen J, Sun J (2009) Indentation creep of an Fe-based bulk metallic glass. Intermetallics 17(4):190–194

Hull D, Bacon D (2011) Introduction to dislocations (5th ed.). Butterworth-Heinemann, Oxford, ISBN 9780080966724

Iyer M, Radhakrishnan B, Gavini V (2015) Electronic-structure study of an edge dislocation in aluminum and the role of macroscopic deformations on its energetics. J Mech Phys Solids 76:260–275

Jumonji K, Ueta S, Miyahara A, Kato M, Sato A (1996) Rapid work hardening caused by cube cross slip in Ni_3Al single crystals. Philos Mag A 73(2):345–364

Kabir M, Lau TT, Rodney D, Yip S, Van Vliet KJ (2010) Predicting dislocation climb and creep from explicit atomistic details. Phys Rev Lett 105(9):095501

Kioussis NG, Ghoniem NM (2010) Modeling of dislocation interaction with solutes, nano-precipitates and interfaces: a multiscale challenge. J Comput Theor Nanosci 7(8):1317–1346

Klueh R (2005) Elevated temperature ferritic and martensitic steels and their application to future nuclear reactors. Int Mater Rev 50(5):287–310

Kocks UF, Argon AS, Ashby MF (1975) Thermodynamics and Kinetics of Slip (9th ed.). In: Chalmers B, Christian JW, Massalski TB (eds) Progress in Materials Science, Pergamon Press, Oxford, 19(9):110

Krisponeit J-O, Pitikaris S, Avila KE, Küchemann S, Krüger A, Samwer K (2014) Crossover from random three-dimensional avalanches to correlated nano shear bands in metallic glasses. Nat Commun 5:3616

Kumar A, Hauser FE, Dorn JE (1968) Viscous drag on dislocations in aluminum at high strain rates. Acta Metall 16(9):1189–1197

Kushima A, Lin X, Li J, Eapen J, Mauro JC, Qian X, Diep P, Yip S (2009a) Computing the viscosity of supercooled liquids. J Chem Phys 130(22):224504

Kushima A, Lin X, Li J, Qian X, Eapen J, Mauro JC, Diep P, Yip S (2009b) Computing the viscosity of supercooled liquids. II. Silica and strong-fragile crossover behavior. J Chem Phys 131(16):164505–164509

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci 99(20):12562–12566

Lebensohn RA, Tomé CN, CastaÑeda PP (2007) Self-consistent modelling of the mechanical behaviour of viscoplastic polycrystals incorporating intragranular field fluctuations. Philos Mag 87(28):4287–4322

Lebyodkin MA, Brechet Y, Estrin Y, Kubin LP (1995) Statistics of the catastrophic slip events in the portevin-le chatelier effect. Phys Rev Lett 74(23):4758–4761

Lu J, Ravichandran G, Johnson WL (2003) Deformation behavior of the Zr 41.2 Ti 13.8 Cu 12.5 Ni 10 Be 22.5 bulk metallic glass over a wide range of strain-rates and temperatures. Acta Mater 51(12):3429–3443

Maaß R, Klaumünzer D, Löffler J (2011) Propagation dynamics of individual shear bands during inhomogeneous flow in a Zr-based bulk metallic glass. Acta Mater 59(8):3205–3213

Matsukawa Y, Zinkle SJ (2004) Dynamic observation of the collapse process of a stacking fault tetrahedron by moving dislocations. J Nucl Mater 329–333, Part B(0):919–923

Mendelev MI, Han S, Srolovitz DJ, Ackland GJ, Sun DY, Asta M (2003) Development of new interatomic potentials appropriate for crystalline and liquid iron. Philos Mag 83(35):3977–3994

Needleman A, Van der Giessen E (2005) Elasticity: finite element modeling A2 – Buschow, K.H. Jürgen. Elsevier, Oxford, pp 1–6

Nieh T, Wadsworth J (2006) Homogeneous deformation of bulk metallic glasses. Scr Mater 54(3):387–392

Norfleet DM, Dimiduk DM, Polasik SJ, Uchic MD, Mills MJ (2008) Dislocation structures and their relationship to strength in deformed nickel microcrystals. Acta Mater 56(13):2988–3001

Onimus F, Béchade J-L (2009) A polycrystalline modeling of the mechanical behavior of neutron irradiated zirconium alloys. J Nucl Mater 384(2):163–174

Onimus F, Monnet I, Béchade JL, Prioul C, Pilvin P (2004) A statistical TEM investigation of dislocation channeling mechanism in neutron irradiated zirconium alloys. J Nucl Mater 328(2–3):165–179

Osetsky YN, Matsukawa Y, Stoller RE, Zinkle SJ (2006a) On the features of dislocation–obstacle interaction in thin films: large-scale atomistic simulation. Philos Mag Lett 86(8):511–519

Osetsky YN, Rodney D, Bacon DJ (2006b) Atomic-scale study of dislocation–stacking fault tetrahedron interactions. Part I: mechanisms. Philos Mag 86(16):2295–2313

Preston DL, Tonks DL, Wallace DC (2003) Model of plastic deformation for extreme loading conditions. J Appl Phys 93(1):211–220

Proville L, Rodney D, Marinica M-C (2012) Quantum effect on thermally activated glide of dislocations. Nat Mater 11(10):845–849

Regazzoni G, Kocks UF, Follansbee PS (1987) Dislocation kinetics at high strain rates. Acta Metall 35(12):2865–2875

Remington BA, Allen P, Bringa EM, Hawreliak J, Ho D, Lorenz KT, Lorenzana H, McNaney JM, Meyers MA, Pollaine SW, Rosolankova K, Sadik B, Schneider MS, Swift D, Wark J, Yaakobi B (2006) Material dynamics under extreme conditions of pressure and strain rate. Mater Sci Technol 22(4):474–488

Rodney D, Proville L (2009) Stress-dependent peierls potential: influence on kink-pair activation. Phys Rev B 79(9):094108

Schuh C, Nieh T, Kawamura Y (2002) Rate dependence of serrated flow during nanoindentation of a bulk metallic glass. J Mater Res 17(7):1651–1654

Schuh CA, Hufnagel TC, Ramamurty U (2007) Mechanical behavior of amorphous alloys. Acta Mater 55(12):4067–4109

Schwarz KW (1999) Simulation of dislocations on the mesoscopic scale. I. Methods and examples. J Appl Phys 85(1):108–119

Sentjabrskaja T, Chaudhuri P, Hermes M, Poon W, Horbach J, Egelhaaf S, Laurati M (2015) Creep and flow of glasses: strain response linked to the spatial distribution of dynamical heterogeneities. Sci Rep 5:11884

Shan Z, Li J, Cheng Y, Minor A, Asif SS, Warren O, Ma E (2008) Plastic flow and failure resistance of metallic glass: insight from in situ compression of nanopillars. Phys Rev B 77(15):155419

Shin CS, Fivel MC, Verdier M, Robertson C (2005) Dislocation dynamics simulations of fatigue of precipitation-hardened materials. Mater Sci Eng A 400–401:166–169

Siebenbürger M, Ballauff M, Voigtmann T (2012) Creep in colloidal glasses. Phys Rev Lett 108(25):255701

Sorensen MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 112(21):9599–9606

Spaepen F (1977) A microscopic mechanism for steady state inhomogeneous flow in metallic glasses. Acta Metall 25(4):407–415

Steinberg DJ, Lund CM (1989) A constitutive model for strain rates from 10[sup - 4] to 10[sup 6] s[sup - 1]. J Appl Phys 65(4):1528–1533

Suo Z (2000) Evolving material structures of small feature sizes. Int J Solids Struct 37(1):367–378

Tao W, Cao P, Park HS (2018a) Atomistic simulation of the rate-dependent ductile-to-brittle failure transition in bicrystalline metal nanowires. Nano Lett 18(2):1296–1304

Tao W, Cao P, Park HS (2018b) Superplastic creep of metal nanowires from rate-dependent plasticity transition. ACS Nano 12(5):4984–4992

Terentyev D, Grammatikopoulos P, Bacon DJ, Osetsky YN (2008) Simulation of the interaction between an edge dislocation and a$<$ 1 0 0$>$ interstitial dislocation loop in α-iron, Acta Materialia 56(18):5034–5046

Terentyev D, Osetsky YN, Bacon DJ (2010a) Competing processes in reactions between an edge dislocation and dislocation loops in a body-centred cubic metal. Scr Mater 62(9):697–700

Terentyev D, Bonny G, Domain C, Pasianot RC (2010b) Interaction of a 1/2< 111> screw dislocation with Cr precipitates in BCC Fe studied by molecular dynamics. Phys Rev B 81(21):214106

Turner PA, Tomé CN, Christodoulou N, Woo CH (1999) A self-consistent model for polycrystals undergoing simultaneous irradiation and thermal creep. Philos Mag A 79(10):2505–2524

Van der Giessen E, Deshpande VS, Cleveringa HHM, Needleman A (2001) Discrete dislocation plasticity and crack tip fields in single crystals. J Mech Phys Solids 49(9):2133–2153

Voter AF (1997) A method for accelerating the molecular dynamics simulation of infrequent events. J Chem Phys 106(11):4665–4677

Wang H, Xu DS, Rodney D, Veyssire P, Yang R (2013) Atomistic investigation of the annihilation of non-screw dislocation dipoles in Al, Cu, Ni and γ-TiAl. Model Simul Mater Sci Eng 21(2):025002

Wei Y, Zhigang S (1996) Global view of microstructural evolution: energetics, kinetics and dynamical systems. Acta Mech Sinica 12(2):144–157

Wright WJ, Byer RR, Gu X (2013) High–speed imaging of a bulk metallic glass during uniaxial compression. Appl Phys Lett 102(24):241920

Wright WJ, Liu Y, Gu X, Van Ness KD, Robare SL, Liu X, Antonaglia J, LeBlanc M, Uhl JT, Hufnagel TC et al (2016) Experimental evidence for both progressive and simultaneous shear during quasistatic compression of a bulk metallic glass. J Appl Phys 119(8):084908

Yamakov V, Wolf D, Phillpot SR, Mukherjee AK, Gleiter H (2002) Dislocation processes in the deformation of nanocrystalline aluminium by molecular-dynamics simulation. Nat Mater 1(1):45–49

Yan X, Cao P, Tao W, Sharma P, Park HS (2016) Atomistic modeling at experimental strain rates and timescales. J Phys D Appl Phys 49(49):493002

Yip S (2003) Synergistic science. Nat Mater 2(1):3–5

Yip S (2016) Understanding the viscosity of supercooled liquids and the glass transition through molecular simulations. Mol Simul 42(16):1330–1342

Yip S, Short MP (2013) Multiscale materials modelling at the mesoscale. Nat Mater 12(9):774

Zbib HM, de la Rubia TD (2002) A multiscale model of plasticity. Int J Plast 18(9):1133–1163

Zerilli FJ, Armstrong RW (1987) Dislocation-mechanics-based constitutive relations for material dynamics calculations. J Appl Phys 61(5):1816–1825

Zhong Y, Zhu T (2008) Simulating nanoindentation and predicting dislocation nucleation using interatomic potential finite element method. Comput Methods Appl Mech Eng 197(41–42):3174–3181

Zhu T, Li J, Samanta A, Leach A, Gall K (2008) Temperature and strain-rate dependence of surface dislocation nucleation. Phys Rev Lett 100(2):025502

Quantized Dislocations for Functional and Quantum Materials

12

Mingda Li and Ricardo Pablo-Pedro

Contents

Abstract

Dislocations have a profound influence on materials' functional properties. In this perspective, we discuss the recent development of quantized dislocations – a theoretical tool that aims to compute the role of dislocations on materials' functionalities, at a full quantum field theoretical level. After a brief discussion of the motivation and a pedagogical introduction of quantization, we focus on a few case studies of dislon theory, to see how dislon can be applied to solve a given materials' functionality problem and lead to new predictions. We conclude by envisioning a few more open questions. With the aid of the powerful quantum field theory, the dislon approach may enable plenty of opportunities to compute multiple functional and quantum properties in a dislocated crystal at a new level of clarity.

M. Li (✉) · R. Pablo-Pedro
Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: mingda@mit.edu; ripablo@mit.edu

1   Introduction

Dislocations affect materials' mechanical properties and functionalities, such as electronic structure, optical properties, thermal transport, magnetic ordering, and superconductivity. For decades, the dislocation research has been centering on mechanical behaviors, such as crystal plasticity (Hirth and Lothe 1992; Nabarro 1967). However, given the flourishing development of novel functional materials in recent years, the proper modeling of dislocation functionalities beyond mechanical aspects is gaining more and more importance, yet faces a series of challenges. As an extended defect with internal structure, a realistic dislocation is not only a quenched disorder composed of a distribution of strain field, but also subject to strong dynamical vibration and material-dependent Coulomb interaction in a complex interaction environment. In this sense, many dislocation studies are only describing a partial feature of a realistic dislocation, where its definition $\oint_L d\mathbf{u} = -\mathbf{b}$ is often not respected. While all these factors are taken into account to fully characterize a realistic dislocation, the calculation on the functionalities poses another level of challenge. On the one hand, due to the long-range nature of a dislocation's stress field, the first-principles calculation with dislocations requires a large supercell – could be as high as $N \sim 1000$ atoms (Wang et al. 2014). To carry out the response calculations, however, the computational complexity $\sim O(N^4)$ using the density functional perturbation theory is simply too high to be realistic (Baroni et al. 2001). On the other hand, since almost all those functionalities can be traced back to a microscopic quantum origin, a classical description of dislocation may simply be incapable to be integrated into a quantum theory in order to describe a complex quantum phenomenon.

In fact, a number of open questions remain in the field of dislocation functionalities. For electronic structure, it is known that the electrical resistivity in dislocated metals can have a particular type of electron-line defect resonance scattering, shown in Fig. 1a (Brown 1977; Karolik and Luhvich 1994). Although the resonance scattering model did explain some experimental data, the origin of such resonance is unclear: "There must be some general mechanism underlying this phenomenon... This is still very much an open question..." (Brown 1977). For optical properties, it is known that dislocation can induce luminescence (Drozdov et al. 1976) with four significant luminescence peaks (called D1 – D4) at low-temperature, which may survive even at room temperature after sample treatment (Fig. 1b). This phenomenon enabled dislocation based light-emitting diode (LED) applications (Kveder et al. 2004). However, even after three-decades-long research, the microscopic origin of these peaks is not yet fully understood, particularly "The origin of the D2-line is still under discussion" (Reiche and Kittler 2016). For the thermal transport, there has been a decades-long debate, arguing whether the dislocation-phonon interaction is static or dynamic in nature. Despite different temperature dependence, that $k_{\text{static}} \propto T^2$ while $k_{\text{dynamic}} \propto T^3$, carefully planned experiments may exhibit a mixed behavior (Fig. 1c). In particular, one recent first-principles calculations of dislocations mentioned that "Because of the break-
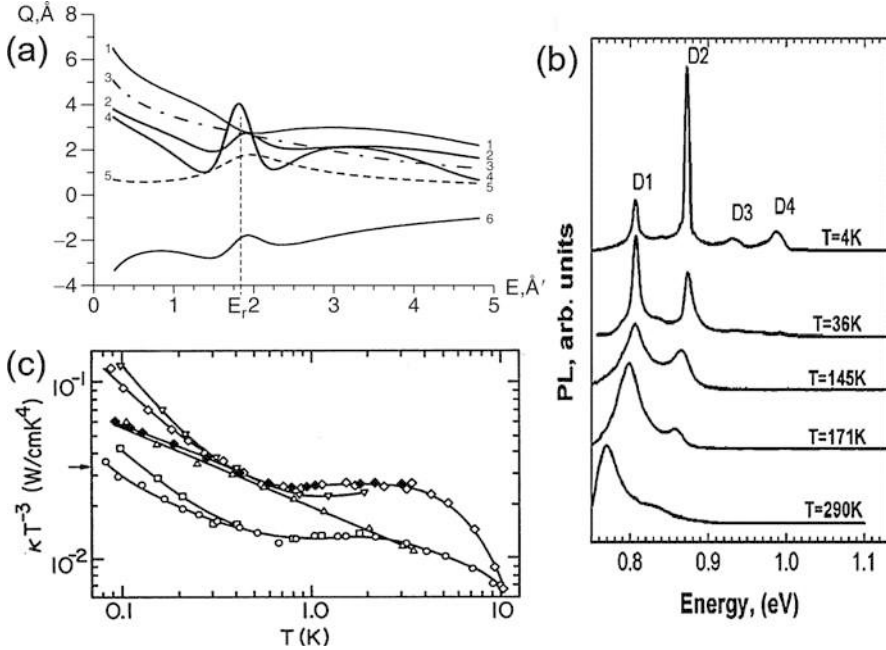
Fig. 1 (**a**) The computed cross section of electron-dislocation scattering, showing the resonance feature. (**b**) The temperature dependence of the dislocation luminescence in silicon. (**c**) The temperature-dependent thermal conductivity of prototypical dislocated LiF, showing a mixed T2 and T3 behavior. (Figures are adapted from (Kveder et al. 2005; Roth and Anderson 1979))

down of the Born approximation, earlier literature models fail, even qualitatively" (Wang et al. 2017). As to superconductivity, P.W. Anderson asserted that "The (superconducting) transition temperature (with defects) will always be slightly smaller for the scattered states than they would be in the pure case" (Anderson 1959), yet many experiments show otherwise. The anisotropic superconducting gap explains the transition temperature enhancement effect for point impurities (Markowitz and Kadanoff 1963), but a quantitative comparison for dislocations without using empirical parameter has been missing. These open questions, along with the unbearably high computational cost, indeed call for a much better approach to tackle the dislocation functionality problems at a more fundamental level.

This perspective article is a self-contained introduction on the recent theoretical progress of the so-called dislon theory. Dislon is a quasiparticle that aims to solve the above dislocation functionalities issues by directly quantizing a classical dislocation (Fu and Li 2017; Li 2019; Li et al. 2017a, b, c, 2018; Pablo-Pedro et al. 2018). We believe that there are a number of unique advantages adopting this quantized dislocation approach, mainly formalism simplicity and strong predictive power.

Formalism simplicity: The simplicity has a multifold meaning. First is the procedure to incorporate a dislocation into an existing system. Apparently, there are plenty of existing approaches that can introduce a dislocation. For instance, for electronic property studies, a dislocation is often modeled as a line charge. With a scattering Coulomb potential $V(r)$, its Fourier transformation $V(q)$ gives the scattering strength, from which the electron-dislocation relaxation time can be obtained using Fermi's golden rule (Jena et al. 2000; Look and Sizelove 1999; Weimann et al. 1998). Taking the dislon approach, a dislocation is introduced by adding the dislon Hamiltonian H_D and the interaction Hamiltonian H_I, into the original Hamiltonian H_0 for a pristine system. Given the arbitrary freedom to choose H_0, a dislocation can always be introduced properly. This leads to the second fold of simplicity that the Hamiltonian approach standardizes the procedure to compute all functional properties. With the knowledge of a given H_0, the electrical conductivity, thermal conductivity, thermopower, optical absorption, etc., are all computable using standard quantum many-body approaches, such as linear response theory (Mahan 2000). Now the understanding of the role that dislocations may play is then reduced to a standard linear-response calculation of functionalities, but with the total Hamiltonian $H_0 + H_I + H_D$. The third meaning of simplicity lies in the form of dislon Hamiltonian H_D, which has great mathematical simplicity to tackle, but with all dislocation effects – strain, dynamic, Coulomb – incorporated simultaneously. This comprehensive description of dislocation is in sharp contrast with a potential scattering approach. Taking the popular line charge model of dislocation as an example, we see that the Burgers vector **b** does not even appear explicitly in the expression of relaxation time, which is unphysical to some extent.

Strong predictive power: Besides the formalism simplicity, the main advantage of dislon theory lies in the strong predictive power. This is a natural consequence by adopting a quantum field theoretical approach, since it can seamlessly incorporate all other interactions and correlation effects to an arbitrarily high order. The interaction effects are essential for a realistic scenario, for instance, dislocations may coexist with point defects, phonons, and electrons in a real crystal, and we could always pick up the relevant degrees of freedom that are of interest for a given problem. For instance, if we are interested in phonon-dominant thermal transport, then the Coulomb charge model of dislocation cannot interact with phonon, hence cannot be used for thermal transport study. The dislon theory, on the other hand, is a unified approach. Since it already contains the strain effect in the dislon field, the interaction with phonon becomes straightforward. The correlation effect and higher-order scattering, on the other hand, are both important toward qualitatively novel phenomena (Rammer 1998). This also distinguishes the dislon theory from a semiclassical model. Again taking the line charge model as an example, we will know for sure that Coulomb scattering is guaranteed to happen, but meanwhile, full predictability is lost since there is no information how dynamic and strain effect affects the electronic structure. A quantum field theory, on the other hand, is still ab initio in nature, which retains a full predictive power without loss of information from the starting point.

2   Dislon as Quantized Dislocation

To see how a dislocation can be quantized into a quantized operator form, we noticed that both a dislocation and a phonon are atomic lattice displacement **u**, with the major difference coming from the dislocation's topological constraint $\oint_L d\mathbf{u} = -\mathbf{b}$. Therefore, two major pillars are needed to quantize a dislocation. One is the lesson of quantum procedure, learned from the more familiar phonon quantization; the other is the unique features of a classical dislocation that distinguishes from a classical lattice wave.

The lesson from phonon quantization: We briefly outline the familiar phonon quantization first since the dislocation quantization shares some formalism similarity. A comprehensive procedure for phonon quantization can be found in quantum many-body monographs (Fetter and Walecka 2003; Mahan 2000). For a system with N atoms with mass m, each atom located at coordinate $\mathbf{R}_n$ has its own displacement $\mathbf{u}_n$ and momentum $\mathbf{p}_n (n = 1, 2, \ldots N)$. The total kinetic energy gives

$$K = \sum_{n=1}^{N} \frac{|\mathbf{p}_n|^2}{2m} = \sum_{n=1}^{N} \sum_{i=1}^{3} \frac{(p_{ni})^2}{2m} \tag{1}$$

where p_{ni} is the ith Cartesian component of the vector $\mathbf{p}_n$. The total potential energy can be written as

$$U = \frac{1}{2} \sum_{m,n=1}^{N} \sum_{i,j=1}^{3} u_{mi} V_{ij}^{mn} u_{ni} \tag{2}$$

where u_{ni} is the ith Cartesian component of displacement $\mathbf{u}_n$, V_{ij}^{mn} is a generalized "spring constant" for a harmonic oscillator. To obtain a quantum theory, we promote the classical dynamical variables $\mathbf{u}_n$ and $\mathbf{p}_n$ into first-quantized operators, with the following canonical quantization condition:

$$[u_{ni}, p_{mj}] = i\hbar \delta_{ij} \delta_{mn} \tag{3}$$

i.e., different atomic locations ($m \neq n$) and directions ($i \neq j$) commute with each other.

As we see, the phonons in an N-atom system Eqs. (1) and (2) appear cumbersome in a first-quantized form. However, all these can be simplified using a second quantization approach. To do so, we perform a Fourier transform:

$$\begin{aligned}
\mathbf{u}_n &= \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathbf{u}_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}_n}, \quad \mathbf{u}_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathbf{u}_n e^{i\mathbf{k}\cdot\mathbf{R}_n} \\
\mathbf{p}_n &= \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathbf{p}_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}_n}, \quad \mathbf{p}_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \mathbf{p}_n e^{i\mathbf{k}\cdot\mathbf{R}_n}
\end{aligned} \tag{4}$$

where $\mathbf{u_k}$ and $\mathbf{p_k}$ are the canonical displacement and momentum operator labeled by quantum number $\mathbf{k}$ (called crystal momentum), respectively. Then, substituting Eq. (4) back to Eqs. (1) and (2), and performing a Fourier transform to the expansion coefficient V_{ij}^{mn},

$$D_{ij}(\mathbf{k}) = \sum_{\mathbf{R} \equiv \mathbf{R}_m - \mathbf{R}_n} V_{ij}^{mn} e^{-i\mathbf{k} \cdot (\mathbf{R}_m - \mathbf{R}_n)} \tag{5}$$

where since a crystalline solid is periodic with translation symmetry, we only need to sum over the position *difference* between two lattice positions $\mathbf{R}_n$ and $\mathbf{R}_m$. Now since $D_{ij}(\mathbf{k}) = D_{ji}(\mathbf{k})$, the $3 \times 3 D(\mathbf{k})$ matrix ($\mathbf{k}$ can be considered as a parameter) can be diagonalized with real eigenvalues. The eigen-equation can be written as

$$D(\mathbf{k}) \varepsilon_{\mathbf{k}\lambda} = m\omega_{\mathbf{k}\lambda}^2 \varepsilon_{\mathbf{k}\lambda} \tag{6}$$

in which $\varepsilon_{\mathbf{k}\lambda}$ is the eigenvector ($\lambda = 1,2,3$) called polarization vector, and $m\omega_{\mathbf{k}\lambda}^2$ is the eigenvalue. The second-quantized form can be defined by the particle occupation number formalism:

$$\begin{aligned} \mathbf{u_k} &= \sum_{\lambda} \sqrt{\frac{\hbar}{2m\omega_{\mathbf{k}\lambda}}} \left(a_{\mathbf{k}\lambda}^+ + a_{\mathbf{k}\lambda} \right) \varepsilon_{\mathbf{k}\lambda} \\ \mathbf{p_k} &= i \sum_{\lambda} \sqrt{\frac{\hbar m\omega_{\mathbf{k}\lambda}}{2}} \left(a_{\mathbf{k}\lambda}^+ - a_{\mathbf{k}\lambda} \right) \varepsilon_{\mathbf{k}\lambda} \end{aligned} \tag{7}$$

where $a_{\mathbf{k}\lambda}$ and $a_{\mathbf{k}\lambda}^+$ annihilate and create a phonon at state $\mathbf{k}\lambda$, respectively, satisfying the following commutation relation:

$$\left[a_{\mathbf{k}\lambda}, a_{\mathbf{k}\lambda}^+ \right] = 1 \tag{8}$$

Now substituting Eqs. (4), (7) and (8) back to Eqs. (1) and (2), we finally obtain a second-quantized Hamiltonian for 3D phonons:

$$H = \sum_{\mathbf{k}\lambda} \hbar\omega_{\mathbf{k}\lambda} \left(a_{\mathbf{k}\lambda}^+ a_{\mathbf{k}\lambda} + \frac{1}{2} \right) \tag{9}$$

which now has a much simpler form and enhanced power to deal with complex interaction problems. Such formalism and power enhancement also happens to a classical dislocation upon second quantization.

The lesson from Classical Dislocation: Another major pillar that supports the quantized dislocation lies in a few aspects of its classical counterpart (Fig. 2). First, a dislocation exists in a crystalline solid, but not in amorphous materials (Fig. 2a). This greatly facilitates the electron and phonon interaction problems in that the Bloch's theorem is valid. Second, a dislocation not only can be defined in discrete
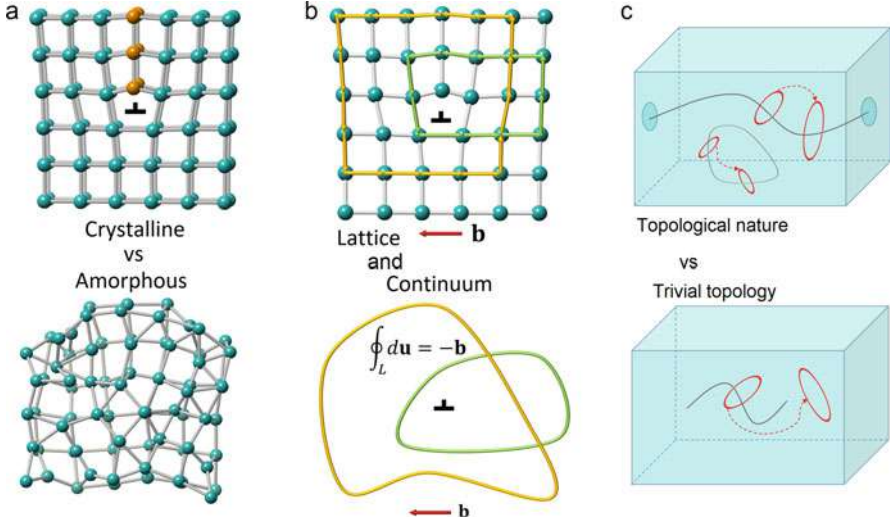
Fig. 2 A few essential elements for a classical dislocation, including crystalline, continuum compatible, and topological. (Figures from Li (2019))

crystals, but also in a continuous medium $\oint_L d\mathbf{u} = -\mathbf{b}$, where $\mathbf{u} = \mathbf{u}(\mathbf{R})$, and $\mathbf{R}$ is a continuous spatial coordinate (Fig. 2b). This facilitates a treatment to develop a simpler low-energy effective quantum theory. Third, a dislocation's definition $\oint_L d\mathbf{u} = -\mathbf{b}$ indicates a topological invariance and forbids a dislocation to end inside a crystal bulk (Fig. 2c). This facilitates the description of dislocation as a long straight line, without worrying a short or segmented dislocation line, if we leave the complex dislocation loop out of the picture.

With the method of canonical quantization procedure and the concept of topological invariance, the dislon theory can be considered as a natural merge of these two pillars. The detailed dislocation quantization procedure is introduced in a recent review (Li 2019). Briefly speaking, a dislocated lattice system still contains kinetic energy and lattice strain potential energy, just like Eqs. (1) and (2). The dislocation's definition $\oint_L d\mathbf{u} = -\mathbf{b}$ is evolved into a simple boundary condition of the quantized operator. The incorporation of both the kinetic energy and potential energy is a critical move. This is so since many dislocation models contain only the potential energy part (e.g., stress field) in a frozen-lattice configuration. However, for phonon-dislocation scattering, it is well known that the dominant mechanism is the dynamic process, which could be orders-of-magnitude higher than the strain field (Anderson and Malinowski 1972; Maurel et al. 2004). By simultaneously incorporating both the kinetic energy and the potential energy, not only the formalism is greatly simplified in second-quantized form, but also both dynamic and static interactions are taken into account on an equal footing, which is more appropriate for a realistic dislocation.

2.1   The Dislon Hamiltonian

The major difference between the phonon and dislocation quantization lies in the Fourier transform step. Since a dislocation is an extended line defect, we have another generic expansion instead of plane-wave expansion:

$$\mathbf{u}_{dis}\left(\mathbf{R}\right) = \frac{1}{L^2} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}} \mathbf{F}\left(\mathbf{k}\right) u_{\mathbf{k}} \tag{10}$$

where L denotes the system size, that there is one dislocation at present within a square of L^2. For acoustic phonons, we have $\mathbf{F}(\mathbf{k}) = \varepsilon_{\mathbf{k}}$. Moreover, under the long-wavelength limit $\mathbf{k} \to 0$, we have $\lim_{\mathbf{k}\to 0} u_{\mathbf{k}} = 0$ for phonons, i.e., it reduces to a perfect lattice under the static limit without displacement. The situation of dislocation is distinct. The $\mathbf{F}(\mathbf{k})$ is an expansion coefficient with local modes,

$$\mathbf{F}\left(\mathbf{k}\right) = \frac{1}{k_x k^2} \left(\mathbf{n}\left(\mathbf{b}\cdot\mathbf{k}\right) + \mathbf{b}\left(\mathbf{n}\cdot\mathbf{k}\right) - \frac{1}{(1-\nu)} \frac{\mathbf{k}\left(\mathbf{n}\cdot\mathbf{k}\right)\left(\mathbf{b}\cdot\mathbf{k}\right)}{k^2}\right) \tag{11}$$

if we assume xz plane as slip-plane, and ν is the Poisson ratio.

On the other hand, $u_{\mathbf{k}}$ satisfies a different boundary condition $\lim_{k_z \to 0} u_{\mathbf{k}} = 1$ if we assume the dislocation is along the z-direction (Li et al. 2018). This simple boundary condition is a natural result to ensure the compatibility with the definition $\oint_L d\mathbf{u} = -\mathbf{b}$. In the end, the dislon Hamiltonian with dislon excitation $\Omega_{\mathbf{k}}$ in the second quantized form can be written as

$$H_D = \sum_{\mathbf{k}\geq 0} \Omega_{\mathbf{k}} \left(d_{\mathbf{k}}^+ d_{\mathbf{k}} + \frac{1}{2}\right) + \sum_{\mathbf{k}\geq 0} \Omega_{\mathbf{k}} \left(f_{\mathbf{k}}^+ f_{\mathbf{k}} + \frac{1}{2}\right) \tag{12}$$

where the operators satisfy

$$\left[d_{\mathbf{k}}, d_{\mathbf{k}'}^+\right] = \delta_{\mathbf{k}\mathbf{k}'}, \quad \left[f_{\mathbf{k}}, f_{\mathbf{k}'}^+\right] = \delta_{\mathbf{k}\mathbf{k}'} \tag{13}$$

with a boundary condition

$$\lim_{k_z \to 0} d_{\mathbf{k}} = \lim_{k_z \to 0} d_{\mathbf{k}}^+ = C_{k_\parallel} \tag{14}$$

in which $C_{k_\parallel}$ is a boundary term taking care of the effects along the in-plane directions perpendicular to the dislocation line, such as Coulomb scattering and strain field scattering. Whenever a classical effect needs to be taken into account, we can generalize $C_{k_\parallel}$ for a specific problem.

To sum up, we need two Bosonic fields to describe a quantized dislocation, which are subject to different boundary conditions, in contrast to the phonon case where one Bosonic field suffices (Fig. 3a). This resembles another topological defect of
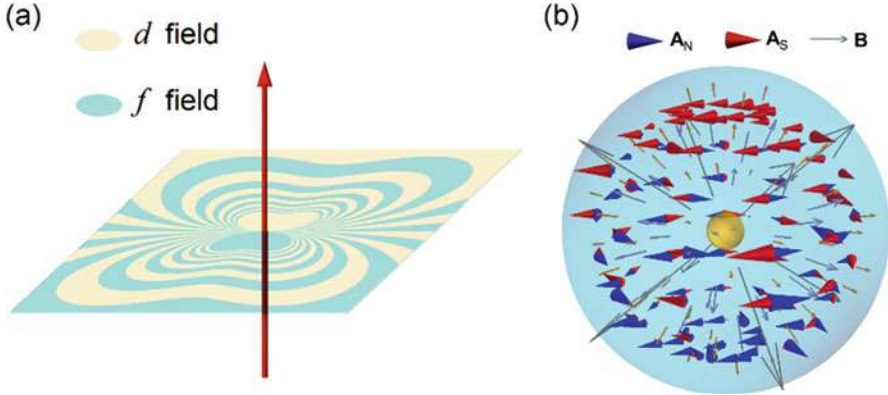
Fig. 3 (**a**) Two fields of the dislon. (**b**) A magnetic monopole with two classical vector fields. (Figure adapted Li (2019))

Fig. 4 The general workflow applying the dislon theory
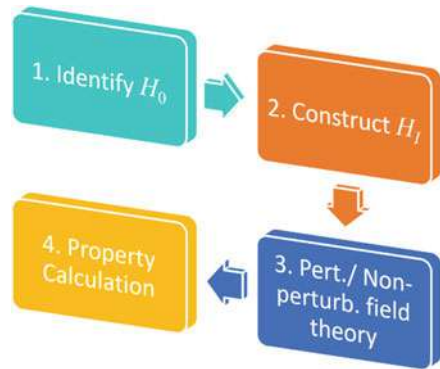
magnetic monopole (Fig. 3b), where two classical fields (magnetic vector potential), A_N and A_S, are needed to capture the intrinsic topology.

Equations (12), (13), and (14) are the central results of a dislon Hamiltonian. Equation (13) takes care of all dynamic effects, while Eq. (14) is responsible for the strain field scattering and the classical Coulomb scattering, where eventually $C_{k_\parallel}$ can be directly related to the classical dislocation-electron scattering amplitude.

2.2   General Workflow to Apply the Dislon Theory

With the dislon Hamiltonian in hand, we are in good shape to introduce the general workflow applying the dislon theory to a general functionality problem, summarized in Fig. 4.

Step 1. Identify the Hamiltonian H_0 that describes the dislocation-free system. If we are interested in electrons in a metal, then H_0 is just non-interacting electrons; if we want to study the phonon transport, H_0 is the free-phonon Hamiltonian Eq. (9). If

we want to study optical and magnetic properties that are influenced by dislocation, then a multi-band Hamiltonian can be applied, for instance, a 2-band model,

$$H_0 = \sum_{\mathbf{k}} \begin{pmatrix} c_{\mathbf{k}a}^+ & c_{\mathbf{k}b}^+ \end{pmatrix} \begin{pmatrix} E_{\mathbf{k}a} - \mu & 0 \\ 0 & E_{\mathbf{k}b} - \mu \end{pmatrix} \begin{pmatrix} c_{\mathbf{k}a} \\ c_{\mathbf{k}b} \end{pmatrix} \qquad (15)$$

where $(a, b) = (\uparrow, \downarrow)$ for a spin-1/2 system, while $(a, b) = (c, v)$ for a 2-band model with conduction and valence band.

Step 2. Identify the interacting Hamiltonian H_I. The non-interacting dislon Hamiltonian H_D itself is not enough to affect H_0, without having an interaction term. As we will see in the next section, in many situations, we need to write down the classical interaction Hamiltonian and then perform the corresponding quantization. Fortunately, Since dislocation contains strain field, dynamic vibration, and Coulomb interaction, the number of such interaction Hamiltonians are finite, some are quantized recently (Li et al. 2018). For instance, for the prototypical deformation potential scattering between dislocation displacement field and the gradient of Coulomb interaction, we have $\mathbf{u}_{dis} \cdot \nabla V_{Coulomb}$; for the velocity-velocity fluttering interaction with phonon, we have $\dot{\mathbf{u}}_{dis} \cdot \dot{\mathbf{u}}_{ph}$; while for the anharmonic dislocation-phonon interaction, we have $\mathbf{u}_{dis}\mathbf{u}_{ph}^2$, etc. The goal is to rewrite these classical interaction Hamiltonians in terms of a combination of creation and annihilation operators from H_0 (electron and phonon operators) and from H_D (dislon operators).

Step 3. Perturbative and Non-perturbative calculations. With the non-interacting and interacting Hamiltonians, we are ready to carry out relevant calculations, by setting up perturbative and non-perturbative calculations. By non-perturbative, we refer to the functional integral approach (Negele and Orland 1988), which is convenient to take into account the constraint Eq. (14). If, on the other hand, we only want to understand the dynamical quantum effect of dislocations, then no constraint is needed.

Step 4. Property calculation. Through step 3, almost all major physical quantities, such as DC and AC electrical conductivity, thermal conductivity, dielectric function, optical absorption, Seebeck coefficient, magnetic susceptibility, superconducting transition temperature, etc., can be computed systematically based on linear response theory, with controllable approximations. An example of DC conductivity workflow is shown in Fig. 5.

3   Cases Studies Using the Dislon Theory

This tribute intends to provide a few concrete examples and see how the dislon theory can be applied in a given interaction scenario. We will provide three examples, including the computation of the electron-dislon relaxation time, the calculation of superconducting transition temperature T_c, and also the phonon energy shift and relaxation time upon dislocation interaction. We elaborate the first example given its simplicity and outline the remaining two by summarizing the main results and major predictions.

Scheme for Computing Electrical Transport with Electron Scattering from Parallel Dislocation Arrays

DC Conductivity

$$\mathrm{Re}[\sigma_{ab}(T)] = -\frac{e^2}{L^3}\sum_{\mathbf{k}}\Gamma_{0,a}(\mathbf{k},\mathbf{k})\left.\frac{\partial n_F(\varepsilon)}{\partial\varepsilon}\right|_{E_{\mathbf{k}}-\mu}$$

$$\times\,\mathrm{Re}[\Gamma_b(\mathbf{k},\mathbf{k};E_{\mathbf{k}}-\mu+i\delta,E_{\mathbf{k}}-\mu-i\delta)]$$

Relaxation Rate

$$\frac{\hbar}{\tau_{\mathbf{k}}} = \frac{2\pi n_{\mathrm{dis}}}{L^2}\sum_{\mathbf{k}'}\left|\frac{A_{\mathbf{k}-\mathbf{k}'}}{\varepsilon(\mathbf{k}-\mathbf{k}')}\right|^2\delta_{kz,k'z}\delta(E_{\mathbf{k}}-E_{\mathbf{k}'})$$

Dislocation Potential

Current Vertex

$$\Gamma_b(\mathbf{k},\mathbf{k};\varepsilon+i\delta,\varepsilon\pm i\delta) = \Gamma_{0,b}(\mathbf{k},\mathbf{k})$$

$$+\frac{n_{\mathrm{dis}}}{L^2}\sum_{\mathbf{k}'}\left|\frac{A(\mathbf{k}'-\mathbf{k})}{\varepsilon(\mathbf{k}'-\mathbf{k})}\right|^2\delta_{kz,k'z}G(\mathbf{k}';\varepsilon+i\delta)$$

$$\times\,G(\mathbf{k}';\varepsilon\pm i\delta)\,\Gamma_b(\mathbf{k}',\mathbf{k}';\varepsilon+i\delta,\varepsilon\pm i\delta)$$

Self-Energy

$$\Sigma(\mathbf{k},\varepsilon\pm i\delta) = \frac{n_{\mathrm{dis}}}{L^2}\sum_{\mathbf{k}'}\frac{|A_{\mathbf{k}-\mathbf{k}'}/\varepsilon(\mathbf{k}-\mathbf{k}')|^2\delta_{kz,k'z}}{\varepsilon-E_{\mathbf{k}'}+\mu\pm i\delta}$$

Green's Function

$$G(\mathbf{k}';\varepsilon\pm i\delta) = \frac{1}{\varepsilon-(E_{\mathbf{k}'}-\mu)-\Sigma(\mathbf{k},\varepsilon\pm i\delta)\pm i\delta}$$
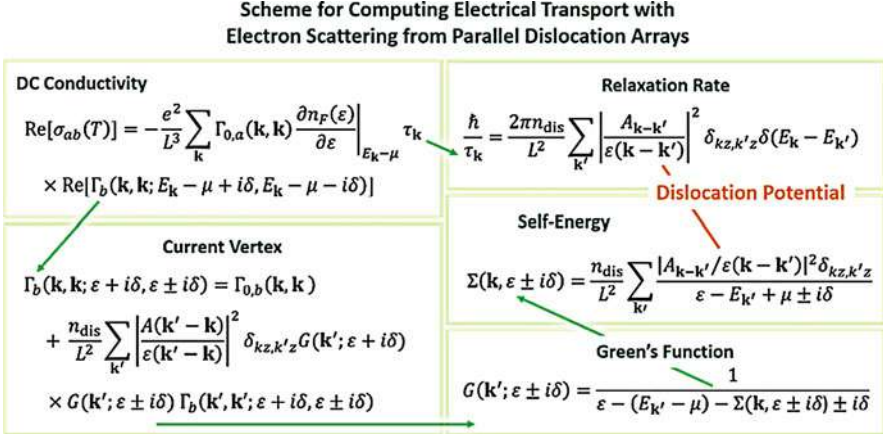
Fig. 5 The example workflow of electrical conductivity calculation using a quantum many-body approach with the presence of dislocations

3.1   Computation of Relaxation Time

In the classical dislocation theory, a dislocation is modeled as a scattering potential $V(\mathbf{r})$, and the relaxation rate from state i to f can be computed using Fermi's golden rule:

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar}\left|\int\psi_f^*(\mathbf{r})\,V(\mathbf{r})\,\psi_i(\mathbf{r})\,d\mathbf{r}\right|^2\delta\left(\varepsilon_f-\varepsilon_i\right) \qquad (16)$$

where $\psi_i(\mathbf{r})$ and $\psi_f(\mathbf{r})$ are the initial and final wavefunctions with energy ε_i and ε_f, respectively. If we assume a plane wave wavefunctions $\psi_i(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}/L^{D/2}$ and $\psi_f(\mathbf{r}) = e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}}/L^{D/2}$, and define a Fourier transform $V(\mathbf{q}) = \int V(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}d^D\mathbf{r}$, the relaxation rate for an electron with momentum $\mathbf{k}$ being scattered into $\mathbf{k}+\mathbf{q}$ using Eq. (16) can be written as

$$\Gamma_{\mathbf{k}\to\mathbf{k}+\mathbf{q}} = \frac{1}{L^{2D}}\frac{2\pi}{\hbar}|V(\mathbf{q})|^2\delta\left(\varepsilon(\mathbf{k}+\mathbf{q})-\varepsilon(\mathbf{k})\right) \qquad (17)$$

Since the momentum change $\mathbf{q}$ is arbitrary, if we assume a total number of N_{dis} dislocations, the total relaxation rate as a function of electron momentum $\mathbf{k}$ can be written as

$$\Gamma_{\mathbf{k}} = N_{dis}\sum_{\mathbf{q}}\Gamma_{\mathbf{k}\to\mathbf{k}+\mathbf{q}} = \frac{2\pi}{\hbar}n_{dis}\int\frac{d^D\mathbf{q}}{(2\pi)^D}|V(\mathbf{q})|^2\delta\left(\varepsilon(\mathbf{k}+\mathbf{q})-\varepsilon(\mathbf{k})\right) \qquad (18)$$

where n_{dis} denotes the dislocation density.

Equation (18) is a common approach to link dislocation's potential $V(\mathbf{r})$, electron energy $\varepsilon(\mathbf{k})$ to the relaxation rate $\Gamma_{\mathbf{k}}$. However, problems remain: (a) The relation between $V(\mathbf{q})$ and the definition $\oint_L d\mathbf{u} = -\mathbf{b}$ is obscure. If we model a dislocation as a line charge, where $V(\mathbf{r})$ is Coulomb potential, it has nothing to do with dislocation's definition. (b) When the electrons are under interactions in the solid, the wavefunctions are usually too difficult to obtain. In addition, it seems impossible to incorporate other interactions into this formalism. (c) When dislocation density is high and multiple scattering emerges, it would be nice to have a formalism to treat multiple scattering, especially infinite order where a qualitative change of system such as Anderson localization, emerges (Rammer 1998). (d) Regardless of interaction and high-order effects, many other effects, such as temperature dependence, are also challenging to be incorporated. As we will see, the quantum field approach can solve all these problems instantly in an elegant way.

In a quantum field language, instead of using wavefunctions, the Green's functions, aka propagators, serve as building blocks. We will work with the imaginary-time Green's function

$$G\left(\mathbf{k}, \tau\right) = -\left\langle c_{\mathbf{k}}\left(\tau\right) c_{\mathbf{k}}^{+}(0)\right\rangle \tag{19}$$

which demonstrates a process that an electron with momentum $\mathbf{k}$ is created at time 0 ($c_{\mathbf{k}}^{+}(0)$ term), and then annihilated at time τ ($c_{\mathbf{k}}(\tau)$ term), hence represents the electron propagation probability amplitude under an interacting environment. τ is called imaginary time, $0 < \tau < \beta = 1/k_B T$, whose Fourier transform gives the so-called Matsubara frequency $p_n, p_n = (2n + 1)\pi/\beta, n = 0, \pm 1, \pm 2 \ldots$ for fermions. Such formalism is a convenient approach to solve finite-temperature problems.

For non-interacting spinless electrons with $H_0 = \sum_{\mathbf{k}} (E_{\mathbf{k}} - \mu) c_{\mathbf{k}}^{+} c_{\mathbf{k}}$, the non-interacting Green's function in Matsubara frequency domain is written as

$$G_0\left(\mathbf{k}, ip_n\right) = \frac{1}{ip_n - E_{\mathbf{k}} + \mu} \tag{20}$$

from which the most straightforward process that leads to relaxation time $\tau_{\mathbf{k}}$ can be considered as a virtual process where an electron $\mathbf{k}$ is first scattered into $\mathbf{k} + \mathbf{q}$ by a dislon with momentum $\mathbf{q}$ with interaction strength $g_{\mathbf{q}}$, then the $\mathbf{k} + \mathbf{q}$ electron is scattered back to the original momentum $\mathbf{k}$ through a dislon with momentum $-\mathbf{q}$. Such a process can be diagrammatically represented as:
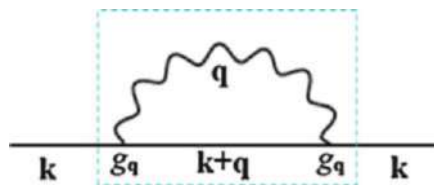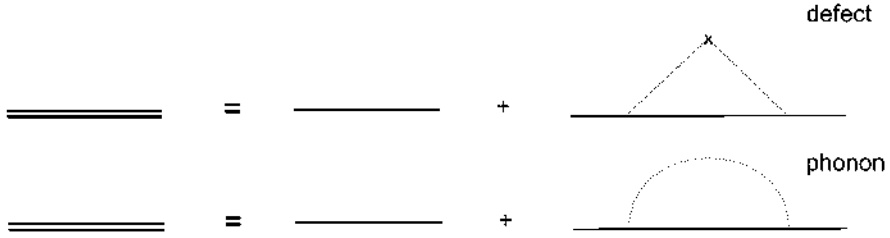
In above figure, the straight line segment and wavy line denotes the electron and dislon propagator, G_0 and D_0, respectively, in which the part inside the dashed rectangle is the lowest-order self-energy, which can be written as

$$\Sigma^1 \left(\mathbf{k}, i p_n \right) = -\frac{1}{\beta} \sum_{\mathbf{q} m} g_{\mathbf{q}}^2 D^0 \left(\mathbf{q}, i \omega_m \right) G^0 \left(\mathbf{k} - \mathbf{q}, i p_n - i \omega_m \right) \tag{21}$$

where the internal momentum $\mathbf{q}$ and Matsubara frequency $i\omega_m$ are summed over. Finally, the relaxation time $\tau_{\mathbf{k}}$ can be computed from a self-energy calculation:

$$\frac{1}{\tau_{\mathbf{k}}} = -2 \operatorname{Im} \left[\Sigma \left(\mathbf{k}, 0 \right) \right] \tag{22}$$

which can lead to consistent result with Eq. (18). Despite seemingly more cumbersome than a Fermi's golden rule approach, the above quantum field scheme can indeed conquer all difficulties faced by Fermi's golden rule approach.

(a) Dislocation's definition is well respected in the dislon Hamiltonian Eqs. (12), (13) and (14).
(b) If other interaction mechanisms are needed, no matter point defects or phonon interactions with electrons, or even Coulomb interaction between electrons, we only need to rewrite the corresponding electron propagator taking into account the relevant interactions. For instance, the lowest-order correction of point defect or phonon scattering can be diagrammatically represented as:
where the double-line on the left-hand side denotes the dressed electron propagator upon interaction. The total self-energy taking into account other interactions can be represented as:
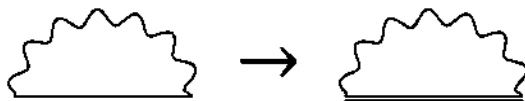
while the Eq. (22) is still valid by replacing the electron propagator $G_0 \to G$ in the updated self-energy.

(c) Higher-order scattering processes are easily incorporable as well. For instance, the second-order electron-dislon scattering can be written pictorially as:
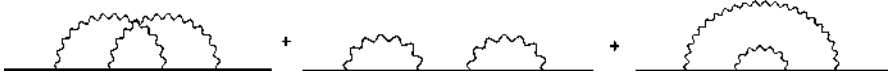
(d) Other effects can also be incorporated in a systematic way. In particular, temperature effects are fully taken into account within the finite-temperature Matsubara formalism.

3.2   Electron-Dislon Interaction: Explaining T_c

The power of adopting the dislon approach to study the electron-dislon interaction problem is beyond the calculation of relaxation time, but can also be used to study the effects of dislocations on dislocated superconductors. To see this, we notice that the electron-dislocation interaction Hamiltonian can well be described by a deformation potential scattering Hamiltonian:

$$H_{e-ion} = \int d^3\mathbf{R}\rho_e\left(\mathbf{R}\right) \sum_{n=1}^{N} \nabla_\mathbf{R} V_{ei}\left(\mathbf{R} - \mathbf{R}_n^0\right) \cdot \mathbf{u}_{dis}\left(\mathbf{R}_n^0\right) \tag{23}$$

where $\rho_e(\mathbf{R})$ is the electron charge density, V_{ei} is the Coulomb interaction, $\mathbf{R}_n^0$ is the atomic location of atom number n. After quantization, we have

$$H_{e-dis} = \sum_{\mathbf{k}'\sigma;\mathbf{k}\geq 0} g_\mathbf{k} c_{\mathbf{k}'+\mathbf{k}\sigma}^+ c_{\mathbf{k}'\sigma} d_\mathbf{k} + \text{h.c.} \tag{24}$$

where $g_\mathbf{k} \propto \mathbf{k} \cdot \mathbf{F}\left(\mathbf{k}\right) \propto \frac{1-2\nu}{1-\nu} b$ is the electron-dislocation coupling strength, which gives highly consistent result from semiclassical theory (Dexter and Seitz 1952).

In the end, we were able to compute the superconducting transition temperature T_c, quantitatively in a dislocated superconductor, for the first time. The T_c equation can be written as (Li et al. 2017c)

$$\frac{1}{g_{ph} + g_D} = N\left(\mu\right) \int_0^{\theta_D'} d\xi \sum_{s=\pm 1} \tanh\left(\frac{\xi + is\Gamma_D}{2T_c}\right) \bigg/ 2\xi \tag{25}$$

with two major coefficients (classical potential scattering Γ_D and quantum fluctuation g_D):
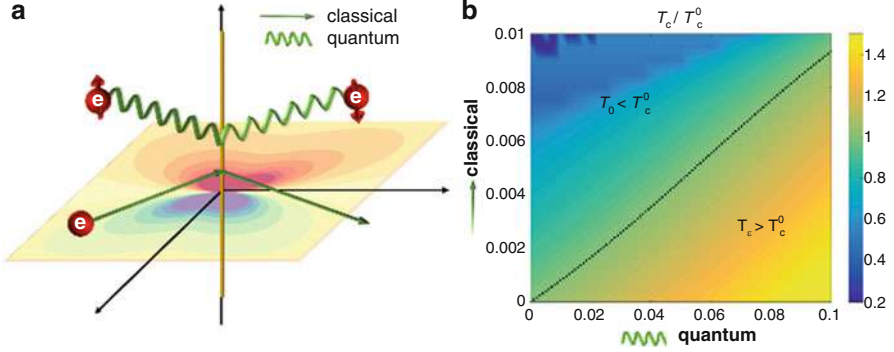
Fig. 6 (**a**) Schematics of the classical vs quantum type of electron-dislocation interaction, whose competition will determine the T_c in a dislocated superconductor (**b**). (Figure adapted from Li et al. (2017c))

$$\Gamma_D = \frac{\pi m^*}{4\hbar^2 k_F^2 k_{TF}^4} \left(Ze^2 n\right)^2 n_{dis} b^2 \left(\frac{1-2\nu}{1-\nu}\right)^2$$
$$g_D = n_{dis}^{3/2} \frac{\left(4\pi Ze^2 n\right)^2}{2k_{TF}^4 (\lambda + 2\mu)} \tag{26}$$

where g_{ph} is the electron-phonon coupling constant, θ'_D is the renormalized Debye frequency, m^* is the effective mass, k_F is Fermi wavevector, k_{TF} is Thomas-Fermi screening wavevector, n is atomic number density, n_{dis} is dislocation density, b is Burgers vector, ν is Poisson ratio, λ and μ are Lame parameters, which are all common materials parameters that can be looked up in database. The Eq. (25) implies a completion effect between classical scattering and quantum fluctuation, as illustrated in Fig. 6, which shows excellent quantitative agreement comparing with experimental data without free fitting parameter (Li et al. 2017c).

3.3   Phonon-Dislon Interaction: Beyond Perturbation

The general workflow is also applicable to the dislocation-phonon interaction problems. The classical velocity-velocity drag-like fluttering interaction Hamiltonian between dislocation and phonon can be written as

$$H_{flu} = \rho \int \dot{\mathbf{u}}_{ph}\left(\mathbf{R}\right) \cdot \dot{\mathbf{u}}_{dis}\left(\mathbf{R}\right) d^3\mathbf{R} \tag{27}$$

After the second quantization procedure, we have

$$H_{flu} = \sum_{\mathbf{k}>0} g_{\mathbf{k}}^{\text{dyn}} \boldsymbol{\varepsilon}_{\mathbf{k}} \cdot \mathbf{F}\left(\mathbf{k}\right) \left(-a_{\mathbf{k}} + a_{-\mathbf{k}}^+\right) f_{\mathbf{k}} + \text{h.c.} \tag{28}$$

where $g_{\mathbf{k}}^{\mathrm{dyn}}$ is the dislocation-phonon coupling dynamic coefficient that depends on phonon and dislon dispersions $\omega_{\mathbf{k}}$ and $\Omega_{\mathbf{k}}$. The interaction Hamiltonian Eq. (28) turns out to be able to shift phonon dispersion and meanwhile result in a finite phonon lifetime (Li et al. 2017b).

The phonon dispersion changed by dislocation is shown in Fig. 7. In a dislocated crystal (Fig. 7a), the dislon theory predicts anisotropic phonon softening and TA modes splitting (Fig. 7b, yellow lines), which are confirmed by both lattice dynamics simulations (Fig. 7b, blue lines) and ab initio first-principles calculations (Fig. 7c). As to the finite phonon lifetime, the dislon theory reveals a fruitful structure, depending on the dislocation type, phonon type, and phonon polarization.

In particular, a special type of saturation and resonance are predicted (Fig. 8a, b). These go far beyond any classical theory, where the relaxation rate τ^{-1} only contains monotonic dependence with phonon frequency ω:

$$
\begin{aligned}
\tau_{\mathrm{fluttering}}^{-1} &\propto \omega^{-1} \\
\tau_{\mathrm{strain}}^{-1} &\propto \omega \\
\tau_{\mathrm{core}}^{-1} &\propto \omega^3 \\
\tau_{\mathrm{isotope}}^{-1} &\propto \omega^4
\end{aligned}
\tag{29}
$$
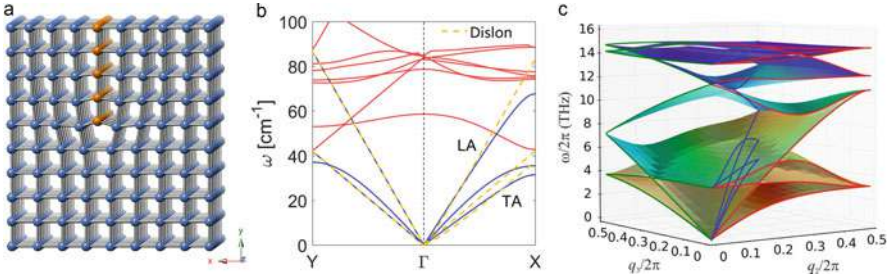
Fig. 7 (**a**) A schematic of a dislocated simple cubic crystal. (**b**) An isotropic phonon softening and transverse acoustic (TA) modes splitting with dislocations. (**c**) Verification through *ab initio* calculations. (Figures adapted from Li et al. (2017b) and Wang et al. (2017))
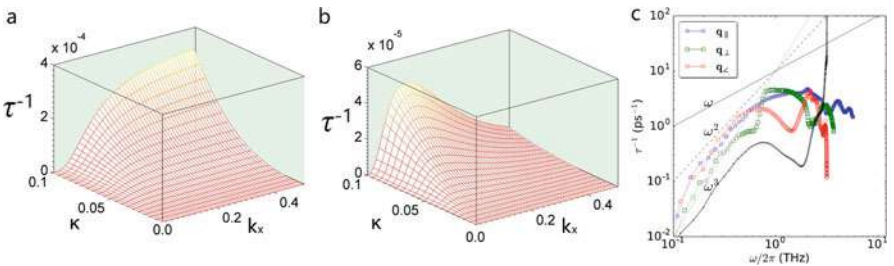
Fig. 8 The dislocation-phonon scattering relaxation rates for a transverse-like (**a**) and a longitudinal-like (**b**) with an edge dislocation. A similar saturation behavior in (**a**) and a resonance behavior in (**b**) are seen in the ab initio calculations in (**c**). (Figures adapted from Li et al. (2017b) and Wang et al. (2017))

where the four terms denote phonon-dislocation dynamic "fluttering" scattering, phonon-dislocation strain-field scattering, phonon-dislocation core scattering, and phonon-isotope point defect scattering, respectively. Both the saturation behavior and the resonance behavior are revealed in the ab initio calculations (Fig. 8c), although the simulation and theory are not precisely computing the same system. The simulation contains a dislocation pair to reduce the system size, while the dislon theory is computed in a simpler system with single dislocation and thus has a slightly different definition on the phonon modes.

4   Outlook and Perspective

Despite some initial success of the dislon theory, as a framework on which new theories can be built upon, the dislon theory is still at its infant stage. Here we sketch a few more examples that appear challenging for a classical dislocation theory, but are expected to be directly computable using the dislon theory:

Electronic bandgap: how dislocations can change the bandgap in semiconductors and insulators. We notice that the multi-band Hamiltonian Eq. (15) contains an energy gap. By introducing an interaction Hamiltonian, such as Eq. (24), the bandstructures will change, $E_{\mathbf{k}a} \rightarrow E'_{\mathbf{k}a}$, which can be obtained from perturbation theory, and will result in a bandgap modulation effect.

Optical absorption: the role of dislocations to the optical absorption coefficient $\alpha(\omega)$. Optical absorption leads to intensity distinction $I(x) = I_0 \exp(-\alpha x)$. To compute $\alpha(\omega)$, we first follow the scheme like Fig. 5 to compute the AC conductivity $\sigma(\omega)$, using the original Hamiltonian H_0 and total Hamiltonian $H_0 + H_I + H_D$. Then, the dielectric function $\varepsilon(\omega)$ can be written as $\varepsilon(\omega) = 1 + 4\pi i\sigma(\omega)/\omega$. Finally, defining an extinction coefficient $k(\omega)$ as $k^2(\omega) \equiv (|\varepsilon(\omega)| - \mathrm{Re}\,[\varepsilon(\omega)])/2$, the optical absorption coefficient can be written as $\alpha(\omega) = 2\omega k(\omega)/c$.

Phase transition: Although challenging to summarize within a few steps, one major application that distinguishes the dislon theory from a classical dislocation model lies in the capability to study dislocation-induced phase transition problems, such as metal-insulator transition. Early experiments reported the dislocation-induced semiconductor-superconductor phase transition (Fogel et al. 2001), which has been explained using the dislon theory, quantitatively (Li et al. 2017c). In a more recent example, a quantum phase transition of phonons is revealed (Pablo-Pedro et al. 2018), caused by a competition effect between the topological protection of dislocation and a topological-breaking inelastic scattering. By understanding the relationship between dislocations and phase transitions, dislocations will gain more importance from materials imperfections to a new dimension to tailor the phase diagram, including unexplored new electronic, phononic, and photonic phases.

Indeed, the dislon theory offers a systematic approach to understand the interplay between crystal dislocations and materials electronic, spintronic, phononic, and photonic degrees of freedom at a microscopic quantum level. However, why the dislon, or say the quantization procedure, could work in the first place? Intuitively, the term "quantum field" contains both the "quantum" part, which deals with the internal dynamics and excitation, and the "field" part, which deals with spatial

extension. This dual nature seems appropriate to describe a dislocation, which contains both internal dynamics, such as dynamic fluttering, and spatial extension that arises naturally as an extended defect. In fact, the spatial extension naturally and unavoidably leads to the internal dynamic structure. This brings up a more general question, that whether any extended defect can, or even should, be described by some quantum field:

$$\text{Large Defects} = \text{Quantum Fields} \tag{30}$$

The answer to this grand question may greatly empower our approach to tackle complex defects problems.

References

Anderson AC, Malinowski ME (1972) Interaction between thermal phonons and dislocations in LiF. Phys Rev B 5:3199–3210. https://doi.org/10.1103/PhysRevB.5.3199

Anderson PW (1959) Theory of dirty superconductors. J Phys Chem Solids 11:26–30. https://doi.org/10.1016/0022-3697(59)90036-8

Baroni S, de Gironcoli S, Dal Corso A, Giannozzi P (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73:515–562. https://doi.org/10.1103/RevModPhys.73.515

Brown RA (1977) Electrical-resistivity of dislocations in metals. J Phys F Met Phys 7:1283–1295

Dexter DL, Seitz F (1952) Effects of dislocations on mobilities in semiconductors. Phys Rev 86:964–965. https://doi.org/10.1103/PhysRev.86.964

Drozdov NA, Patrin AA, Tkachev VD (1976) Recombination radiation on dislocations in silicon. ZhETF Pisma Redaktsiiu 23:651

Fetter AL, Walecka JD (2003) Quantum theory of many-particle systems. Dover Publications, Mineola

Fogel NY, Pokhila AS, Bomze YV, Sipatov AY, Fedorenko AI, Shekhter RI (2001) Novel superconducting semiconducting superlattices: dislocation-induced superconductivity? Phys Rev Lett 86:512–515. https://doi.org/10.1103/PhysRevLett.86.512

Fu C, Li M (2017) Oscillatory deviations from Matthiessen's rule due to interacting dislocations. J Phys-Condens Matter 29:325702

Hirth JP, Lothe J (1992) Theory of dislocations, 2nd edn. Krieger, Malabar

Jena D, Gossard AC, Mishra UK (2000) Dislocation scattering in a two-dimensional electron gas. Appl Phys Lett 76:1707. https://doi.org/10.1063/1.126143

Karolik AS, Luhvich AA (1994) Calculation of electrical-resistivity produced by dislocations and grain-boundaries in metals. J Phys-Condens Matter 6:873–886

Kveder V, Badylevich M, Steinman E, Izotov A, Seibt M, Schroter W (2004) Room-temperature silicon light-emitting diodes based on dislocation luminescence. Appl Phys Lett 84:2106–2108. https://doi.org/10.1063/1.1689402

Kveder V, Badylevich M, Schröter W, Seibt M, Steinman E, Izotov A (2005) Silicon light-emitting diodes based on dislocation-related luminescence. Phys Status Solidi A 202(5):901–910. https://doi.org/10.1002/pssa.200460512

Li M (2019) Quantized dislocations. J Phys Condens Matter 31:083001. https://iopscience.iop.org/article/10.1088/1361-648X/aaf6e1

Li M, Cui W, Dresselhaus MS, Chen G (2017a) Electron energy can oscillate near a crystal dislocation. New J Phys 19:013033

Li M et al (2017b) The non-perturbative quantum nature of the dislocation-phonon interaction. Nano Lett 17:1587. https://doi.org/10.1021/acs.nanolett.6b04756

Li M, Song Q, Liu T-H, Meroueh L, Mahan GD, Dresselhaus MS, Chen G (2017c) Tailoring superconductivity with quantum dislocations. Nano Lett 17:4604. https://doi.org/10.1021/acs.nanolett.7b00977

Li M, Tsurimaki Y, Meng Q, Andrejevic N, Zhu Y, Mahan GD, Chen G (2018) Theory of electron–phonon–dislon interacting system—toward a quantized theory of dislocations. New J Phys 20:023010. https://doi.org/10.1088/1367-2630/aaa383

Look DC, Sizelove JR (1999) Dislocation scattering in GaN. Phys Rev Lett 82:1237–1240

Mahan GD (2000) Many-particle physics. Physics of solids and liquids, 3rd edn. Kluwer Academic/Plenum Publishers, New York

Markowitz D, Kadanoff LP (1963) Effect of impurities upon critical temperature of anisotropic superconductors. Phys Rev 131:563–575. https://doi.org/10.1103/PhysRev.131.563

Maurel A, Mercier J-F, Lund F (2004) Elastic wave propagation through a random array of dislocations. Phys Rev B 70. https://doi.org/10.1103/PhysRevB.70.024303

Nabarro FRN (1967) Theory of crystal dislocations, International series of monographs on physics. Clarendon P, Oxford

Negele JW, Orland H (1988) Quantum many-particle systems. Frontiers in physics, vol 68. Addison-Wesley, Redwood City

Pablo-Pedro R et al (2018) Phonon quantum phase transition. ArXiv e-prints:1809.06495

Rammer J (1998) Quantum transport theory. Frontiers in physics, vol 99. Perseus Books, Reading

Reiche M, Kittler M (2016) Electronic and optical properties of dislocations in silicon. Crystals 6:74 ARTN 74. https://doi.org/10.3390/cryst6070074

Roth EP, Anderson AC (1979) Interaction between thermal phonons and dislocations in LiF. Phys Rev B 20:768. https://doi.org/10.1103/PhysRevB.20.768

Wang Z, Saito M, McKenna KP, Ikuhara Y (2014) Polymorphism of dislocation core structures at the atomic scale. Nat Commun 5:3239. https://doi.org/10.1038/ncomms4239

Wang T, Carrete J, van Roekeghem A, Mingo N, Madsen GKH (2017) *Ab initio* phonon scattering by dislocations. Phys Rev B 95:245304. https://doi.org/10.1103/PhysRevB.95.245304

Weimann NG, Eastman LF, Doppalapudi D, Ng HM, Moustakas TD (1998) Scattering of electrons at threading dislocations in GaN. J Appl Phys 83:3656–3659

Part III
Photovoltaics: First Principles Modeling

Photovoltaics: Advances in First Principles Modeling – Overview

13

Marco Bernardi and Jeffrey C. Grossman

Contents

Abstract

Photovoltaic devices are challenging to model due to their multi-scale physics, which ranges from light absorption to charge carrier generation and transport. This overview chapter outlines recent developments in first-principles modeling of solar cells. Besides the traditional focus on predicting the structure, bandgap, and optical properties of materials, first-principles methods can now also model charge transport and excited carrier dynamics. New materials are breaking the established paradigms of conventional solar cells, leading to exciting discoveries of broad relevance to materials science and physics. We discuss three recent directions in photovoltaics research, including ultrathin solar cells, complex

M. Bernardi (✉)
Department of Applied Physics and Materials Science, California Institute of Technology, Pasadena, CA, USA
e-mail: bmarco@caltech.edu

J. C. Grossman
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: jcg@mit.edu

287

materials, and ultrafast electron dynamics, and outline future challenges. First-principles calculations are playing a key role in pushing these frontiers of photovoltaics research.

1   Modeling Solar Cells

Photovoltaic (PV) solar cells convert sunlight to electricity through a cascade of physical processes spanning multiple length and timescales (Bernardi and Grossman 2016). In a conventional solar cell, light absorption generates electron and hole carriers, which are then transported through the active layer and extracted at the contacts. In excitonic solar cells, such as those made of organic or nanostructured materials, sunlight generates excitons (bound electron-hole pairs), so that additional exciton transport and dissociation steps are needed for energy conversion. Modeling PV devices entails understanding the energetics and dynamics of the charge carriers, excitons, and photons involved in the energy conversion process.

We recently reviewed the computational methods available for modeling solar cell materials and devices (Bernardi and Grossman 2016). Our chapter identified several open challenges, including modeling from first-principles charge transport, ultrafast electron dynamics, and the impact of defects and impurities on solar cell performance. New trends and computational tools have emerged in the last three years since our review was published. Some of the most notable recent research directions, which are outlined in the next section, can be summarized with three words: *ultrathin, ultrafast, complex*. These ideas pose new challenges to first-principles calculations. New complex PV materials and ultrafast electron dynamics are introduced in Sect. 2 and reviewed in the contributed chapters (see below). Research efforts on atomically thin solar cells are summarized in Sect. 3.

2   New Frontiers: Ultrathin, Ultrafast, Complex

Industry and academia approach solar cell research from complementary angles. In academia, PV is a playground for materials synthesis and for solid state and device physics. First-principles calculations play a key role in understanding energy conversion in new PV materials discovered in academia. By contrast, industry efforts are more practical – they are centered on maximizing device efficiency while reducing module and installation cost per Watt. Conventional semiconductors – mainly silicon, but also GaAs, chalcogenides, and amorphous Si – are still the main focus of industry research, though new materials are also being explored. Equipping solar cells with new functionalities, such as making them bendable and transparent, is also a recent research direction toward making solar cells ubiquitous. We explore recent trends in academic PV research, focusing on three themes – complex materials, ultrafast timescales, and ultrathin solar cells.

Complex materials   Despite the dominant role of conventional semiconductors, there is a renewed focus on structurally and chemically complex materials, which

can enable new engineering paradigms not possible with simple elemental or binary inorganic semiconductors (Jean et al. 2015). Examples are hybrid halide perovskites, which have been a primary focus of recent PV research (Liu et al. 2013; Green et al. 2014). Halide perovskites possess several desirable properties, including easy deposition from solution or vapor, a near-ideal bandgap that can be tuned with composition, and strong sunlight absorption in thin films. Despite the dramatic rise in the power conversion efficiency of halide perovskites, the origin of their exceptional performance and charge transport properties remains the subject of debate, largely due to their complexity. The atomic and electronic structures of halide perovskites are uniquely complex compared to conventional solar cell materials. Their electronic, vibrational, and electron spin degrees of freedom are strongly coupled, and their structure exhibits phase transitions controlled by an interplay between the inorganic framework and the organic ion rotations.

Long carrier lifetimes and diffusion lengths have been recognized early on as a key factor for device performance in halide perovskites; understanding this unusual charge transport constitutes a main challenge for first-principles calculations. Methylammonium lead iodide ($MAPbI_3$) is a key halide perovskite compound. It has been studied extensively, but the mechanisms governing its long carrier lifetimes are not completely understood. A leading hypothesis is that a large bulk Rashba effect (due to spin-orbit coupling) makes the bandgap indirect, thus slowing down carrier recombination. One of us (Bernardi) recently demonstrated that at room temperature, $MAPbI_3$ is centrosymmetric and has only a very small Rashba splitting, thus ruling out the role of bulk Rashba effects (Frohna et al. 2018). This work also established the correct structure and band structure for future ab initio studies of charge transport in $MAPbI_3$. There are important open challenges for theory in halide perovskites besides charge transport; they are outlined in the contributed chapter by Giustino and in a recent review article (Martiradonna 2018).

Ultrafast timescales Ultrafast spectroscopies are able to probe charge carrier dynamics at the femtosecond to nanosecond timescale, over which key PV physical processes occur. These experiments are providing new insights into the microscopic energy conversion mechanisms in solar cells. Among other benefits, understanding the sub-picosecond dynamics of carriers and excitons can lead to more efficient exciton dissociation and to extracting hot carriers. Time-dependent density functional theory (TD-DFT) has so far been the primary method for studying excited electron dynamics. Efforts on understanding charge transfer in model donor-acceptor interfaces using TD-DFT are reviewed in the contributed chapter by Rozzi and Pittalis.

There are also new rigorous methods to compute excited electron dynamics from first principles. One of us (Bernardi) combined electron-phonon scattering calculations (Bernardi 2016) with the real-time Boltzmann transport equation to investigate excited electron dynamics and hot carriers effects in a range of semiconductors and metals (Bernardi et al. 2014, 2015a, b; Jhalani et al. 2017). The time and length scale for hot carrier relaxation and the energy loss mechanisms of the excited electrons can be accurately predicted with these techniques, thus providing valuable information on a regime (sub-picosecond time and nanometer

length scales) that is hard to access experimentally. Using similar techniques, charge transport (e.g., the electron mobility as a function of temperature) can now be accurately computed in materials ranging from simple inorganic to complex organic semiconductors and transition metal oxides (Mustafa et al. 2016; Zhou and Bernardi 2016; Lee et al. 2018; Zhou et al. 2018). These new first-principles methods for investigating carrier dynamics complement the traditional ab initio toolkit employed for studying solar cells – mainly, DFT and GW for bandgaps and interface band alignments and TD-DFT and the Bethe-Salpeter equation (BSE) for computing optical absorption and excitons (Bernardi and Grossman 2016).

Ultrathin solar cells   The last trend we mention here is the recent focus on novel two-dimensional (2D) materials with potential to enable ultrathin (∼10 nm thick) solar cells. The rise of graphene has spurred investigations of 2D materials, which have led to the discovery of a wide range of atomically thin semiconductors. These materials and their potential for solar cells are discussed in the next section as a selected case study from our own research.

3   Ultrathin Solar Cells

Recently discovered atomically thin semiconductors possess yet untapped technological potential. Much research has focused on group-6 transition metal dichalcogenides (TMDs) with a chemical formula of MX_2 (where M is a group-6 transition metal and X a chalcogen atom). This family of semiconductors, which includes MoS_2, WS_2, and related compounds, can be prepared routinely in monolayer and few-layer form and exhibits unique electronic and optical properties (Bernardi et al. 2017). Solar cells based on 2D-TMDs provide an example of how theory can guide the design of new PV devices.

In an article we published in 2013 (Bernardi et al. 2013), we predicted that 2D-TMDs possess an extraordinary sunlight absorption – they absorb as much sunlight in a sub-nm thick monolayer as GaAs (one of the best solar cell materials) can absorb in 10 nm. These ultrathin light absorbers show promise for applications in solar cells, as is seen by analyzing PV devices based on just two stacked monolayers. We investigated a Schottky barrier solar cell between MoS_2 and graphene and an excitonic solar cell based on a MoS_2/WS_2 hetero-bilayer (Bernardi et al. 2013). Our calculations demonstrated that such 1 nm thick active layers can attain power conversion efficiencies of up to ∼1%, which corresponds to power densities (Watts/kg) higher by a factor of 100–1,000 than the best existing thin-film solar cells based on conventional semiconductors (Bernardi et al. 2013).

Several results predicted in our work have now been confirmed experimentally, including the type-II alignment and the presence of interlayer excitons at the MoS_2/WS_2 hetero-bilayer interface. Solar cells consisting of a MoS_2/WS_2 bilayer with close to 1% efficiency have been fabricated (Furchi et al. 2014; Pospischil et al. 2014), confirming our calculations. Note that these nanometer thick devices are perhaps the thinnest solar cells ever made. While the 1% power conversion

efficiency of a TMD bilayer is far too small for PV applications, our work predicted that light absorption and the power conversion efficiency could be increased by an order of magnitude by stacking multiple TMD monolayers and managing light through a back metallic contact (Bernardi et al. 2013). Inspired by our idea, researchers have recently fabricated ~15 nm thick multi-layer stacks of TMDs exhibiting near-unity absorption (Jariwala et al. 2016), in agreement with our predictions.

These results demonstrate the potential of first-principles calculations for designing novel PV materials and devices. The future of atomically thin solar cells is still unclear, mainly because fabricating large-area and high-quality TMDs remains challenging. Yet, first-principles calculations on 2D-TMDs are making significant strides. An example from our research is the accurate prediction of radiative lifetimes in monolayer TMDs (Palummo et al. 2015) and the development of new methods to investigate their anisotropic light emission (Chen et al. 2018). These recent advances are an example of how methods developed to investigate PV materials are also broadly relevant to the materials theory community.

4   Outlook and Perspective

The textbook model of a solar cell is a semiconductor *pn* junction under illumination. This simplified picture is not representative of devices investigated in modern PV research, where *interfaces, defects, impurities, contacts*, and *degradation* play a key role. First-principles calculations able to capture such complexity of real devices are needed for understanding solar cells and improving their efficiency.

Heterojunctions and interfaces are present in nearly all inorganic and organic solar cells. For organic solar cells, the optimization of transport and buffer layers can turn a poor device into a record one, and designing donor/acceptor interfaces is essential to optimizing the efficiency. Defects, impurities, and grain boundaries are particularly important in inorganic solar cells. They are challenging to model from first principles both because they require simulation cells with a large number of atoms and because new theory is needed.

For example, the impact of defects and impurities on charge transport cannot be accurately computed from first principles. Electron-defect scattering and carrier trapping at impurities have received little attention from the first-principles community despite being crucial issues in devices. New methods are emerging to rigorously treat electron-defect scattering; these efforts parallel the recent advances on electron-phonon interactions and will be highly valuable for PV devices. Predicting the stability of materials exposed to sunlight and moisture is also a challenge, though new computational methods (mainly developed in the context of photocatalysis) are advancing this frontier.

Carefully designed experiments are necessary to advance the field. Experiments focusing on deeper understanding of PV materials and devices (rather than merely on increasing their efficiency) could dramatically move the field forward. Recent experiments are moving in this direction, in particular since the advent of complex

perovskites has stimulated experimental studies on the physics of PV materials. By contrast, when a decade ago the community was focusing on organic and dye-sensitized solar cells, the goal of much research was merely improving the efficiency, often by optimizing active layer materials by trial and error.

As discussed above, remarkable progress has been made in the last few years on charge carrier dynamics calculations. Two main challenges need to be addressed to make these methods more broadly applicable. One is computing carrier dynamics in the presence of excitons, which is relevant to solar cells. Since exciton transport is typically understood through heuristic diffusion models, first-principles calculations of exciton dynamics would significantly improve understanding of solar cells. The BSE (and to some extent, TD-DFT) can compute the exciton energy and wave function, but not exciton dynamics; new methods to compute exciton-phonon scattering and exciton dynamics are emerging as a new frontier in the field. Computing charge transport in the polaron regime is the second challenge. In organic semiconductors and oxides, localized carriers can couple strongly with lattice vibrations, forming bound states – called polarons – which can be thought of as electrons moving along with a phonon cloud. In the polaron transport regime, materials typically exhibit a low mobility with a peculiar temperature dependence. Poor charge transport due to polarons is the reason why, in spite of several attempts, oxide solar cells have never emerged as a viable alternative and organic semiconductors still struggle to make double-digit efficiencies. Predicting the properties of polarons is a long-standing challenge due to the strong (non-perturbative) electron-phonon interactions. While empirical models and even DFT calculations of polaron formation and hopping can be found in the literature, rigorous treatments of charge transport in the polaron regime remain an open challenge. In summary, new theory and first-principles methods are needed to advance both practical aspects and the physics of solar cells.

5   Contributions

This short section contains two contributions that discuss recent advances in modeling complex solar cell materials and ultrafast electron dynamics in photovoltaics. The authors use a variety of first-principles approaches beyond density functional theory to investigate the properties of novel materials and to microscopically understand solar energy conversion:

- ▶ Chapter 15, "Prototyping Ultrafast Charge Separation by Means of Time-Dependent Density Functional Methods" by Carlo Andrea Rozzi and Stefano Pittalis. This chapter focuses on charge separation, a key process in bulk heterojunction solar cells. The authors identify prototypical molecular systems for studying charge separation and provide new insights into the ultrafast dynamics of coupled electrons and nuclei using real-time time-dependent density functional theory. Method advances and strategies for improving device performance are also discussed.

- ▶ Chapter 14, "Hybrid Halide Perovskites: Fundamental Theory and Materials Design" by Flip, Volonakis and Giustino. This chapter reviews efforts by the author and his collaborators to model complex hybrid halide perovskites with promising power conversion efficiencies. Density functional theory, GW, and electron-phonon calculations are employed to shed light on the structure-property relations of this fascinating new family of compounds. Materials design efforts to develop novel halide perovskites where Pb is replaced by nontoxic elements are also discussed.

References

Bernardi M (2016) First-principles dynamics of electrons and phonons. Eur Phys J B 89(11):239

Bernardi M, Grossman JC (2016) Computer calculations across time and length scales in photovoltaic solar cells. Energy Environ Sci 9:2197–2218

Bernardi M, Palummo M, Grossman JC (2013) Extraordinary sunlight absorption and one nanometer thick photovoltaics using two-dimensional monolayer materials. Nano Lett 13(8):3664–3670

Bernardi M, Vigil-Fowler D, Lischner J, Neaton JB, Louie SG (2014) Ab initio study of hot carriers in the first picosecond after sunlight absorption in silicon. Phys Rev Lett 112:257402

Bernardi M, Mustafa J, Neaton JB, Louie SG (2015a) Theory and computation of hot carriers generated by surface plasmon polaritons in noble metals. Nat Commun 6:7044

Bernardi M, Vigil-Fowler D, Ong CS, Neaton JB, Louie SG (2015b) Ab initio study of hot electrons in GaAs. Proc Natl Ac Sci USA 112(17):5291–5296

Bernardi M, Ataca C, Palummo M, Grossman JC (2017) Optical and electronic properties of two-dimensional layered materials. Nanophotonics 6(2):479–493

Chen HY, Palummo M, Sangalli D, Bernardi M (2018) Theory and ab initio computation of the anisotropic light emission in monolayer transition metal dichalcogenides. Nano Lett 18(6):3839–3843

Frohna K, Deshpande T, Harter J, Peng W, Barker BA, Neaton JB, Louie SG, Bakr OM, Hsieh D, Bernardi M (2018) Inversion symmetry and bulk Rashba effect in methylammonium lead iodide perovskite single crystals. Nat Commun 9(1):1829

Furchi MM, Pospischil A, Libisch F, Burgdorfer J, Mueller T (2014) Photovoltaic effect in an electrically tunable van der Waals heterojunction. Nano Lett 14(8):4785–4791

Green MA, Ho-Baillie A, Snaith HJ (2014) The emergence of perovskite solar cells. Nat Photon 8(7):506–514

Jariwala D, Davoyan AR, Tagliabue G, Sherrott MC, Wong J, Atwater HA (2016) Near-unity absorption in van der Waals semiconductors for ultrathin optoelectronics. Nano Lett 16(9):5482–5487

Jean J, Brown PR, Jaffe RL, Buonassisi T, Bulović V (2015) Pathways for solar photovoltaics. Energy Environ Sci 8(4):1200–1219

Jhalani VA, Zhou JJ, Bernardi M (2017) Ultrafast hot carrier dynamics in GaN and its impact on the efficiency droop. Nano Lett 17(8):5012–5019

Lee NE, Zhou JJ, Agapito LA, Bernardi M (2018) Charge transport in organic molecular semiconductors from first principles: the bandlike hole mobility in a naphthalene crystal. Phys Rev B 97(11):115203

Liu M, Johnston MB, Snaith HJ (2013) Efficient planar heterojunction perovskite solar cells by vapour deposition. Nature 501(7467):395–398

Martiradonna L (2018) Riddles in perovskite research. Nat Mater 17(5):377

Mustafa JI, Bernardi M, Neaton JB, Louie SG (2016) Ab initio electronic relaxation times and transport in noble metals. Phys Rev B 94(15):155105

Palummo M, Bernardi M, Grossman JC (2015) Exciton radiative lifetimes in two-dimensional transition metal dichalcogenides. Nano Lett 15(5):2794–2800

Pospischil A, Furchi MM, Mueller T (2014) Solar-energy conversion and light emission in an atomic monolayer pn diode. Nat Nanotech 9(4):257–261

Zhou JJ, Bernardi M (2016) Ab initio electron mobility and polar phonon scattering in GaAs. Phys Rev B 94(20):201201

Zhou JJ, Hellman O, Bernardi M (2018) Electron-phonon scattering in the presence of soft modes and electron mobility in SrTiO$_3$ perovskite from first principles. Phys Rev Lett 121:226603 https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.121.226603

Hybrid Halide Perovskites: Fundamental Theory and Materials Design

14

Marina R. Filip, George Volonakis, and Feliciano Giustino

Contents

Abstract

Hybrid organic-inorganic halide perovskites have emerged as a disruptive new class of materials, exhibiting optimum properties for a broad range of opto-electronic applications, most notably for photovoltaics. The first report of highly efficient organic-inorganic perovskite solar cells in 2012 (Lee et al., Science 338:643–647, 2012) marked a new era for photovoltaics research, reporting a power conversion efficiency of over 10% (NREL, National renewable energy laboratory: best research-cell efficiencies. http://www.nrel.gov/ncpv/images/efficiency_chart.jpg, Accessed Nov 2017). Only five years after this

M. R. Filip · G. Volonakis
Department of Materials, University of Oxford, Oxford, UK

F. Giustino (✉)
Department of Materials Science and Engineering, Cornell University, Ithaca, NY, USA

Department of Materials, University of Oxford, Oxford, UK
e-mail: feliciano.giustino@materials.ox.ac.uk

discovery, perovskite photovoltaic devices have reached a certified efficiency of 22.7%, making them the first-solution processable technology to surpass thin film and multi-crystalline silicon solar cells (NREL, National renewable energy laboratory: best research-cell efficiencies. http://www.nrel.gov/ncpv/images/efficiency_chart.jpg, Accessed Nov 2017). The remarkable development of perovskite solar cells is due to the ideal optoelectronic properties of organic-inorganic lead-halide perovskites. The prototypical compound, methylammonium lead iodide, $CH_3NH_3PbI_3$ (Stranks and Snaith, Nat Nanotechnol 10:391–402, 2015), is a direct band gap semiconductor with a band gap in the visible, high charge carrier mobility, long diffusion length, and low excitonic binding energy (Johnston and Herz, Acc Chem Res 49(1):146–154, 2016). Due to these ideal properties, $CH_3NH_3PbI_3$ is also drawing interest across many other applications beyond photovoltaics, such as light-emitting devices (Tan et al., Nat Nanotechnol 9:687–692, 2014), lasers (Wehrenfennig et al., J Phys Chem Lett 5:1300–1306, 2014), photocatalysts (Chen et al., J Am Chem Soc 137(2):974–981, 2015), and transistors (Ward et al., ACS Appl Mater Interf 9(21):18,120–18,126, 2017).

The continued progress of metal-halide perovskite optoelectronics relies not only on a detailed understanding of the electronic and optical properties of materials in this class but also on the development of practical strategies to tune their properties by controlling parameters such as chemical composition. In this context, ab initio computational modelling can play a key role in providing a physical interpretation of experimental measurements and guiding the design of novel halide perovskites with tailored properties.

In this chapter we will present an account of the contributions to this fast-developing field of research from our computational modelling group. The chapter is organized in two sections. The first section focuses on the structural and optoelectronic properties of $CH_3NH_3PbI_3$. Here, we expand on some of the challenging aspects of modelling the electronic and vibrational properties of $CH_3NH_3PbI_3$ and discuss the main theoretical results alongside experimental data. The second section discusses the recent computationally led materials design of novel halide perovskites and the principal challenges in replacing Pb^{2+} in $CH_3NH_3PbI_3$ by nontoxic elements.

1   Methylammonium Lead Iodide

Methylammonium lead iodide, $CH_3NH_3PbI_3$, belongs to the ABX_3 perovskite structural family (Poglitsch and Weber 1987). As shown in Fig. 1, the Pb^{2+} and I^- ions form a three-dimensional network of corner-sharing octahedra. The organic $CH_3NH_3^+$ cations occupy the center of the cuboctahedral cavities enclosed by the inorganic PbI_6 network (Poglitsch and Weber 1987). The crystal structure of $CH_3NH_3PbI_3$ is strongly dependent on temperature and undergoes two phase transitions (Poglitsch and Weber 1987; Baikie et al. 2013; Stoumpos et al. 2013; Weller et al. 2015): from the low-temperature orthorhombic *Pnma* structure to
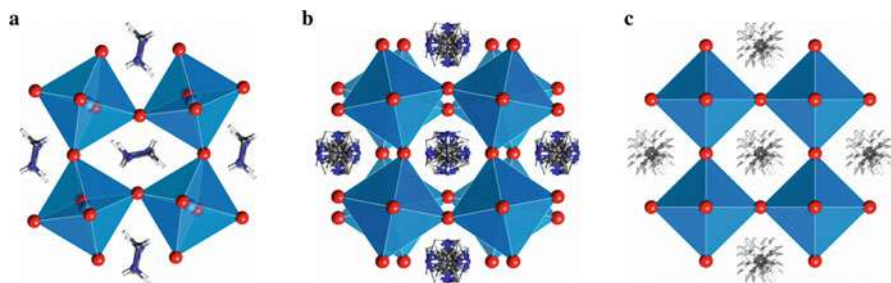
Fig. 1 **Temperature-dependent crystal structure of CH$_3$NH$_3$PbI$_3$.** Polyhedral representation of the low-temperature orthorhombic *Pnma* structure (**a**), the tetragonal *I*4/*mcm* structure (**b**), and the cubic *Pm*$\bar{3}$*m* structure (**c**). In all three panels, the Pb atoms are represented by the light blue spheres at the center of the octahedra, the I atoms are the red spheres at the corners of the octahedra, and the C, N, and H atoms are represented by the dark grey, dark blue, and white spheres, respectively. The figures were constructed using the lattice parameters and atomic positions reported by Weller et al. (2015)

the tetragonal *I*4/*mcm* structure at 160 K and the cubic *Pm*$\bar{3}$*m* structure at 315–330 K. As shown in Fig. 1, all phases of CH$_3$NH$_3$PbI$_3$ maintain the corner-sharing connectivity of the PbI$_6$ octahedra; however the degree of octahedral tilting is reduced as the temperature increases. In addition, the organic CH$_3$NH$_3^+$ cations exhibit ordered orientations throughout the unit cell only in the orthorhombic phases, while in the tetragonal and cubic phase, the cations are orientationally disordered (Poglitsch and Weber 1987; Baikie et al. 2013; Stoumpos et al. 2013; Weller et al. 2015).

The excited state properties of CH$_3$NH$_3$PbI$_3$ have been intensely investigated over the last 5 years. The optical absorption spectrum of CH$_3$NH$_3$PbI$_3$ is characteristic of a direct band gap semiconductor (Herz 2016), with a sharp absorption onset around 1.6 eV. At low temperature, the optical absorption spectrum exhibits an excitonic peak, with an exciton binding energy of ∼20 meV. This feature becomes less visible as the temperature increases (D'Innocenzo et al. 2014; Miyata et al. 2015; Herz 2016). In addition, the absorption line shape at room temperature exhibits an Urbach tail of 15 meV (De Wolf et al. 2014), unexpectedly low given the disordered character of the CH$_3$NH$_3^+$ cation. The optical band gap of CH$_3$NH$_3$PbI$_3$ increases with temperature from 1.62 eV at 4 K to 1.67 eV at 160 K. At the phase transition temperature between the orthorhombic and tetragonal phase, 160 K, the band gap decreases sharply to approximately 1.58 eV and then continues to increase beyond this temperature, as in the orthorhombic phase (Milot et al. 2015).

Free electrons and holes dominate over bound exciton pairs upon photoexcitation, an ideal characteristic for application in solar energy conversion (D'Innocenzo et al. 2014; Herz 2016). The photogenerated charge carriers are long lived in CH$_3$NH$_3$PbI$_3$ and exhibit charge carrier mobilities of approximately 35 cm$^2(Vs)^{-1}$ (Johnston and Herz 2016) at room temperature. Bimolecular recombination in CH$_3$NH$_3$PbI$_3$ is found to be the predominant mechanism for charge carrier

recombination at standard photovoltaic device operation conditions, exhibiting a highly non-Langevin behavior (Johnston and Herz 2016; Herz 2016). This means that the ratio between the bimolecular recombination rate $k_2 \sim 10^{-10}\,\mathrm{cm^3 s^{-1}}$ and the charge carrier mobility $\mu \sim 10\,\mathrm{cm^2 V^{-1} s^{-1}}$ is four orders of magnitude lower than the value $e/(\epsilon_0 \epsilon_r)$ predicted by the Langevin model (Herz 2016; Wehrenfennig et al. 2014). This effect explains the long charge carrier diffusion lengths ($\sim 1\,\mu\mathrm{m}$) measured for halide perovskites (Herz 2016).

The band gap of $CH_3NH_3PbI_3$ can be tuned via chemical substitution (Herz 2016). The replacement of the $CH_3NH_3^+$ cation with $CH(NH_2)_2^+$ (formamidinium), Cs^+, or Rb^+ leads to a change in the optical band gap over a range of 0.3 eV (Eperon et al. 2014; Filip et al. 2014); the substitution of Pb^{2+} by Sn^{2+} reduces the band gap by approximately 0.4 eV (Hao et al. 2014), and the replacement of I^- by the smaller anions Cl^- and Br^- blue-shifts the optical band gap by up to 0.7 eV (Comin et al. 2015; Noh et al. 2013). In practice, mixed cation, mixed metal, and mixed halide perovskites are currently explored in order to optimize light absorption over the entire visible spectrum, but also to improve the stability of perovskite solar cells (McMeekin et al. 2016; Eperon et al. 2016; Saliba et al. 2016).

Theoretical investigations of the fundamental structural, electronic, and vibrational properties of $CH_3NH_3PbI_3$ have played an important role in the interpretation of the experimental data summarized above. In the following we discuss some of the theoretical results obtained by us and our coworkers in the study of $CH_3NH_3PbI_3$.

1.1   Basic Electronic Structure of $CH_3NH_3PbI_3$

To construct the electronic band structure of $CH_3NH_3PbI_3$, it is necessary to know the lattice parameters and the atomic positions for the perovskite unit cell. For the low-temperature orthorhombic phase, the unit cell was fully resolved for the first time by Baikie et al. (2013). The unit cell contains four PbI_6 octahedra and four $CH_3NH_3^+$ cations, 48 atoms in total, arranged according to the *Pnma* space group.

In the tetragonal and cubic phases, the unit cell contains four and one PbI_6 octahedra, respectively, and the shape of the inorganic octahedral networks are well described (Stoumpos et al. 2013; Weller et al. 2015) by the space groups $I4/mcm$ and $Pm\bar{3}m$, respectively. At variance with the inorganic scaffold, the organic cations have a disordered orientation inside the perovskite cavity, in the tetragonal and cubic phases, and pose difficulties for any electronic structure calculation that employs periodic boundary conditions. For example, structural relaxation of the unit cell starting from arbitrary orientations of the $CH_3NH_3^+$ cations leads to significant distortions of the unit cell and considerable changes to the electronic band structure that are highly dependent on the choice of the orientation (Motta et al. 2015). This effect is purely an artefact of the chosen orientation of the cation and bears no physical meaning.

To ensure that we can obtain meaningful results of the electronic properties of $CH_3NH_3PbI_3$, we have two options. One approach is to simulate the disordered

orientation of the $CH_3NH_3^+$ by averaging over its many possible orientations. This approach is best addressed using molecular dynamics (Carignano and Kachmar 2015; Lahnsteiner et al. 2016) and requires a careful study of the size of the supercell in order to correctly represent the disordered orientations of $CH_3NH_3^+$. The second approach relies on the general property of ABX_3 perovskites by which the A-site cations do not participate in bonds, and as a consequence, they only contribute to electronic states far away from the band edges (Megaw 1973). In this latter approach, we can first calculate the electronic properties of $CH_3NH_3PbI_3$ in the well-known orthorhombic phase. Then we can investigate how the change in the shape of the inorganic octahedral network affects these properties. In the following, we describe the latter strategy.

In Fig. 1a we show the projected density of states of $CH_3NH_3PbI_3$ in the low-temperature orthorhombic *Pnma* case. The valence band top is occupied by I-p and Pb-s electrons, while the conduction band bottom is of I-p and Pb-p character. As expected, the states localized on the $CH_3NH_3^+$ cation are located away from the band edges (approximately 5 eV). Based on Fig. 2a we can construct a tight-binding argument to explain the tunability of the band gap by substitution of I anions or Pb cations. The replacement of I by Cl or Br leads to an increase in the band gap, due to the lower-lying Br-p and Cl-p electronic states that populate the valence band top; similarly, the substitution of Pb by Sn lowers the conduction band bottom and reduces the band gap due to the lower-lying Sn-p states. In addition, from Fig. 2a, it is clear that the $CH_3NH_3^+$ cations do not contribute to the optical absorption, and their role is to act as a structural filler and to ensure the charge neutrality of the unit cell. In practice, the $CH_3NH_3^+$ cation can be replaced by a positive background charge, and the electronic band structure would not change significantly (Filip et al. 2014).

Taking advantage of this property, we can study how the shape of the cuboctahedral cavity influences the electronic structure of the perovskite. Filip et al. (2014) showed that the unit cell of the perovskite lattice can be uniquely described by a "Platonic" model, whereby the PbI_6 octahedra are ideal and identical throughout the lattice, and their orientation within the unit cell is uniquely described by two angles: the equatorial and the apical bond angles (Fig. 2c). Figure 2b depicts the variation of the band gap with the two bond angles of the "Platonic" perovskite lattice. The gap increases as the bond angles decrease.

Two observations can be drawn from Fig. 2b. Firstly, the size of the A-site cation in the perovskite lattice influences the shape of the octahedral lattice, by changing the apical and equatorial bond angles, which in turn enables the control of the band gap. This leads to the following simple trend: the smaller the cation, the smaller the bond angles, the larger the band gap. The band gap trend predicted in Fig. 2b is in agreement with the trend observed experimentally and explains the band gap tunability with the substitution of the A-site cation (Eperon et al. 2014; Filip et al. 2014). Secondly, based on Fig. 2b we can explain the discontinuity in the band gap dependence on temperature observed at the transition temperature of 160 K (Milot et al. 2015). When the $CH_3NH_3PbI_3$ transitions from an orthorhombic to
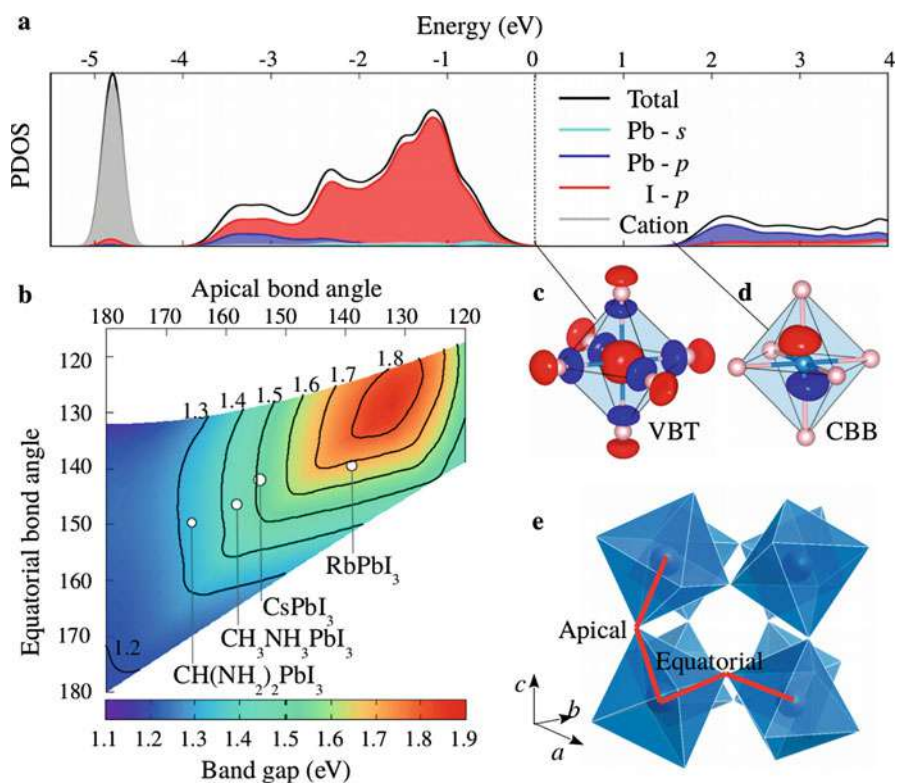
Fig. 2 Electronic structure and band gap tunability of CH₃NH₃PbI₃. (**a**) Projected density of states of CH$_3$NH$_3$PbI$_3$ calculated within the local density approximation to density functional theory (DFT/LDA) without spin-orbit coupling for the orthorhombic crystal structure. The bands are aligned to the valence band top (VBT) represented by the dotted line. (Adapted with permission from Filip et al. (2014), Copyright (2014) Nature Publishing Group). (**b**) Two-dimensional map of the DFT band gap as a function of the apical and equatorial bond angles. The band gaps for each point of the map were calculated for Platonic model structures, as described by Filip et al. (2014), which do not contain the A-site cation, as shown in (**e**). The white disks represent the angular coordinates (apical and equatorial bond angles) of realistic models of CH$_3$NH$_3$PbI$_3$, CH(NH$_2$)$_2$PbI$_3$, CsPbI$_3$ and RbPbI$_3$. (Adapted with permission from Filip et al. (2014), Copyright (2014) Nature Publishing Group). (**c–d**) Wave functions corresponding to the valence band top (**c**) and the conduction band bottom of a lead iodide perovskite (**d**). (**e**) Polyhedral representation of the Platonic model for the orthorhombic perovskite structure. The apical and equatorial bond angles are represented by the red lines in the figure

a tetragonal phase, the apical bond angle increases from 160° to 180°. According to the map in Fig. 2b, the band gap should decrease by up to 0.2 eV, in good agreement with the sudden drop of 0.1 eV measured around 160 K.

The core conclusion that can be drawn from Fig. 2a is that the fundamental roles of the organic cation CH$_3$NH$_3^+$ in the lead iodide perovskite lattice are to balance the charge and to stabilize the shape of the perovskite lattice. Therefore, in order

to ensure that we are studying a physically meaningful crystal structure, in the following discussions, we focus on the low-temperature orthorhombic *Pnma* phase, as refined by Baikie et al. (2013).

1.2   The Quasiparticle Band Structure of $CH_3NH_3PbI_3$

The calculations shown so far in Fig. 2 were obtained within DFT in the scalar-relativistic approximation. However, the presence of Pb in $CH_3NH_3PbI_3$ requires a relativistic treatment. In Fig. 3a we show that the spin-orbit coupling effect is crucial for the correct calculation of the band edge topology. Not only relativistic effects lower the conduction band bottom of $CH_3NH_3PbI_3$ by approximately 1 eV (Even et al. 2013), but also they change the shape of the band edges. In particular, the effective masses become fully isotropic upon inclusion of spin-orbit coupling, by
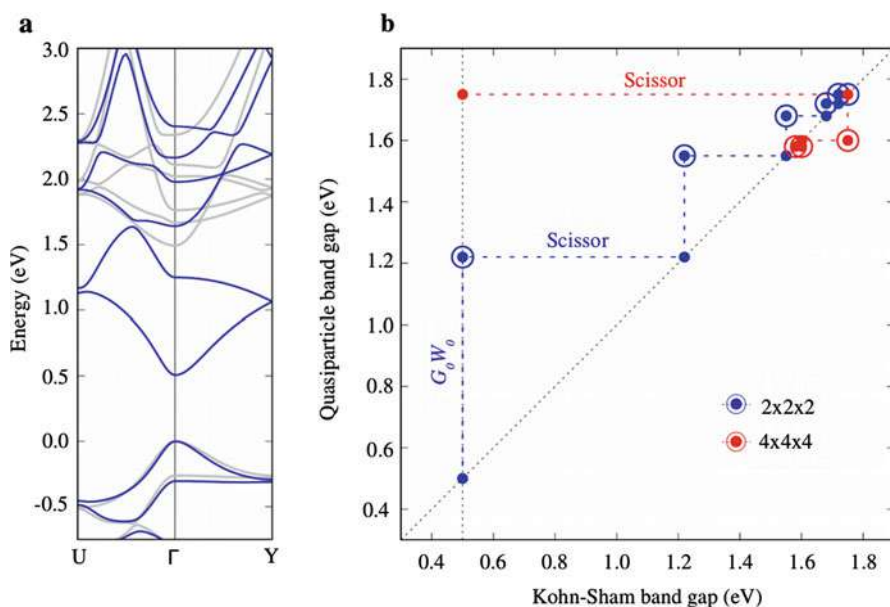
Fig. 3 Band structure of $CH_3NH_3PbI_3$. (**a**) DFT/LDA band structure of $CH_3NH_3PbI_3$ calculated with (blue line) and without (grey line) including spin-orbit coupling effects. The calculations were performed for the orthorhombic *Pnma* phase using norm-conserving Troullier-Martins pseudopotentials as described by Filip and Giustino (2014). (**b**) Convergence of the self-consistent scissor GW scheme for the quasiparticle band gap of $CH_3NH_3PbI_3$. The self-consistent scissor GW scheme is described in detail by Filip and Giustino (2014). For all quasiparticle calculations, we have used 1000 bands and plane wave cutoffs of 35 Ry and 6 Ry for the exchange and polarizability, respectively. The diagram shows the self-consistent scissor scheme calculated for a Γ-centered $2 \times 2 \times 2$ **k**-point grid (blue) and a $4 \times 4 \times 4$ **k**-point grid (red). All quasiparticle calculations include semicore d electrons for both Pb and I, as discussed by Filip and Giustino (2014)

contrast with those calculated in the scalar-relativistic approach (Menéndez-Proupin et al. 2014). As shown in Table 1, fully relativistic DFT calculations underestimate the band gap by ∼1 eV and the reduced effective mass by a factor of two.

The underestimation of the band gap and effective masses of DFT can be corrected by including quasiparticle effects. The quasiparticle band gap and band structure can be calculated within the GW approximation (Hedin 1965). The quasiparticle eigenvalues $E_{n\mathbf{k}}$ are calculated starting from the single-particle DFT/LDA eigenvalues $\epsilon_{n\mathbf{k}}$ as (Hedin 1965; Hybertsen and Louie 1986):

$$E_{n\mathbf{k}} = \epsilon_{n\mathbf{k}} + Z(\epsilon_{n\mathbf{k}})\langle n\mathbf{k}|\hat{\Sigma}(\epsilon_{n\mathbf{k}}) - V_{xc}|n\mathbf{k}\rangle, \qquad (1)$$

where $|n\mathbf{k}\rangle$ denotes a single-particle eigenstate with band index n and crystal momentum $\mathbf{k}$, $Z(\omega) = (1 - \partial \mathrm{Re}\hat{\Sigma}/\partial\omega)^{-1}$ is the quasiparticle renormalization, V_{xc} is the exchange-correlation potential, and $\hat{\Sigma} = iG_0W_0$ is the quasiparticle self-energy (Hybertsen and Louie 1986). In the expression of $\hat{\Sigma}$, G_0 is the non-interacting Green's function calculated using single-particle eigenvalues and eigenstates as obtained from DFT. The screened Coulomb interaction has the expression $W_0 = \epsilon^{-1}v$, where v is the bare Coulomb interaction and ϵ^{-1} is the inverse dielectric matrix.

In the case of $CH_3NH_3PbI_3$, it has been shown that the one-shot G_0W_0 quasiparticle band gap is underestimated with respect to experiment by up to

Table 1 Calculated values of the band gap, high- and low-frequency dielectric constants, and electron, hole, and reduced effective masses within DFT/LDA and within the GW approximation including the self-consistent scissor correction of Filip and Giustino (2014). The band gaps are expressed in eV, and the effective masses are expressed in the unit of electron masses (m_e). The reduced effective mass is calculated using the following expression: $\mu^* = \frac{m_h^* m_e^*}{m_h^* + m_e^*}$, where m_e^*, m_h^*, and μ^* are the electron, hole, and reduced effective masses, respectively. The exciton binding energy is calculated using the expression $E_b = \mu^*/\epsilon_\infty^2 \times E_{Ry}$ eV, with $E_{Ry} = 13.60565$ eV, the reduced effective mass $\mu^* = 0.11\,m_e$, and the high-frequency dielectric constant $\epsilon_\infty = 5.8$

	DFT	G_0W_0	GW	Exp.
Band gap (eV)	0.50[a,b,c]	1.20[a,b,c]	1.57[c]	1.65[c,d]
ϵ_∞		5.8[b,e]		6.5[f]
ϵ_0		25.3[e]		30.5[g]
Hole effective mass (m_e)	0.14[b,c]	0.18[b,c]	0.23[b,c]	N/A
Electron effective mass (m_e)	0.12[b,c]	0.16[b,c]	0.22[b,c]	N/A
Reduced effective mass (m_e)	0.06[b,c]	0.09[b,c]	0.11[b,c]	0.104±0.003[h]
Exciton binding energy (meV)		44		16[h]

[a]Filip and Giustino (2014)
[b]Filip et al. (2015)
[c]Davies et al. (2018)
[d]Milot et al. (2015)
[e]Pérez-Osorio et al. (2015)
[f]Hirasawa et al. (1994)
[g]Poglitsch and Weber (1987)
[h]Miyata et al. (2015)

0.4 eV (Brivio et al. 2014; Filip and Giustino 2014), as shown in Table 1. This is primarily due to the underestimated single-particle band gap calculated at the DFT level, which leads to an overestimation of the dielectric screening in W_0 (Brivio et al. 2014; Filip and Giustino 2014). The quasiparticle band gap underestimation at the G_0W_0 level appears irrespective of the exchange-correlation functional employed for the DFT starting point, and it tends to worsen when more core electrons are included in the valence configuration of I (Scherpelz et al. 2016). However, this sensitivity to the DFT starting point can be mitigated by calculating the electron self-energy self-consistently (van Schilfgaarde et al. 2006).

Figure 3b shows an approximate self-consistency scheme introduced by Kioupakis et al. (2008) and applied to the case of $CH_3NH_3PbI_3$ by Filip and Giustino (2014), the self-consistent scissor GW. In this approach the G_0W_0 band gap correction Δ is first calculated, and then a scissor correction of magnitude Δ is applied to all eigenstates in the conduction band. These two steps are repeated until the band gap is converged, as shown in Fig. 3b (Kioupakis et al. 2008; Filip and Giustino 2014). The final converged band gap is of 1.57 eV, within 80 meV of the measured optical band gap of 1.65 eV (Milot et al. 2015; Davies et al. 2018). This small difference is accounted for by the absence of electron-hole interaction effects and electron-phonon effects.

Figure 4a shows the comparison between the GW quasiparticle band structure and the DFT/LDA band structure obtained from Wannier interpolation (Marzari and Vanderbilt 1997; Souza et al. 2001). Qualitatively the band dispersions calculated within DFT/LDA and GW are similar, with the quasiparticle valence band exhibiting a noticeably more pronounced curvature. A more quantitative comparison between the DFT/LDA and GW band dispersions is given by the isotropic effective masses, as shown in Table 1. The effective mass tensor is calculated as the inverse of the second derivatives of the energy with respect to the crystal momentum, $m_{ij}^* = \hbar^2(\partial^2\epsilon/\partial k_i\partial k_j)^{-1}$, and the values reported in Table 1 are the averages of the eigenvalues of the effective mass tensors. From Table 1 we can see that the DFT/LDA and GW effective masses differ by almost a factor of two. In addition, the GW reduced effective mass of $0.11\,m_e$ is in excellent agreement with the experimental value of $0.10\,m_e$ reported by Miyata et al. (2015), thus establishing the accuracy of the quasiparticle band structure calculations.

An important consequence of the quasiparticle effects in the calculation of the band structure of $CH_3NH_3PbI_3$ is the band edge parabolicity. While both DFT/LDA and GW yield isotropic band edges, the DFT/LDA bands are parabolic only in a narrow energy range near the band edges. On the other hand, the quasiparticle valence and conduction band edges exhibit a parabolic profile over a range of 300 meV (Fig. 4b). This observation is most clearly visualized in the joint density of states (JDOS), as shown in Fig 4b, where the parabolic shape of the bands extends over approximately 0.3 eV for the quasiparticle JDOS, while the JDOS calculated from DFT/LDA departs from the parabolic shape within 50 meV from the onset.

The parabolicity of the conduction and valence band edges has important implications in the understanding of both absorption and recombination mechanisms in $CH_3NH_3PbI_3$. Using a combined experimental and theoretical analysis of the
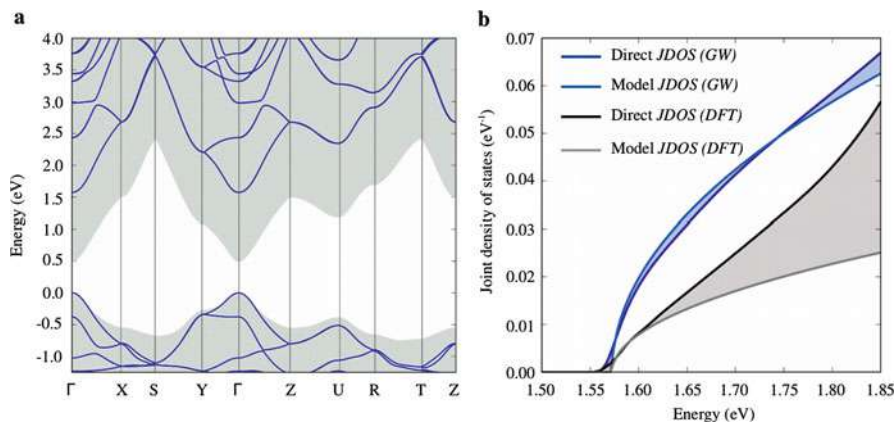
Fig. 4 Band parabolicity. (**a**) Quasiparticle band structure of $CH_3NH_3PbI_3$. The quasiparticle band structure is represented by the blue lines; the grey background is the DFT/LDA band structure. The quasiparticle band structure was obtained by performing a Wannier interpolation of the quasiparticle eigenvalues calculated for a $4 \times 4 \times 4$ Γ-centered grid, using the GW and the self-consistent scissor scheme (Filip et al. 2015; Davies et al. 2018). (**b**) Joint density of states of $CH_3NH_3PbI_3$ calculated within DFT (black line) and GW (dark blue line). The direct joint densities of states were calculated using the definition $J(\omega) = \sum_{cv\mathbf{k}} \delta(\epsilon_{c\mathbf{k}} - \epsilon_{v\mathbf{k}} - \omega)$, where c and v are the band indices for the conduction and valence band states, respectively, and $\mathbf{k}$ is the crystal momentum. The summation over the crystal momenta is performed by discretizing the Brillouin zone on a Γ-centered grid of $40 \times 40 \times 40$ points. The DFT and GW eigenvalues are obtained on this grid from Wannier interpolation (Filip et al. 2015; Davies et al. 2018). The model joint densities of states (light blue line for GW and grey for DFT) are calculated using the assumption that the bands are parabolic using, $J^{\mathrm{DFT}/GW}(\omega) = 1/4\pi^2 (2\mu^{\mathrm{DFT}/GW}/\hbar^2)^{3/2} (\omega - E_g)^{1/2}$, where μ is the reduced effective mass calculated from DFT ($0.06\,m_e$) or GW ($0.11\,m_e$). Both the model and the direct JDOS calculated from DFT were rigidly shifted in order to match the GW onset, for clarity

optical absorption line shape, Davies et al. (2018) showed that the optical absorption spectrum of $CH_3NH_3PbI_3$ can be modelled using Elliot's theory (Elliott 1957) up to 100 meV above the absorption onset. Within this premise, Davies et al. (2018) decoupled the excitonic and continuum part of the absorption spectrum and calculated the bimolecular recombination rate using the van Roosbroeck-Shockley formalism (van Roosbroeck and Shockley 1954). By comparing this model to direct transient spectroscopic measurements, Davies et al. (2018) demonstrated that bimolecular recombination in $CH_3NH_3PbI_3$ is an inverse absorption process. This result reinforces the conclusion that $CH_3NH_3PbI_3$ is in many respects very similar to conventional semiconductors, such as GaAs (Herz 2016).

1.3   Vibrational Properties of $CH_3NH_3PbI_3$

The vibrational spectrum is a key piece of information in the study of the optoelectronic properties of $CH_3NH_3PbI_3$. Lattice vibrations can be analyzed experimentally from IR (Pérez-Osorio et al. 2015) and Raman (Ledinský et al. 2015) spectroscopy,

as well as inelastic neutron scattering (Druzbicki et al. 2016), and can be directly compared with ab initio calculations. Compared to the latter two techniques, IR spectroscopy is most accessible, as high-resolution measurements can be achieved in a short time frame over a wide range of frequencies (Pérez-Osorio et al. 2015).

The vibrational properties of $CH_3NH_3PbI_3$ can be calculated within density functional perturbation theory (DFPT). The vibrational spectrum of $CH_3NH_3PbI_3$ in the orthorhombic *Pnma* phase consists of 144 normal modes at the Γ point, with frequencies extending up to $3100\,cm^{-1}$ (Pérez-Osorio et al. 2015). The vibrational modes of $CH_3NH_3PbI_3$ can be classified into three main categories: vibrations of the $CH_3NH_3^+$ cations, vibrations of the inorganic PbI_3 network, and mixed modes. In addition, the vibrations of the $CH_3NH_3^+$ cations ca be decomposed into rigid translations, spinning vibrations around the C-N axis, librations around the center of mass of the cation, and internal vibrations. Up to $60\,cm^{-1}$, the main contribution to the vibrational phonon modes comes from the internal vibrations of the inorganic network, while above $60\,cm^{-1}$, the vibrational phonon modes are almost exclusively due to the vibrations of the organic cations (Pérez-Osorio et al. 2015). Pérez-Osorio et al. (2015) describes the precise contribution of these three categories of vibrations to each of the vibrational modes. These contributions were calculated by a systematic decomposition of each mode followed by a factor-group analysis of their symmetries.

Using the calculated vibrational frequencies of $CH_3NH_3PbI_3$, the IR spectrum is obtained by calculating the absorption spectrum in the IR frequency range as (Pérez-Osorio et al. 2015):

$$I(\omega) = \frac{e^2}{M_0} \sum_{\alpha v} |Z_{\alpha v}^*|^2 \delta(\omega - \omega_v), \tag{2}$$

where M_0 is the average mass over the unit cell, e is the electronic charge, ω_v are the vibrational eigenfrequencies at the zone center, and $Z_{\alpha v}^* = \sum_{\kappa \beta} \sqrt{\frac{M_0}{M_\kappa}} e_{\kappa \beta, v} Z_{\kappa, \alpha \beta}^*$, with M_κ being the nuclear mass of atom κ in the unit cell; $e_{\kappa \beta, v}$ the vibrational eigenmode of atom κ in the Cartesian direction β, with vibrational frequency ω_v; and $Z_{\kappa, \alpha \beta}^*$ the Born effective charge (Pérez-Osorio et al. 2015).

In Fig. 5 we show a comparison between the experimental and theoretical IR spectra of $CH_3NH_3PbI_3$, as measured and calculated by Pérez-Osorio et al. (2015). Each of the features in the experimental IR spectrum is assigned to a peak and a vibrational symmetry in the calculated spectrum. The theoretical and experimental spectra agree well for almost the entire frequency range. The two main discrepancies are shown in Fig. 5c–f. At high frequency, the three peaks are redshifted by approximately $70\,cm^{-1}$. This discrepancy is shown by Pérez-Osorio et al. (2015) to be corrected by the use of the generalized gradient approximation (GGA) instead of LDA for the calculation of the exchange-correlation potential in DFPT. In the low-frequency (Fig. 5c, d) region of the spectrum, DFPT calculations appear to overestimate the vibrational frequencies of the librational modes of $CH_3NH_3^+$ around $150\,cm^{-1}$. This discrepancy (Fig. 5e, f) is consistent with other studies in
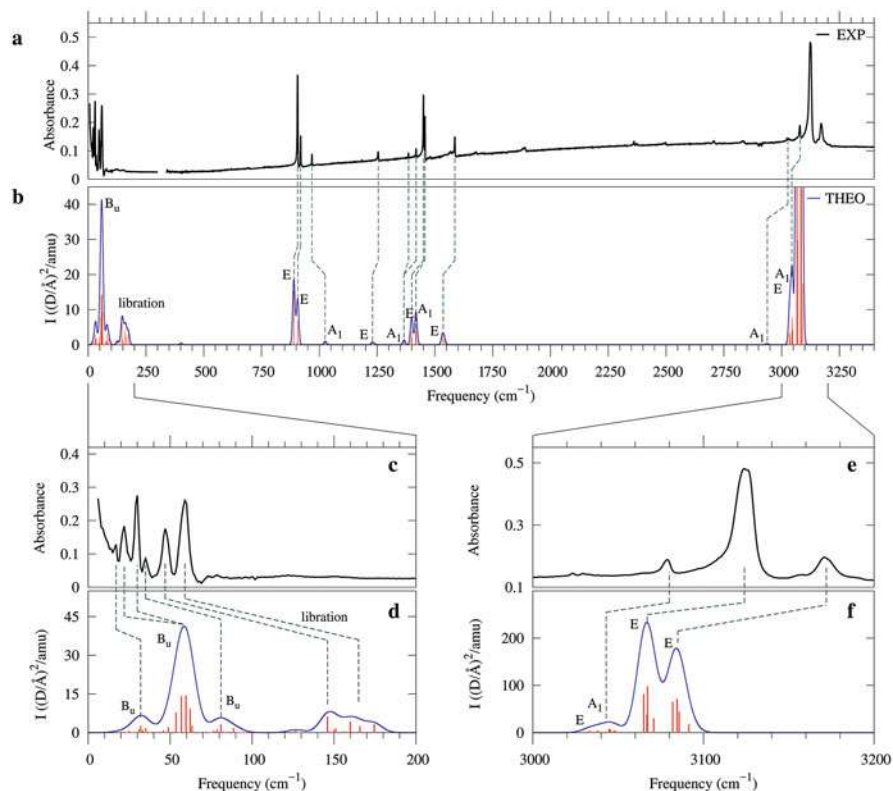
Fig. 5 Experimental and theoretical IR spectra of $CH_3NH_3PbI_3$ at low temperature. (a) IR spectrum of CH_3NH_3 measured on thin film at 10 K. The dashed lines indicate the assignment of the most prominent peaks of the spectrum to features in the calculated IR spectrum in (**b**). The IR spectrum is measured at normal incidence, and only TO normal modes can be excited in this geometry. (**b**). Calculated IR spectrum of $CH_3NH_3PbI_3$ in the low-temperature orthorhombic structure. The red sticks are the calculated infrared intensities, and the blue line corresponds to the spectrum calculated by representing each absorption peak with a Gaussian broadening. To be able to compare with the experimental data obtained within the geometry described in (**a**), only TO modes are included in the calculated IR spectrum, as discussed by Pérez-Osorio et al. (2015). (**c, d**). Close-up of the comparison between the experimental (**c**) and theoretical (**d**) IR spectra in the low-frequency region. (**e, f**) Close-up of the comparison between experimental (**e**) and theoretical (**f**) IR spectra in the high-frequency region. (Reprinted with permission from Pérez-Osorio et al. (2015), Copyright (2015) American Chemical Society)

literature (Mosconi et al. 2014; Družbicki et al. 2016), and it appears regardless of the choice of exchange-correlation functional (LDA vs. GGA), inclusion of relativistic spin-orbit coupling effects, van der Waals interactions, or anharmonic effects (Pérez-Osorio et al. 2017). Moreover, Pérez-Osorio et al. (2017) show that the calculated vibrational density of states is in good agreement with inelastic neutron scattering spectra (Družbicki et al. 2016), suggesting that the discrepancy

between the experimental and calculated IR spectra may be related to a mismatch between the experimental and theoretical IR intensities.

Having established both the electronic and vibrational properties of $CH_3NH_3PbI_3$, we have access to the high- and low-frequency dielectric permittivities, ϵ_∞ and ϵ_0, respectively. The high-frequency permittivity does not depend on lattice vibrations, but its value does depend on the calculated band gap. For this reason, the closest agreement with the experimental value $\varepsilon_\infty = 6.5$ (Hirasawa et al. 1994) is obtained from finite-electric field calculations within scalar-relativistic DFT (Pérez-Osorio et al. 2015; Filip et al. 2015) (5.8, see Table 1). The low-frequency permittivity depends on the vibrational frequencies of $CH_3NH_3PbI_3$, with the vibrations of the inorganic PbI_6 network carrying the largest contribution (Pérez-Osorio et al. 2015). Pérez-Osorio et al. (2015) reports a value of $\varepsilon_0 = 25.3$, as calculated from DFPT, in very good agreement with the experimental low-frequency dielectric constant $\varepsilon_0 = 30.5$ (Poglitsch and Weber 1987). As shown by Pérez-Osorio et al. (2015), the large difference between the high- and low-frequency dielectric permittivities is due almost entirely to the vibrations of the inorganic PbI_6 network.

1.4   Electron-Phonon Coupling in $CH_3NH_3PbI_3$

Understanding the interaction between electrons and lattice vibrations is essential in the investigation of the transport properties and charge carrier recombinations in $CH_3NH_3PbI_3$. In particular, electron-phonon coupling manifests itself prominently in the temperature-dependent charge carrier mobility and in the broadening of the photoluminescence spectra (Johnston and Herz 2016). The charge carrier mobility decreases with increasing temperature as T^m, where m takes values between -1.4 and -1.6 (Brenner et al. 2015; Wright et al. 2016). Such behavior is typical of nonpolar semiconductors such as silicon or germanium, where electron-phonon interactions at room temperature are almost exclusively governed by the interaction with acoustic phonons. It is, however, unexpected in the case of $CH_3NH_3PbI_3$, given the polar nature of the this compound.

As shown by Wright et al. (2016), the nature of electron-phonon interactions in $CH_3NH_3PbI_3$ can be investigated by analyzing the photoluminescence linewidth as a function of temperature. In their study, Wright et al. (2016) find that the contribution of acoustic phonons to the broadening of the photoluminescence linewidth is negligible compared to that of the longitudinal-optical (LO) modes. This argument is further corroborated by first principles calculations of the electron-phonon coupling, as shown in Fig. 6 (Wright et al. 2016).

The photoluminescence linewidth can be directly associated with the imaginary part of the electron-phonon self-energy, calculated using the following expression (Wright et al. 2016):

$$\Sigma_{n\mathbf{k}} = \sum_{m\nu\mathbf{q}} |g_{mn}^\nu(\mathbf{k}, \mathbf{q})|^2 \left[\frac{n_{\mathbf{q},\nu} + f_{m\mathbf{k}+\mathbf{q}}}{\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}+\mathbf{q}} - \omega_{\mathbf{q}\nu} - i\eta} + \frac{n_{\mathbf{q},\nu} + 1 - f_{m\mathbf{k}+\mathbf{q}}}{\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}+\mathbf{q}} - \omega_{\mathbf{q}\nu} - i\eta} \right],$$

$$(3)$$
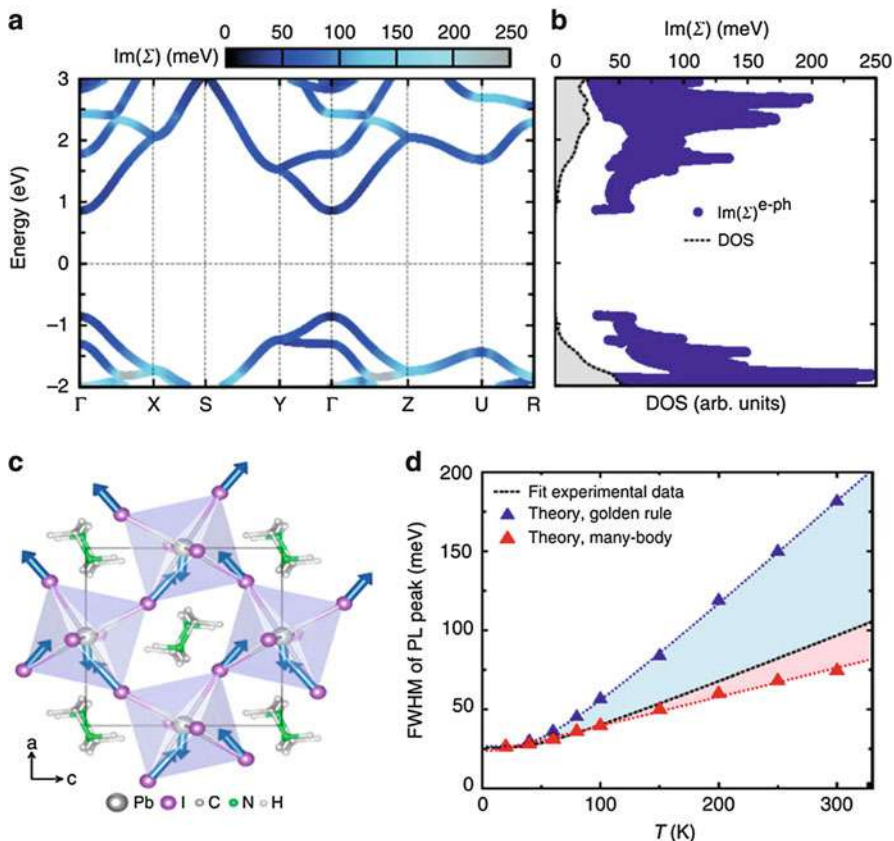
Fig. 6 Electron-phonon coupling and PL broadening of $CH_3NH_3PbI_3$. (**a**) GW quasiparticle band structure of the orthorhombic $CH_3NH_3PbI_3$, combined with the heat map of the imaginary part of the electron-phonon self-energy, $Im(\Sigma)$ at T = 200 K. The zero of the energy is placed in the middle of the band gap. (**b**) Imaginary part of the electron-phonon self-energy (dark blue line) and the electronic density of states (dotted black line). $2Im(\Sigma)$ corresponds to the electron-hole linewidth arising from electron-phonon coupling (without the quasiparticle renormalization factor Z). (**c**) Ball-and-stick representation of the LO vibration responsible for the broadening of the PL peaks. (**d**) Temperature dependence of the full width at half maximum (FWHM) of the PL peak in $CH_3NH_3PbI_3$. The dashed black line corresponds to the fit to experimental data, and the blue and red triangles correspond to theoretical calculations using Fermi's golden rule and the Brillouin Wigner perturbation theory (red triangles), respectively. The theoretical broadening is obtained as the sum of $2Im(\Sigma)$ at the valence and conduction band edges, rigidly shifted by a constant inhomogeneous broadening of 26 meV. (Reprinted with permission from Wright et al. (2016), Copyright (2016) Nature Publishing Group)

where $f_{m\mathbf{k}+\mathbf{q}}$ and $n_{\mathbf{q},\nu}$ are the Fermi-Dirac and Bose-Einstein occupations; $\epsilon_{n\mathbf{k}}$ and $\hbar\omega_{\mathbf{q}\nu}$ are the electron and phonon energies, respectively; η is a small broadening parameter; and $g_{mn}^{\nu}(\mathbf{k}, \mathbf{q})$ is the electron-phonon matrix element calculated by taking into account only the interaction with the LO phonons, as discussed by Verdi and Giustino (2015).

Figure 6a shows how the imaginary part of the electron and hole linewidth is distributed across the electronic states in the quasiparticle band structure (Wright et al. 2016). The linewidth increases as the density of states increases (Fig. 6b), because more states become available for electronic transitions. The predominant contribution to the electron-phonon self-energy comes from the LO phonon mode with an energy of 13 meV, which corresponds to the vibrations shown in Fig. 6c. This value is in excellent agreement with the dominant phonon energy of 11 meV determined from the experiments (Wright et al. 2016). Figure 6d shows that the measured and the calculated trend for the PL broadening as a function of temperature is in very good agreement, thereby confirming that the predominant electron-phonon interaction is that between charge carriers and longitudinal optical phonons. In addition, both theoretical and experimental results reported by Wright et al. (2016) show that $CH_3NH_3PbBr_3$ exhibits an electron-phonon coupling which is 40% stronger than for the iodide, pointing to potential directions for controlling such effects through chemical substitution.

1.5 Desirable Electronic Structure Properties and the Basis for Materials Design

Experimental and theoretical studies of the electronic and optical properties of $CH_3NH_3PbI_3$ have so far elucidated many of its fundamental properties. Perhaps the most striking finding is that $CH_3NH_3PbI_3$ exhibits optoelectronic properties which are remarkably similar to the best inorganic semiconductors such as GaAs. Despite much progress on the understanding of the fundamental properties of $CH_3NH_3PbI_3$ and related compounds, several important questions remain to be addressed. For example, it would be important to investigate charge carrier transport and recombination processes (Herz 2016), degradation mechanisms (Leijtens et al. 2017), and defect physics (Ball and Petrozza 2017). All these topics are active areas of current research in this field.

Much of the current success of halide perovskites owes to the continued in-depth investigation of their fundamental optoelectronic properties, but also to the development of practical strategies to control these properties through chemical substitution and to design new materials with targeted functionalities. For example, chemical substitution has been systematically explored as a means to tune the optical absorption properties (Filip et al. 2014; Eperon et al. 2016), improve the stability (McMeekin et al. 2016), and reduce the toxicity (Giustino and Snaith 2016) of lead-halide perovskites.

In particular, there is currently an ongoing search for alternative lead-free halide perovskites which retain the optoelectronic properties of $CH_3NH_3PbI_3$ and remove any potential environmental concerns due to the presence of lead. The nearly ideal optoelectronic properties of lead-halide perovskites set a very high bar for potential lead-free candidates: new compounds must exhibit a band gap in the visible range, long carrier lifetimes, good charge carrier mobilities, low exciton binding energies,

and shallow or electrically inactive defects. In the next section, we review our efforts in the search for novel halide perovskites that might exhibit this unique combination of optoelectronic properties.

2   Design of Lead-Free Perovskites

The presence of lead in lead-halide perovskites has raised questions on the environmental impact of perovskite solar cells. While the content of lead by weight in these devices is well below the limits set by governmental agencies, the prospect of developing lead-free perovskites is very attractive. Until now a lead-free material that can compete with $CH_3NH_3PbI_3$ and related compounds has not been found, but substantial progress has been made in the development of new lead-free perovskites.

   Among the different approaches to develop new lead-free perovskites, first-principles computational design has emerged as a powerful tool for screening new materials and identifying promising candidates. In fact, over the last two years, several new lead-free halide double perovskites have been first proposed by theoretical studies and subsequently synthesized in the lab. In the next sections, representative examples in this area are discussed.

2.1   Homovalent Pb Replacement

In order to replace lead in the $APbX_3$ structure, the simplest starting point is to search for alternative cations in the +2 oxidation state. Hypothetical new perovskites should be stable toward air, moisture, illumination, and heat and should exhibit band gap and carrier effective masses similar to those of $CH_3NH_3PbI_3$. The most obvious choice would be to replace Pb^{2+} with another atom of Group IV, such as Sn^{2+}. Indeed Sn-based perovskites were shown to exhibit ideal band gaps for photovoltaics applications, as well as good charge carrier mobilities (Stoumpos et al. 2013). However, perovskites containing Sn^{2+} tend to degrade rapidly due to the oxidation of the cation to Sn^{4+}. This limitation was recently overcome by developing mixed Sn/Pb halide perovskites, and the first applications to tandem device architectures are very promising (McMeekin et al. 2016; Eperon et al. 2016).

   A comprehensive search for other +2 B-site cations was performed by Filip and Giustino (2016). In this work the Periodic Table was screened by means of DFT calculations, using the procedure illustrated in Fig. 7. The candidate cations include metals which exhibit a stable +2 oxidation state and which form BX_2 salts with $X = Cl$, Br, or I. This choice leads to 116 hypothetical compounds for which the electronic structure was calculated using DFT/LDA, using the same orthorhombic perovskite structure as for $CH_3NH_3PbI_3$. The calculated band structures were then used to reduce the materials space, and only those insulating compounds with a band gap smaller than 3.5 eV were retained, leaving 40 potential candidates. These compounds were tested for dynamic stability using a "shake-and-relax" approach, meaning that the structures are randomly distorted and subsequently re-optimized.
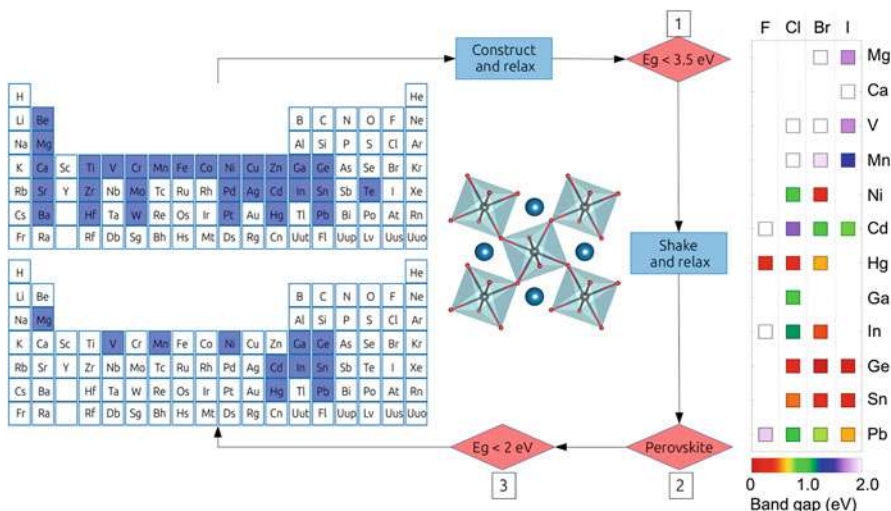
Fig. 7 Homovalent replacement of lead. Schematic illustration of the methodology and the selection criteria (in red triangles) employed for screening the Periodic Table for stable lead-free ABX_3 halide perovskites. At first, all perovskite structures with a DFT/LDA band gap larger than 3.5 eV were eliminated. The remaining compounds were randomly distorted ("shake-and-relax") to probe their dynamic stability, and only compounds which retained the perovskite structure and exhibited a band gap of less than 2.0 eV were considered for the final GW calculations. These compounds are marked with the blue squares in the lower Periodic Table. The DFT/LDA band gaps for all the halide perovskites which satisfied all selection criteria are summarized at the left side of the figure. (Adapted with permission from Filip and Giustino (2016), Copyright (2016) American Chemical Society)

Only structures that preserve the perovskite connectivity are retained in this step. For the 32 compounds that pass this test, more refined calculations including spin-orbit coupling effects were performed, and 25 hypothetical perovskites with band gaps smaller than 2 eV were identified. For the compounds with the most promising electronic properties, Filip and Giustino (2016) performed additional calculations at the GW level. However in all the cases considered, the quasiparticle band gaps were too wide for solar cell applications. This result suggests that homovalent Pb replacement may not be the best option for eliminating Pb in halide perovskites. A possible promising exception is the use of Sn^{2+}, provided the stability issues can be resolved. Similar conclusions were reached by an extensive, high-throughput search of novel ABX_3 perovskites by Korbel et al. (2016).

2.2   Heterovalent Pb Replacement

In the previous section, we considered the possibility of replacing Pb with another divalent cation. In order to broaden the search space, several authors considered the potential replacement of Pb^{2+} by cations in their +3 or +4 formal oxidation

states. This choice leads to the formation of perovskite-related structures such as $Cs_3Sb(III)_2I_9$ and $Cs_3Bi_2I_9$ (Saparov et al. 2015; Park et al. 2015), $Cs_2Sn(IV)I_6$, and Cs_2PdBr_6 (Lee et al. 2014; Sakai et al. 2017). In all these structures, the three-dimensional corner-sharing connectivity of the octahedra is disrupted. In particular, A_2BX_6 compounds can be described as vacancy-ordered double perovskites, where the B-sites of the perovskite lattice are replaced alternatively by the $+4$ cation and by a vacancy. In this structure the BX_6 octahedra are structurally disconnected. Similarly $A_3B_2X_9$ compounds can be thought of as if obtained from the perovskite lattice by removing one layer of octahedra. In both cases the resulting compounds behave electronically as low-dimensional systems (zero-dimensional and two-dimensional, respectively), and as a consequence, the optoelectronic properties are not ideal (Xiao et al. 2017b).

Another potential strategy for lead replacement is the split-cation approach. This consists of replacing pairs of Pb^{2+} cations by pairs of $+1$ and $+3$ cations, so as to preserve the average oxidation state on the B-site. When the $+1$ and $+3$ cations are arranged in a rock-salt sublattice, the structure is an ordered double perovskite, also known as elpasolite, as shown in Fig. 8. In this case the three-dimensional connectivity of the octahedra is maintained.

Double perovskites are well known in the literature on oxide perovskites (Vasala and Karppinen 2015). Halide double perovskites are less known, but are commonly employed in radiation detectors as scintillators (van Loef et al. 2002; Biswas and Du 2012). Common elpasolites exhibit wide band gaps; therefore they are unsuitable for photovoltaics applications. In the following three sections, we review recent findings on three new classes of halide elpasolites: halide double perovskites based on pnictogens and noble metals (Volonakis et al. 2016; Filip et al. 2016; Slavney et al. 2016; McClure et al. 2016); indium-silver halide double perovskites (Volonakis et al. 2017a; Zhao et al. 2017a); and indium-bismuth double perovskites (Volonakis et al. 2017b; Xiao et al. 2017a; Zhao et al. 2017b).
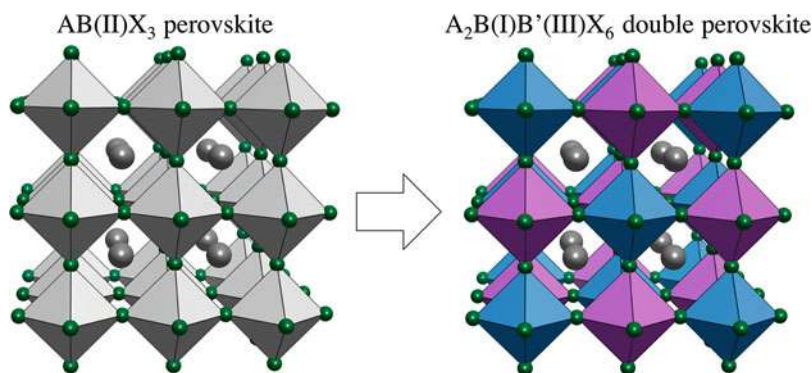
Fig. 8 The double perovskite structure. Two divalent cations at the B-site of the $AB(II)X_3$ perovskite structure (left) are replaced with a monovalent and a trivalent cation to form the double perovskite $A_2B(I)B'(III)X_6$ structure (right). The two types of octahedra in the double perovskite remain corner-sharing. When the B(I) and B′(III) cations are ordered in a rock-salt sublattice, these compounds are called elpasolites (Morss et al. 1970; Meyer and Gaebell 1978; Meyer 1980)

2.2.1   Double Perovskites Based on Pnictogens and Noble Metals

In order to design new lead-free halide double perovskites, Volonakis et al. (2016) started from the known elpasolite $Cs_2NaBiCl_6$ (Morss and Robinson 1972) and considered the replacement of monovalent Na by a noble metal. The rationale behind this strategy is that noble metals are known to be good electrical conductors due to their s^1 electrons. Volonakis et al. (2016) performed DFT calculations for hypothetical compounds Cs_2NmPnX_6, where Nm = Cu, Ag, Au, Pn = Bi, Sb, and X = Cl, Br, I. The calculated band gaps and carrier effective masses for the Bi-based hypothetical perovskites are shown in Fig. 9. All band gaps fall below 2.7 eV, and the effective masses do not exceed $0.4\,m_e$. These results indicate that pnictogen/noble metal halide double perovskites might indeed be promising for optoelectronic applications.

Volonakis et al. (2016) reported the synthesis of $Cs_2AgBiCl_6$. The compound was synthesized by following the same route as for $Cs_2NaBiCl_6$ (Morss and Robinson 1972) and replacing AgCl for NaCl. The structure of this compound was characterized by powder and single crystal X-ray diffraction (XRD) measurements. The authors obtained an ordered double perovskite structure with the $Fm\overline{3}m$ face-centered cubic space group (no. 255) at room temperature. The reported lattice
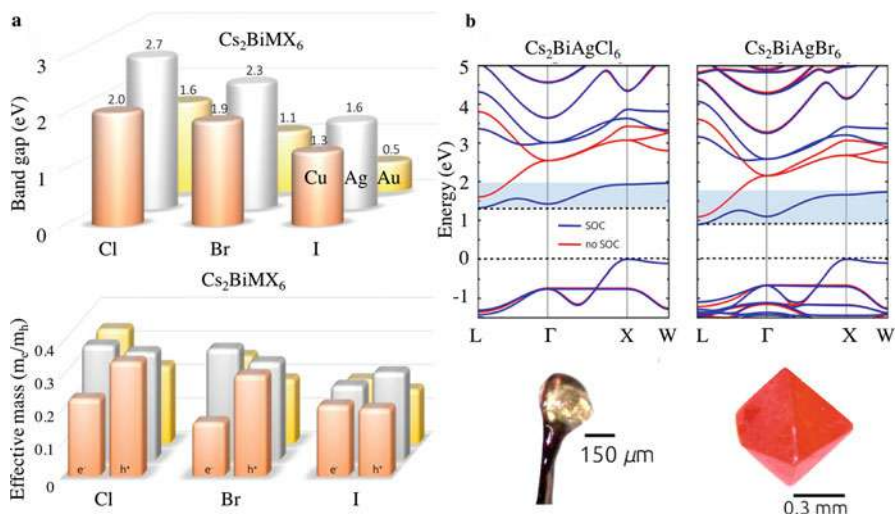
Fig. 9 Electronic properties of the pnictogen/noble metal double perovskites. (**a**) Calculated electronic band gaps (top left) and effective masses (bottom left) of all the hypothetical $Cs_2BiB'X_6$ lead-free halide double perovskites, with B' = Cu, Ag, Au and X = Cl, Br, I. All the band gaps are calculated by employing the PBE0 hybrid functional and are found to be indirect and below 2.7 eV. The effective masses are low ($<0.4\,m_e$), as calculated from the DFT/LDA at the valence band top and the conduction band bottom for the holes and electrons, respectively. (**b**) The DFT/LDA electronic band structure for the synthesized $Cs_2AgBiCl_6$ and $Cs_2AgBiBr_6$ compounds. Calculations with (without) spin-orbit coupling are shown in blue (red), and the shaded area highlights the lowest unoccupied band. The synthesized $Cs_2AgBiCl_6$ and $Cs_2AgBiBr_6$ single crystals are shown below. (Adapted with permission from Volonakis et al. (2016) and Filip et al. (2016), Copyright (2016) American Chemical Society)

constant is 10.78 Å, in good agreement with the predicted DFT/LDA value of 10.50 Å. Following the same synthetic route, Filip et al. (2016) reported the new double perovskites $Cs_2AgBiBr_6$. Also in this case, the XRD measurements confirmed the formation of an elpasolite structure, with a lattice constant of 11.26 Å. This value is slightly larger than for $Cs_2AgBiCl_6$, as expected based on the ionic radii of the halides. The synthesis of these compounds via solution and via solid-state synthesis was also reported independently (and a few weeks earlier) by two other groups (Slavney et al. 2016; McClure et al. 2016).

The electronic band structures of the new double perovskites are shown in Fig. 9. In both cases the band gap is indirect, with the valence band top at the X point of the Brillouin zone and the conduction band bottom at the L point. The band gaps calculated using the PBE0 hybrid functional are 2.7 and 2.3 eV for $Cs_2AgBiCl_6$ and $Cs_2AgBiBr_6$, respectively (Volonakis et al. 2016). Quasiparticle GW calculations yield slightly smaller gaps, namely, 2.4 and 1.8 eV, respectively (Filip et al. 2016). The calculated band gaps are broadly in agreement with the measured optical gaps, which range between 2.2–2.8 eV and 1.9–2.2 eV for $Cs_2AgBiCl_6$ and $Cs_2AgBiBr_6$, respectively (Filip et al. 2016).

More recently Filip et al. (2017) performed a systematic DFT investigation of the thermodynamic stability of the entire family of Cs_2NmPnX_6 compounds. The calculations involve the comparison between the total energy of each perovskite and the total energies of all possible decomposition products. This study predicted that only three double perovskites should be stable, namely, $Cs_2AgBiCl_6$, $Cs_2AgBiBr_6$, $Cs_2AgSbCl_6$, in line with experimental observations (Slavney et al. 2016; McClure et al. 2016; Volonakis et al. 2016; Tran et al. 2017).

Among these new compounds, the double perovskite $Cs_2AgBiBr_6$ attracted considerable interest as a potential new material for photovoltaics and optoelectronic applications. In fact, solar cells using $Cs_2AgBiBr_6$ as the active layer were demonstrated by Greul et al. (2017), and power conversion efficiencies of 2.5% were reported. Furthermore, the possibility of mixing $Cs_2AgBiBr_6$ with $CH_3NH_3PbI_3$ in order to reduce the Pb content of perovskite solar cells (as opposed to completely replace Pb) is also under investigation (Du et al. 2017a, b). Finally, Pan et al. (2017) showed that single crystals of $Cs_2AgBiBr_6$ can be employed to realize X-ray detectors with low detection thresholds (Pan et al. 2017).

2.2.2 Double Perovskites Based on In(III) and Ag(I)

Following up on the successful design and synthesis of lead-free halide double perovskites based on pnictogen/noble metals combinations, several groups set to improve the design by targeting double perovskites with a direct band gap. Volonakis et al. (2017a) reasoned that the origin of the indirect band gap of $Cs_2AgBiBr_6$ lies in the hybridization between Bi s-states and Ag d-states near the valence band top. In order to demonstrate this effect, the authors artificially lowered the energy of the Bi s-states in $Cs_2AgBiBr_6$ using an effective Hubbard potential in the DFT calculations. This strategy led to the appearance of a direct band gap at the zone center, in line with expectations. In order to exploit this finding, Volonakis

et al. (2017a) proposed to replace Bi with a +3 cation with the outermost *s*-shell unoccupied. This proposal, combined with the observation that the synthetic route to prepare $Cs_2InNaCl_6$ via solution was already known (Morss and Robinson 1972; Meyer and Gaebell 1978), motivated the consideration of hypothetical compounds of the type Cs_2InAgX_6 with X = Cl or Br (Volonakis et al. 2017a).

The calculated band structure of $Cs_2InAgCl_6$ is shown in Fig. 10a. As expected the system exhibits a direct gap at the Γ point, and the top of the valence band is comprised of Ag and In *d*-orbitals, hybridized with the halogen *p*-orbitals. The band gap of $Cs_2InAgCl_6$ was predicted to be in the range 2.7±0.6 eV, with the error bar reflecting the spread in band gaps obtained from calculations based on different exchange and correlation functionals. This compound was successfully synthesized via an acidic solution synthesis route by mixing stoichiometric amounts of $InCl_3$, AgCl, and CsCl. Powder XRD measurements confirmed the formation of a double perovskite with a face-centered cubic crystal ($Fm\overline{3}m$ space group) and a lattice constant of 10.47 Å, in good agreement with the DFT predictions (Volonakis et al. 2017a).

The nature of In/Ag cation ordering in $Cs_2InAgCl_6$ is not firmly established yet. The characteristic fingerprints of cation ordering in the XRD spectra of double perovskites are two doublets arising from reflections at the planes containing each cation (Fig. 10b). In contrast to the cases of $Cs_2AgBiCl_6$ and $Cs_2AgBiBr_6$, where such doublets have clearly been identified (Volonakis et al. 2016), the assignment proved more difficult for $Cs_2InAgCl_6$. Volonakis et al. (2017a) calculated the XRD patterns for structures with ordered or disordered cations and compared these results to measured spectra, as shown in Fig. 10b. From this comparison it is clear that the difference between the two structures is too subtle to be resolved by the measurements. More refined XRD data from Zhou et al. (2017), taken on single crystal samples, show a clear doublet at low angle. This supports the notion that $Cs_2InAgCl_6$ is a fully ordered elpasolite, although further investigations would be desirable to settle this question.

The calculated electron and hole effective masses of $Cs_2InAgCl_6$ are relatively light, $0.29\,m_e$ and $0.28\,m_e$, respectively. These values refer to the parabolic bands that can be seen in Fig. 10a. The non-dispersive band that is seen along the ΓX direction in the same figure arises from two-dimensional wave functions comprising of In $4d_{x^2-y^2}$ states and Cl $3p_{x,y}$-states. These states are expected to hinder hole transport along the six equivalent [001] directions, as well as to favor the formation of deep traps.

The nature of the band gap of $Cs_2InAgCl_6$ remains an open question. The absorption data reported by Volonakis et al. (2017a) indicate an absorption onset around 3.3 eV, which is consistent with the upper range obtained from DFT calculations. However the photoluminescence data by Volonakis et al. (2017a) also indicate emission near 2.0 eV, indicating the presence of optically active defects within the gap. Luo et al. (2017) proposed an alternative explanation for these effects, involving parity-forbidden band-to-band transitions.

While the band gap of $Cs_2InAgCl_6$ is too wide for applications in photovoltaics, this new double perovskites is attracting attention for applications in UV detectors
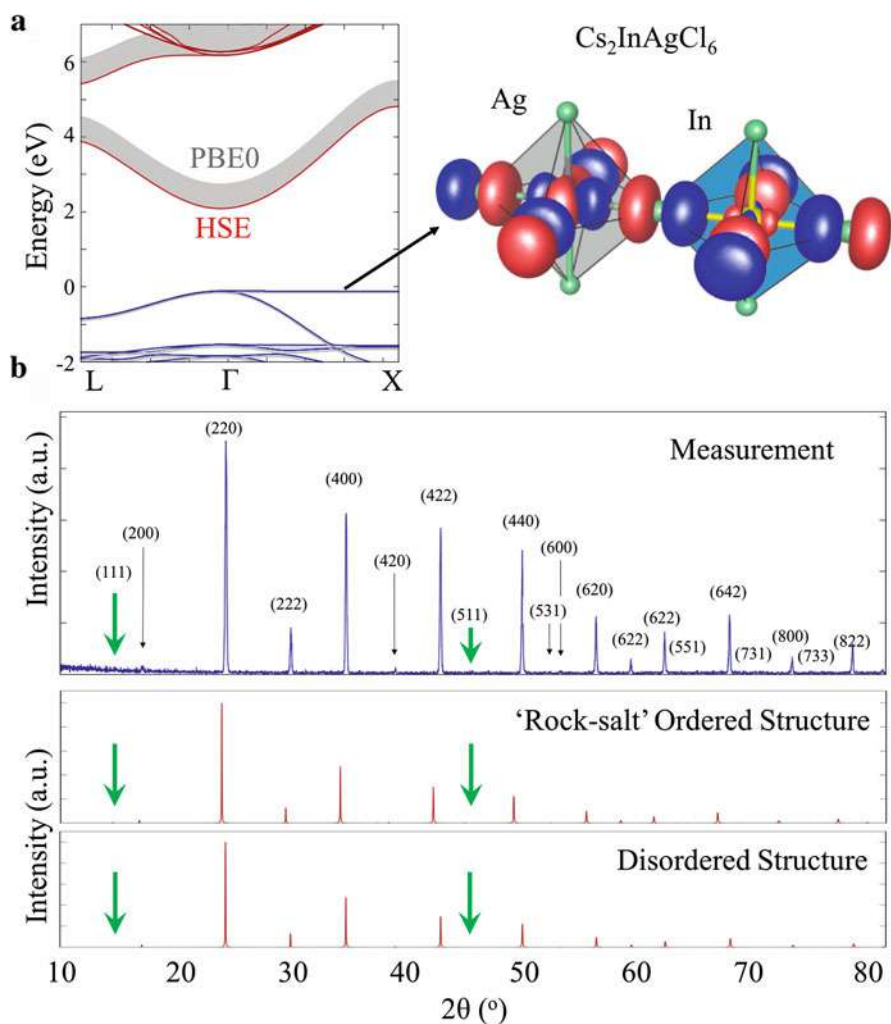
Fig. 10 Electronic and structural properties of $Cs_2InAgCl_6$. (**a**) Electronic band structure of $Cs_2InAgCl_6$ within the DFT/PBE0 and the DFT/HSE hybrid functionals. Square moduli of the electronic wave functions for the highest (top) and second highest (bottom) occupied state along the ΓX direction. (**b**) Measured powder XRD pattern of $Cs_2InAgCl_6$ (top), XRD pattern calculated by assuming ideal rock-salt ordering of the In and Ag cations (middle), and XRD pattern calculated by assuming a fully disordered cation sublattice (bottom). The green arrows show the fingerprints of the cation ordering. The difference between the calculated ordered and disordered patterns is within the noise level of the measured spectrum; therefore it is not possible to unambiguously confirm cation ordering in this case. (Adapted with permission from Volonakis et al. (2017a), Copyright (2017) American Chemical Society)

due to its excellent stability and nontoxicity. In particular Luo et al. (2017) and Zhou et al. (2017) succeeded in growing mm-sized single crystals, and Luo et al. (2017) fabricated UV detectors with high on/off ratios, fast photoresponse, low dark current, and high sensitivity.

2.2.3 Double Perovskites Based on Bi(III) and In(I)

The new double perovskites described in the preceding sections exhibit either indirect band gaps or non-dispersive bands at the band extrema. In order to move a step closer to realizing high-quality semiconducting halide perovskites without lead, Volonakis et al. (2017b) revisited the general design principles of the split-cation approach based on +1 and +3 cations.

The authors argued that an ideal double perovskite should contain cations with the same valence configuration as Pb^{2+}. This implies that candidate elements should have (i) occupied s-orbitals, (ii) unoccupied p-orbitals, and (iii) a filled d shell far from the band edge. The candidates for the monovalent cation are limited to the alkali metals, the noble metals, and the Group III elements (boron group). Among these options only the latter have occupied s-orbitals and unoccupied p-orbitals in their formal +1 oxidation state. However, only two members of this group, In and Tl, are stable as +1 cations, and Tl is to be excluded due to its toxicity. This leaves In^+ as the sole candidate for the monovalent B-site cation. For the trivalent B-site cation, the choice is between pnictogens; some transition metals such as Sc, Y, Cu, and Au that were previously reported to form elpasolites (Giustino and Snaith 2016); and elements of the boron group in the +3 oxidation state, lanthanides and actinides. Among these elements only the pnictogens satisfy the criteria (i)–(iii) above. In this group N and P are too small to coordinate six halogens in an octahedral arrangement, and As is toxic; therefore the only options left are Sb^{3+} and Bi^{3+}. Based on this simple reasoning, Volonakis et al. (2017b) concluded that the best option for replacing Pb, from the electronic structure viewpoint, is to consider In(I)/Sb(III) and In(I)/Bi(III) double perovskites.

Building on this insight, Volonakis et al. (2017b) evaluated the stability of the hypothetical compounds A_2InPnX_6 with A = K, Rb, Cs, Pn = Sb, Bi, and X = F, Cl, Br, I, using both the Goldschmidt tolerance analysis and DFT calculations of decomposition energies in conjunction with the Materials Project Database (Jain et al. 2013). The outcome of this study was that all compounds are unstable with respect to decomposition, in line with the work of Xiao et al. (2017a). However $Cs_2InBiBr_6$ was found to be only marginally unstable, as the calculated decomposition energy of 1 meV per formula unit falls within the uncertainty of the computational method. This finding raises the hope that $Cs_2InBiBr_6$ or a closely related compound might be amenable to synthesis. Volonakis et al. (2017b) also noted that the decomposition energy correlates with the size of the A-site cation and that double perovskites with larger A-site cations tend to be less prone to decomposition. Based on this observation, the authors proposed that In(I)/Bi(III) halide double perovskites might be amenable for synthesis by using large organic cations such as $CH_3NH_3^+$ and $CH(NH_2)_2^+$.
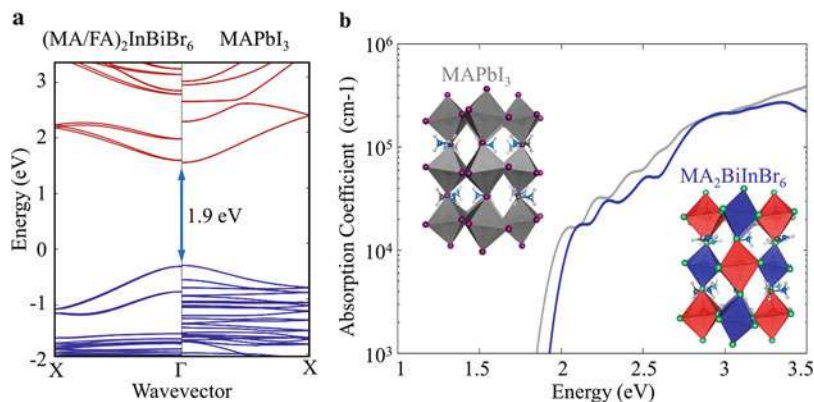
Fig. 11 Indium/Bismuth halide double perovskites. (**a**) Comparison between the electronic band structures of the hypothetical lead-free double perovskite $(MA/FA)_2InBiBr_6$ and the lead halide perovskite $MAPbI_3$. MA stands for methylammonium (CH_3NH_3) and FA for formamidinium [$CH(NH_2)_2$]. Calculations were performed by employing the PBE0 hybrid functional. The double perovskite structures were fully optimized starting from the low-temperature *Pnma* phase of $MAPbI_3$ and by replacing two Pb atoms with In and Bi. (**b**) Calculated absorption coefficients for $MA_2InBiBr_6$ (blue line) and $MAPbI_3$ (grey line). (Adapted with permission from Volonakis et al. (2017b), Copyright (2017) American Chemical Society)

The electronic band structure of the hypothetical $(CH_3NH_3/CH(NH_2)_2)_2InBiBr_6$ double perovskites along with the band structure of $CH_3NH_3PbI_3$ is shown in Fig. 11a. As anticipated, owing to the similar electronic configuration of Pb^{2+}, In^+ and Bi^{3+}, the two compounds exhibit very similar band structures. The band gaps are direct in both cases and of the same magnitude. Similarly the calculated optical absorption spectra for these compounds are very similar, as shown in Fig. 11b. These findings highlight the potential of hybrid In/Bi halide double perovskites for photovoltaics and optoelectronics.

At the time of writing this chapter, synthetic attempts at making In(I)/Bi(III) halide double perovskites have not been successful. The key challenge appears to be the tendency of In(I) to be oxidized to the more stable form In(III); therefore future synthesis studies should devise strategies for maintaining In in its +1 state long enough to enable its incorporation in the double perovskite lattice.

2.3 Conclusions

In this chapter we have presented a detailed account of the contributions made by our materials modelling group to the study of the fundamental properties and design principles of metal-halide perovskites. The results reviewed in this chapter were obtained by employing techniques ranging from standard density functional theory and density functional perturbation theory to many-body perturbation theory treatments of electronic excitations and electron-phonon coupling phenomena. Our studies so far have been directed at two principal goals: understanding the

fundamental properties of lead-halide perovskites and developing design principles to guide the discovery of novel lead-free metal-halide perovskite semiconductors.

The current understanding of the fundamental properties of lead-halide perovskites is due to a complementary effort combining state-of-the art predictive computational modelling and experimental measurements which have allowed us to rationalize the optoelectronic properties of these materials from the perspective of atomistic modelling. In addition, we have shown that the predictive power of computational modelling can be directed at designing new lead-free halide perovskites, which led most notably to the computational discovery of a new family of lead-free halide double perovskites. During the last year more than eight novel lead-free double perovskite compounds have been successfully synthesized (Volonakis et al. 2016, 2017a; Slavney et al. 2016; McClure et al. 2016; Filip et al. 2016; Tran et al. 2017; Wei et al. 2016; Deng et al. 2016; Wei et al. 2017). Among these, compounds such as $Cs_2AgBiCl_6$, $Cs_2AgBiBr_6$, $Cs_2InAgCl_6$, and $Cs_2AgSbCl_6$ were first proposed by us in silico.

Despite the tremendous popularity of halide perovskites, there are still fundamental open questions and technological challenges that remain unresolved. A non-exhaustive list includes the fundamental understanding of charge transport and excitonic properties, elucidating the physics of defects and their impact on the degradation of these materials under standard solar cell operating conditions, and developing a practical solution for reducing the Pb content without impacting the device performance. All of these problems are currently active areas for both experimental and theoretical research.

In a broader context, the rapid rise of hybrid halide perovskites can be seen as a true success story for experimental and theoretical materials design. As the race for ever higher-performing and cost-efficient optoelectronic devices continues, there is an increasing demand for the discovery of new functional materials. Thanks to the continued development of highly accurate ab initio methods, computational materials modelling can now directly guide discovery, by designing materials and predicting their properties in silico ahead of synthesis. These developments place ab initio modelling in a central role in modern materials science and engineering, as an accelerator for the discovery and deployment of new technologies.

References

Baikie T, Fang Y, Kadro JM, Schreyer M, Wei F, Mhaisalkar SG, Grätzel M, White TJ (2013) Synthesis and crystal chemistry of the hybrid perovskite ($CH_3NH_3PbI_3$) for solid-state sensitized solar applications. J Chem Mater A 1:5628

Ball JM, Petrozza A (2017) Defects in perovskite-halides and their effects in solar cells. Nat Energ 1(16149)

Biswas K, Du MH (2012) Energy transport and scintillation of cerium-doped elpasolite Cs_2LiYCl_6: hybrid density functional calculations. Phys Rev B 86(1):014102–014109

Brenner TM, Egger DA, Rappe AM, Kronik L, Hodes G, Cahen D (2015) Are mobilities in hybrid organic-inorganic halide perovskites actually high? J Phys Chem Lett 6:4754–4757

Brivio F, Bulter KT, Walsh A, van Schilfgaarde M (2014) Relativistic quasiparticle self-consistent electronic structure of hybrid halide perovskite photovoltaic absorbers. Phys Rev B 89: 155204

Carignano MA, Kachmar A (2015) Thermal effects on $CH_3NH_3PbI_3$ perovskite from ab initio molecular dynamics simulations. J Phys Chem C 119(17):8991–8997

Chen YS, Manser JS, Kamat PV (2015) All solution-processed lead halide perovskite-$BiVO_4$ tandem assembly for photolytic solar fuels production. J Am Chem Soc 137(2):974–981

Comin R, Walters G, Thibau ES, Voznyy O, Lu ZH, Sargent EH (2015) Structural, optical, and electronic studies of wide-bandgap lead halide perovskites. J Mater Chem C 3:8839–8843

Davies CL, Filip MR, Patel JB, Crothers TW, Verdi C, Wright AD, Milot RL, Giustino F, Johnston MB, Herz LM (2018) Bimolecular recombination in methylammonium lead triiodide perovskite: an inverse absorption process. Nat Commun 9:293

De Wolf S, Holovsky J, Moon SJ, Löper P, Niesen B, Ledinsky M, Haug FJ, Yum JH, Ballif C (2014) Organometallic halide perovskites: sharp optical absorption edge and its relation to photovoltaic performance. J Phys Chem Lett 5:1035–1039

Deng Z, Wei F, Sun S, Kieslich G, Cheetham AK, Bristowe PD (2016) Exploring the properties of lead-free hybrid double perovskites using a combined computational-experimental approach. J Mater Chem A 4(31):12025–12029

D'Innocenzo V, Grancini G, Alcocer MJP, Kandada ARS, Stranks SD, Lee MM, Lanzani G, Snaith HJ, Petrozza A (2014) Excitons versus free charges in organo-lead tri-halide perovskites. Nat Commun 5:3586

Družbicki K, Pinna RS, Rudić S, Jura M, Gorini G, Fernandez-Alonso F (2016) Unexpected cation dynamics in the low-temperature phase of methylammonium lead iodide: the need for improved models. J Phys Chem Lett 7(22):4701–4709

Du KZ, Meng W, Wang X, Yan Y, Mitzi DB (2017a) Bandgap engineering of lead-free double perovskite $Cs_2AgBiBr_6$ through trivalent metal alloying. Angew Chem Int Ed 56(28): 8158–8162

Du KZ, Wang X, Han Q, Yan Y, Mitzi DB (2017b) Heterovalent B-site co-alloying approach for halide perovskite bandgap engineering. ACS Energ Lett 2:2486–2490

Elliott RJ (1957) Intensity of the optical absorption by excitons. Phys Rev 108:1384

Eperon GE, Stranks SD, Menelaou C, Johnston MB, Herz LM, Snaith HJ (2014) Formamidinium lead trihalide: a broadly tunable perovskite for efficient planar heterojunction solar cells. Energ Environ Sci 7:982–988

Eperon GE, Leijtens T, Bush KA, Prasanna R, Green T, Jacob Tse-Wei Wang JTW, McMeekin DP, Volonakis G, Milot RL, May R, Palmstrom A, Slotcavage DJ, Belisle RA, Patel JB, Parrott ES, Sutton RJ, Ma W, Moghadam F, Conings B, Babayigit A, Boyen HG, Bent S, Giustino F, Herz LM, Johnston MB, McGehee MD, J SH (2016) Perovskite-perovskite tandem photovoltaics with optimized bandgaps. Science 354:861–865

Even J, Pedesseau L, Jancu JM, Katan C (2013) Importance of spin-orbit coupling in hybrid organic/inorganic perovskites for photovoltaic applications. J Phys Chem Lett 119: 10161–10177

Filip MR, Giustino F (2014) GW quasiparticle band gap of the hybrid organic-inorganic perovskite $CH_3NH_3PbI_3$. Phys Rev B 90:245145

Filip MR, Giustino F (2016) Computational screening of homovalent lead substitution in organic–inorganic halide perovskites. J Phys Chem C 120:166–173

Filip MR, Eperon GE, Snaith HJ, Giustino F (2014) Steric engineering of metal-halide perovskites with tunable optical band gaps. Nat Commun 5:5757

Filip MR, Verdi C, Giustino F (2015) GW band structure and carrier effective mases of $CH_3NH_3PbI_3$ and hypothetical perovskites of the type $APbI_3$: A = NH_4, PH_4, AsH_4 and SbH_4. J Phys Chem C 119:24209–24219

Filip MR, Hillman S, Haghighirad AA, Snaith HJ, Giustino F (2016) Band gaps of the lead-free halide double perovskites $Cs_2BiAgCl_6$ and $Cs_2BiAgBr_6$ from theory and experiment. J Phys Chem Lett 7:2579–2585

Filip MR, Liu X, Miglio A, Hautier G, Giustino F (2017) Phase diagrams and stability of lead-free halide double perovskites $Cs_2BB'X_6$. J Phys Chem C https://doi.org/10.1021/acs.jpcc.7b10370

Giustino F, Snaith HJ (2016) Toward lead-free perovskite solar cells. ACS Energy Lett 1(6): 1233–1240

Greul E, Petrus ML, Binek A, Docampo P, Bein T (2017) Highly stable, phase pure $Cs_2AgBiBr_6$ double perovskite thin films for optoelectronic applications. J Mater Chem A 5(37): 19972–19981

Hao F, Stoumpos CC, Chang RPH, Kanatzidis MG (2014) Anomalous band gap behavior in mixed sn and pb perovskites enables broadening of absorption spectrum in solar cells. J Am Chem Soc 136(22):8094–8099

Hedin L (1965) New method for calculating the one-particle green's function with application to the electron-gas problem. Phys Rev 139:A796

Herz LM (2016) Charge-carrier dynamics in organic-inorganic metal halide perovskites. Annu Rev Phys Chem 67:3.1–3.25

Hirasawa M, Ishihara T, Goto T, Uchida N Kand Miura (1994) Magnetoabsorption of the lowest exciton in perovskite-type compound ($CH_3NH_3PbI_3$). Phys B 201:427–430

Hybertsen MS, Louie SG (1986) Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. Phys Rev B 34:5390

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) The materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:011002

Johnston MB, Herz LM (2016) Hybrid perovskites for photovoltaics: charge-carrier recombination, diffusion, and radiative efficiencies. Acc Chem Res 49(1):146–154

Kioupakis E, Zhang P, Cohen ML, Louie SG (2008) *GW* quasiparticle corrections to the LDA+U/GGA+U electronic structure of bcc hydrogen. Phys Rev B 77:155114

Korbel S, Marques MAL, Botti S (2016) Stability and electronic properties of new inorganic perovskites from high-throughput ab initio calculations. J Mater Chem C 4(15):3157–3167

Lahnsteiner J, Kresse G, Kumar A, Sarma DD, Franchini C, Bokdam M (2016) Room-temperature dynamic correlation between methylammonium molecules in lead-iodine based perovskites: an ab initio molecular dynamics perspective. Phys Rev B 94:214114

Ledinský M, Löper P, Niesen B, Holovský J, Moon SJ, Yum JH, De Wolf S, Fejfar A, Ballif C (2015) Raman spectroscopy of organic inorganic halide perovskites. J Phys Chem Lett 6: 401–406

Lee B, Stoumpos CC, Zhou N, Hao F, Malliakas C, Yeh C-Y, Marks TJ, Kanatzidis MG, Chang RPH (2014) Air-stable molecular semiconducting iodosalts for solar cell applications: Cs2sni6 as a hole conductor. J Am Chem Soc 136(43):15379–15385

Lee MM, Teuscher J, Miyasaka T, Murakami TN, Snaith HJ (2012) Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites. Science 338:643–647

Leijtens T, Bush K, Cheacharoen R, Beal R, Bowring A, McGehee MD (2017) Towards enabling stable lead halide perovskite solar cells; interplay between structural, environmental, and thermal stability. J Mater Chem A 5:11483

van Loef EVD, Dorenbos P, van Eijk CWE, Krämer KW, Güdel HU (2002) Scintillation and spectroscopy of the pure and Ce^{3+}-doped elpasolites: Cs_2LiYX_6 (X = Cl, Br). Journal of Physics: Condensed Matter 14(36):8481

Luo J, Li S, Wu H, Zhou Y, Li Y, Liu J, Li J, Li K, Yi F, Niu G, Tang J (2017) $Cs_2AgInCl_6$ double perovskite single crystals: Parity forbidden transitions and their application for sensitive and fast UV photodetectors. ACS Photonics https://doi.org/10.1021/acsphotonics.7b00837

Marzari N, Vanderbilt D (1997) Maximally localized generalized Wannier functions for composite energy bands. Phys Rev B 56:12,847

McClure ET, Ball MR, Windl W, Woodward PM (2016) Cs_2AgBiX_6 (X = Br, Cl) - new visible light absorbing, lead-free halide perovskite semiconductors. Chem Mater 28:1348–1354

McMeekin DP, Sadoughi G, Rehman W, Eperon GE, Saliba M, Hörantner MT, Haghighirad AA, Sakai N, Korte L, Rech B, Johnston MB, Herz LM, Snaith HJ (2016) A mixed-cation lead mixed-halide perovskite absorber for tandem solar cells. Science 8:151–155

Megaw H (1973) Crystal Structures. A working approach. W. B. Saunders Company, Philadelphia, London, Toronto

Menéndez-Proupin E, Palacios P, Wahnón P, Conesa JC (2014) Self-consistent relativistic band structure of the $CH_3NH_3PbI_3$ perovskite. Phys Rev B 90:045,207

Meyer G (1980) Halo-elpasolites. VI. the first iodo-elpasolites, $Cs_2B^IM^{III}I_6$ (B^I = Li, Na). Z Naturforsch B 35:268–276

Meyer G, Gaebell H (1978) Halo-elpasolites. iv. on bromo-elpasolites $Cs_2B^IM^{III}Br_6$, B^I = Li, Na; M^{III} = Sc, Y, La–Lu, In, V, Cr). Z Naturforsch B 33(12):1476–1478

Milot RL, Eperon GE, Snaith HJ, Johnston MB, Herz LM (2015) Temperature-dependent charge-carrier dynamics in $CH_3NH_3PbI_3$ perovskite thin films. Adv Func Mater 25:6218–6227

Miyata A, Mitioglu A, Plochocka P, Portugall O, Wang JTW, Stranks SD, Snaith HJ, Nicholas RJ (2015) Direct measurement of the exciton binding energy and effective masses for charge carriers in an organic-inorganic tri-halide perovskite. Nat Phys 11:582–587

Morss LR, Robinson WR (1972) Crystal structure of $Cs_2BiNaCl_6$. Acta Cryst B 28:653–654

Morss LR, Siegal M, Stenger L, Edelstein N (1970) Preparation of cubic chloro complex compounds of trivalent metals: Cs_2NaMCl_6. Inorg Chem 9(7):1771–1775

Mosconi E, Quarti C, Ivanovska T, Ruani G, F DA (2014) Structural and electronic properties of organo-halide lead perovskites: a combined IR-spectroscopy and ab initio molecular dynamics investigation. Phys Chem Chem Phys 16:16137–16144

Motta C, El-Mellouhi F, Kais S, Tabet N, Alharbi F, Sanvito S (2015) Revealing the role of organic cations in hybrid halide perovskite $CH_3NH_3PbI_3$. Nat Commun 6(7026):1–7

Noh JH, Im SH, Heo JH, Mandal TN, Seok SI (2013) Chemical management for colorful, efficient, and stable inorganic–organic hybrid nanostructured solar cells. Nano Lett 13(4):1764–1769

NREL (2017) National renewable energy laboratory: best research-cell efficiencies.

Pan W, Wu H, Luo J, Deng Z, Ge C, Chen C, Jiang X, Yin WJ, Niu G, Zhu L, Yin L, Zhou Y, Xie Q, Ke X, Sui M, Tang J (2017) $Cs_2AgBiBr_6$ single-crystal X-ray detectors with a low detection limit. Nat Photon 11(11):726–732

Park BW, Philippe B, Zhang X, Rensmo H, Boschloo G, Johansson EMJ (2015) Bismuth based hybrid perovskites $A_3Bi_2I_9$ (A: methylammonium or cesium) for solar cell application. Adv Mater 27(43):6806–6813

Pérez-Osorio MA, Milot RL, Filip MR, Patel JB, Herz LM, Johnston MB, Giustino F (2015) Vibrational properties of the organic-inorganic halide perovskite $CH_3NH_3PbI_3$ from theory and experiment: factor group analysis, first principles calculations, and low-temperature infrared spectra. J Phys Chem C 119:25703–25718

Pérez-Osorio MA, Champagne A, Zacharias M, Rignanese GM, Giustino F (2017) Van der Waals interactions and anharmonicity in the lattice vibrations, dielectric constants, effective charges, and infrared spectra of the organic-inorganic halide perovskite $CH_3NH_3PbI_3$. J Phys Chem C 121:18459–18471

Poglitsch A, Weber D (1987) Dynamic disorder in methylammoniumhalogenoplumbates (II) observed by millimeter-wave spectroscopy. J Chem Phys 87:6373

van Roosbroeck W, Shockley W (1954) Photon-radiative recombination of electrons and holes in germanium. Phys Rev 94:1558–1560

Sakai N, Haghighirad AA, Filip MR, Nayak PK, Nayak S, Ramadan A, Wang Z, Giustino F, Snaith HJ (2017) Solution-processed cesium hexabromopalladate(iv), Cs_2PdBr_6, for optoelectronic applications. J Am Chem Soc 139(17):6030–6033

Saliba M, Matsui T, Domanski K, Seo JY, Ummadisingu A, Zakeeruddin SM, Correa-Baena JP, Tress WR, Abate A, Hagfeldt A, Grätzel M (2016) Incorporation of rubidium cations into perovskite solar cells improves photovoltaic performance. Science 354:206–209 https://doi.org/10.1126/science.aah5557

Saparov B, Hong F, Sun JP, Duan HS, Meng W, Cameron S, Hill IG, Yan Y, Mitzi DB (2015) Thin-film preparation and characterization of $Cs_3Sb_2I_9$: a lead-free layered perovskite semiconductor. Chem Mater 27(16):5622–5632

Scherpelz P, Govoni M, Hamada I, Galli G (2016) Implementation and validation of fully relativistic GW calculations: spin-orbit coupling in molecules, nanocrystals, and solids. J Chem Thory Comput 12(8):3523–3544

Slavney AH, Hu T, Lindenberg AM, Karunadasa HI (2016) A bismuth-halide double perovskite with long carrier recombination lifetime for photovoltaic applications. J Am Chem Soc 138:2138–2141

Souza I, Marzari N, Vanderbilt D (2001) Maximally localized Wannier functions for entangled energy bands. Phys Rev B 65:035109

Stoumpos CC, Malliakas CD, Kanatzidis MG (2013) Semiconducting tin and lead iodide perovskites with organic cations: phase transitions, high mobilities, and near-infrared photo-luminescent properties. Inorg Chem 52(15):9019–9038

Stranks S, Snaith HJ (2015) Perovskites for photovoltaic and light-emitting devices. Nat Nanotechnol 10:391–402

Tan ZK, Moghaddam R, Lai ML, Docampo P, Higler R, Deschler F, Price M, Sadhanala A, Pazos LM, Credgington D, Hanush F, Bein T, Snaith HJ, Friend RH (2014) Bright light-emitting diodes based on organometal halide perovskites. Nat Nanotechnol 9:687–692

Tran TT, Panella JR, Chamorro JR, Morey JR, McQueen TM (2017) Designing indirect-direct bandgap transitions in double perovskites. Mater Horiz 4:688–693

van Schilfgaarde M, Kotani T, Faleev S (2006) Quasiparticle self-consistent GW theory. Phys Rev Lett 96:226402

Vasala S, Karppinen M (2015) $A_2B'B''O_6$ perovskites: a review. Prog Solid State Chem 43(1):1–36

Verdi C, Giustino F (2015) Fröhlich electron-phonon vertex from first principles. Phys Rev Lett 115(17):176401

Volonakis G, Filip MR, Haghighirad AA, Sakai N, Wenger B, Snaith HJ, Giustino F (2016) Lead-free halide double perovskites via heterovalent substitution of noble metals. J Phys Chem Lett 7:1254–1259

Volonakis G, Haghighirad AA, Milot RL, Sio WH, Filip MR, Wenger B, Johnston MB, Herz LM, Snaith HJ, Giustino F (2017a) $Cs_2InAgCl_6$: a new lead-free halide double perovskite with direct band gap. J Phys Chem Lett 8(4):772–778

Volonakis G, Haghighirad AA, Snaith HJ, Giustino F (2017b) Route to stable lead-free double perovskites with the electronic structure of $CH_3NH_3PbI_3$: a case for mixed-cation $[Cs/CH_3NH_3/CH(NH_2)_2]_2InBiBr_6$. J Phys Chem Lett 8:3917

Ward JW, Smith HL, Zeidell A, Diemer PJ, Baker SR, Lee H, Payne MM, Anthony JE, Guthold M, Jurchescu OD (2017) Solution-processed organic and halide perovskite transistors on hydrophobic surfaces. ACS Appl Mater Interf 9(21):18120–18126

Wehrenfennig C, Liu M, Snaith HJ, Johnston MB, Herz LM (2014) Homogeneous emission line broadening in the organo lead halide perovskite $CH_3NH_3PbI_{3x}Cl_x$. J Phys Chem Lett 5:1300–1306

Wei F, Deng Z, Sun S, Xie F, Kieslich G, Evans DM, Carpenter MA, Bristowe PD, Cheetham AK (2016) The synthesis, structure and electronic properties of a lead-free hybrid inorganic-organic double perovskite $(MA)_2KBiCl_6$ (MA = methylammonium). Mater Horiz 3(4):328–332

Wei F, Deng Z, Sun S, Zhang F, Evans DM, Kieslich G, Tominaka S, Carpenter MA, Zhang J, Bristowe PD, Cheetham AK (2017) Synthesis and properties of a lead-free hybrid double perovskite: $(CH_3NH_3)_2AgBiBr_6$. Chem Mater 29(3):1089–1094

Weller MT, Weber OJ, Henry PF, DiPumpo AM, Hansen TC (2015) Complete structure and cation orientation in the perovskite photovoltaic methylammonium lead iodide between 100 and 352 K. Chem Commun 51:4180–4183

Wright AD, Verdi C, Milot RL, Eperon GE, Pérez-Osorio MA, Snaith HJ, Giustino F, Johnston MB, Herz LM (2016) Electron-phonon coupling in hybrid lead halide perovskites. Nat Commun 7(11755):1–9

Xiao Z, Du KZ, Meng W, Wang J, Mitzi DB, Yan Y (2017a) Intrinsic instability of $Cs_2In^{(I)}M^{(III)}X_6$ (M = Bi, Sb; X = halogen) double perovskites: a combined density functional theory and experimental study. J Am Chem Soc 139(17):6054–6057

Xiao Z, Meng W, Wang J, Mitzi DB, Yan Y (2017b) Searching for promising new perovskite-based photovoltaic absorbers: the importance of electronic dimensionality. Mater Horiz 4(2):206–216

Zhao XG, Yang D, Sun Y, Li T, Zhang L, Yu L, Zunger A (2017a) Cu-In halide perovskite solar absorbers. J Am Chem Soc 139(19):6718–6725

Zhao XG, Yang JH, Fu Y, Yang D, Xu Q, Yu L, Wei SH, Zhang L (2017b) Design of lead-free inorganic halide perovskites for solar cells via cation-transmutation. J Am Chem Soc 139(7):2630–2638

Zhou J, Xia Z, Molokeev MS, Zhang X, Peng D, Liu Q (2017) Composition design, optical gap and stability investigations of lead-free halide double perovskite $Cs_2AgInCl_6$. J Mater Chem A 5(29):15031–15037

Prototyping Ultrafast Charge Separation by Means of Time-Dependent Density Functional Methods

15

Carlo Andrea Rozzi and Stefano Pittalis

Contents

Abstract

We discuss applications of time-dependent density functional theory (TDDFT) to simulate ultrafast charge separation in organic photovoltaic materials. We first outline both the targeted physics and the employed computational methods with particular regard to the real-time propagation framework. Then, we review some recent successful case studies in which TDDFT-based simulations have helped identify prototypical systems useful both for the basic understanding of the charge separation phenomena and for the systematic improvement of photovoltaic device performances.

1   Introduction

The concept of joining a p-doped and an n-doped semiconductor (typically silicon) in a single crystal is at the heart of traditional inorganic photovoltaics (IPV). In IPV, the charge carriers in a thin volume around the p-n junction diffuse at equilibrium

C. A. Rozzi (✉) · S. Pittalis
Istituto Nanoscienze – CNR, Modena, Italy
e-mail: carloandrea.rozzi@nano.cnr.it
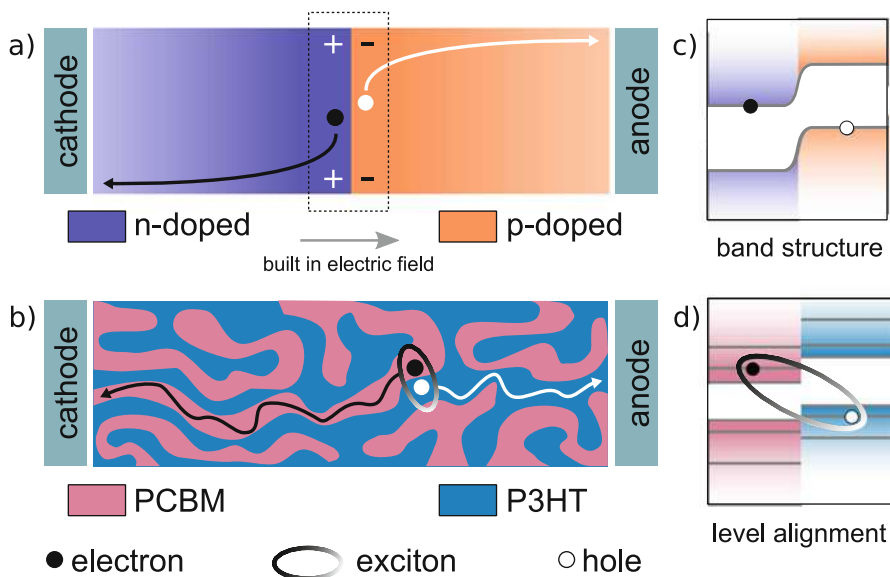
Fig. 1 Comparison between inorganic (first generation) and organic bulk heterojunction photovoltaic cells. Electrons (black), holes (white), and excitons are highlighted. (**a**) Junction between a p- and an n-doped crystalline semiconductor. At the junction, the depletion zone creates an intrinsic electric field that separates the photogenerated carriers. (**b**) Interpenetrating paths in a section of a polymer-acceptor blend (such as P3HT:PCBM). Once the interfacial excitons are dissociated, each type of carriers moves in a distinct phase. (**c**) Band structure of a crystalline p-n junction. Free carriers are directly injected in the conduction and valence bands. (**d**) Energy levels in an organic heterojunction. A strongly bound exciton is formed before the charges are separated

toward neighbor volumes following the gradient of carriers' concentration. At equilibrium, the migrated charges leave behind a depletion zone in which an intrinsic electric field balances the surface charge across the junction. The intrinsic electric field is ultimately responsible for separating electrons and holes when solar light reaches the junction (see Fig. 1a).

The operating principle of solar cells changed dramatically with the advent of the era of organic photovoltaics (OPV) based on dye-sensitized solar cells (DSC) and bulk heterojunctions (BHJ).

In DSC, organic dye molecules are employed to enhance the absorbance of an underlying nanostructured semiconductor (such as TiO_2). The dye is the photoactive material. Electrons are injected from the photoexcited dye to the TiO_2 due to the energy offset between the molecular levels and the valence band of the semiconductor. Electrons are then replaced in the dye by the action of a liquid electrolyte. The process is made unidirectional by the kinetic competition between the (fast) electron injection and the (slower) regeneration by the electrolyte.

In BHJ, the electron donors and acceptors are blended into a solid phase. No liquids appear in the final product, and the material can be printed on a flexible plastic support. Although some variants have been proposed, typically the donor

is a semiconducting polymer or small organic molecules, and the acceptor is a fullerene molecule (Sariciftci et al. 1992). Suitable chemical functionalization of both moieties is performed in order to enhance their mutual mixing and phase separation. The tight interleaving of the donor and acceptor phases increases the charge-separating interfacial area with respect to the film thickness and keeps the surface between the phases within a distance comparable to the exciton diffusion length (see Fig. 1b).

The charge separation induced by light in OPV is fundamentally different from that in IPV in several respects. In traditional p-n junctions, excitation and charge separation occur in the same material. Excitons may be equally created on both sides of the junction. Those formed within the bulk semiconductor can reach the junction if this is within the exciton diffusion length. Otherwise they will recombine with high probability.

Excitons are quite delocalized throughout the crystal lattice (Wannier-type excitons) and have binding energies smaller than the thermal energy at room temperature (of the order of 10 meV for bulk semiconductors (Green 2013). Therefore light excitation is able to directly inject free electrons into the conduction and holes into the valence band. Since the concentration gradient for the two species would drive them back toward the junction, the intrinsic field is the key factor to separate them.

In OPV, instead, two distinct phases exist, but only one of them can efficiently absorb visible light. It is the dye molecule in DSC and the polymer in BHJ. The intermolecular coupling between the components of each phase is weaker than for covalently linked units or bulk solids and the dielectric constant much lower (about 4 in organic (Li et al. 2012) instead of about 15 in bulk semiconductors (Dunlap and Watters 1953)). Therefore excitons typically have Frankel character in the bulk donor phase, i.e., they are localized and have a binding energy (of the order of tenths of an eV (Alvarado et al. 1998). An exothermic pathway to overcome this binding energy and separate the carriers is available provided that the bands of the donor and the acceptor have a suitable offset, namely, with the donor excited state higher in energy than the charge-separated state (see Fig. 1). If this is the case, an exciton can be dissociated at the interface, and the two carriers of opposite sign can be spatially dislocated: electrons move to the acceptor phase, while holes remain in the donor phase. No intrinsic field is needed to keep electrons apart from holes.

Photoexcitation produces free carriers in IPV, but a driving force is needed to avoid their recombination with the holes; in OPV a driving force is necessary to dissociate excitons, but charge carriers remain separated in the appropriate phases. Further important details, however, must be accounted for in order to obtain an accurate description of the charge separation dynamics and, eventually, to gain insight into the actual material performance for OPV. For instance, while the structure of p-n junctions is simply controlled by the crystal purity and the precision of doping, the device morphology of nanostructured blends can be quite intricate and process-dependent. Besides, since the charge separation occurs at the interface, while the carrier transport occurs in each phase separately, morphology affects both stages of the photovoltaic conversion in an interdependent way (Zhong et al. 2015; Jakowetz et al. 2016).

Most importantly, the scheme outlined above does not include the dynamics of exciton dissociation and charge separation. While in IPV the scattering of charge carriers by lattice phonons basically acts as an energy loss channel cooling down electrons toward the bottom of the conduction band, in OPV the role of nuclear motion is significantly more complex. Both low- and high-frequency modes can be strongly coupled to the photoexcited states and may take an active role in driving exciton dissociation. Even assuming that the structure is perfectly known and under control, the description of the aforementioned processes from first principles poses formidable challenges.

In the following sections, we discuss some examples of how the ultrafast regime of charge separation can be successfully analyzed by combining time-dependent density functional theory and suitable schemes for molecular dynamics. After an overview of the theoretical methods, we will review some cases in which the interplay between electronic and nuclear degrees of freedom dramatically affects the behavior of OPV materials. We will show that simple nonempirical calculations can reach a very good agreement with transient spectroscopy experiments and the corresponding simulations offer us important clarifications of the underlying mechanisms.

2 Simulations: Setting Up the Stage

The challenge of computing electronic properties at the nanoscale of even quite sizable systems can be considerably simplified by admitting a reduced description, such as that offered by density functional methods (Hohenberg and Kohn 1964; Kohn and Sham 1965; Runge and Gross 1984). In (time-dependent) density functional theory [(TD)DFT], the basic variable is the (TD) particle density: a real-valued function of the position, which represents the distribution of electrons in space. Crucially, the interacting many-body system can be mapped onto an auxiliary noninteracting many-body system: i.e., the (TD) Kohn-Sham (KS) system. The electrons of the KS system move in an effective potential of which the exchange-correlation (XC) potential is the only component to be approximated. The particle density can be obtained in terms of the contributions due to occupied single-particle orbitals, and thus, the relevant properties of the real system can be computed if they can be practically expressed as functionals of the particle density. The ground state and the excitation energies are among the typical properties which can be determined efficiently and rather accurately. For an extended account on TDDFT, the reader is referred to the contribution of Gross et al. in the Volume 1.

The first and simplest approximation in DFT is the local density approximation (LDA) (Kohn and Sham 1965). The LDA assumes that the XC energy density – and, thus, also the corresponding XC potential – can be approximated locally with that of the uniform gas. The LDA forms the base of a hierarchy of improved approximations: the "Jacobs' ladder" of DFT (Perdew and Schmidt 2001). An approximate time-dependent XC potential can be readily obtained by evaluating an approximation originally designed for ground-state DFT on the instantaneous

particle density (or, more in general, on the corresponding instantaneous KS orbitals). This is the popular adiabatic approximation of TDDFT, which, for example, when applied to the LDA, yields the adiabatic LDA (ALDA). Adiabatic TDDFT functionals neglect the time non-localities of their implicit dependency on particle density. In the Born-Oppenheimer (BO) sense, instead, adiabaticity refers to the possibility of decoupling the dynamics of the nuclear and electronic degrees of freedoms. In the BO approximation, the electronic wave function is restricted to the ground-state manifold at each given nuclear configuration. The cases reported in the remainder of this chapter will demonstrate the usefulness of ALDA when used in a scheme that goes beyond the BO approximation.

Most of the applications of TDDFT are carried out in the linear response regime. The common way of proceeding does not involve a RT propagation but the solution of a (pseudo) eigenvalue problem, from which it is possible to determine the absorption oscillation strengths and the mean polarizability using ground-state KS quantities as an input (Casida 1995). The information obtained in this way identifies the sets of KS transitions which represent the relevant excitations. In the cases presented below, for example, the solutions of the Casida's equations are used to identify the KS states that can mimic the initial excited states, as obtained at the onset of the photoexcitation, to be considered in subsequent RT propagations.

While the linear response of a system is independent of the (small) perturbation, the nonlinear response can vary both quantitatively and qualitatively with the intensity of the perturbation. Nonlinear perturbations are particularly relevant in experimental study of photoinduced charge transfer and include, for example, high-intensity laser pulses such as those employed in transient absorption and 2D spectroscopy experiments (Jonas 2003), suited to highlight emerging coherences.

The response function of the particle density can be accessed by means of real-time (RT) propagation of the TDDFT equations both in the linear and nonlinear regime (Yabana et al. 2006). In the linear response regime, the method based on RT propagation scales better than the method based on the perturbation approach in the frequency space: while the former only involves occupied single-particle orbitals, the latter requires the inclusion of unoccupied orbitals. Moreover, since the propagation in a single-time step of each KS state is independent of the others, the real-time approach is prone to a natural parallelization scheme which is ideally suited to deal with sizable systems. The inter-process communications are only required once per time step to calculate quantities that depend on a sum over states.

An efficient implementation of RT-TDDFT is freely available to users with the `octopus` code (http://octopus-code.org). `octopus` uses a real space grid representation combined with the finite difference approximation for the calculation of derivatives. Describing systems at the atomic level on such a grid ultimately demands the use of pseudo-potentials (Nogueira et al. 2003). The code implements several propagator schemes (Castro et al. 2004); it can handle mixed periodic boundaries and is highly suitable for massively parallel applications (Castro et al. 2006; Andrade et al. 2015). `octopus` is the main computational tool that has been used in the applications reviewed in the next section.

In the remainder of this chapter, we shall consider charge separation occurring after photoexcitation being driven by the vibrational motion of the nuclei. These kind of charge separation should not be confused with the, more commonly analyzed, charge-transfer excitations that can occur at fixed nuclear configurations within the linear regime (Dreuw and Head-Gordon 2004; Hieringer and Görling 2006). The processes we will focus on do not necessarily involve the presence of conical intersections and develop before any thermal equilibration. Therefore, under the assumption of a factorization of the electron-nuclear wave function, relevant dynamical couplings between the electrons and the nuclei can be described by resorting to the Ehrenfest molecular dynamics (EMD). In such a scheme, electrons are evolved according to the TD-KS equations along the trajectory of the instantaneous nuclear positions, while the nuclei are evolved according to Newtonian equations which account for the electron-nuclear interaction as mediated by the instantaneous electron density. The resulting TDDFT+Ehrenfest scheme may be regarded as a dynamical mean-field method. Crucially, the overall dynamics can be started from a KS state which corresponds to either the ground state of the system or to an excited (Tavernelli et al. 2005) or ionized (Gaigeot et al. 2010) configuration.

The time step required in EMD is determined by the maximum electronic frequencies, which means about three orders of magnitude less than the time step required to follow the nuclei by using the adiabatic Born-Oppenheimer approximation. Thus, to enable the application of EMD to realistic systems, efficient numerical implementations are essential. In the remainder of this chapter, we review a few applications of this technique that rest on a highly scalable implementation of a modified EMD (Andrade et al. 2009) included in the `octopus` code.

For the necessary details on the aforementioned theoretical aspects, we refer the reader to the first volume in the MTM set of this handbook and especially to Chap. 6, "TDDFT and Quantum-Classical Dynamics: A Universal Tool Describing the Dynamics of Matter" by Agostini et al. Other pedagogical textbooks discussing the fundamental aspects of TDDFT together with interesting applications are (Ullrich 2012; Marques et al. 2012). A recent review of the state of the art of capturing various types of charge-transfer processes by means of TDDFT approximations is available in Maitra (2017).

3   Charge Separation in Supramolecular Assemblies

The idea of mimicking natural light-harvesting and charge separation processes such as those which occur in photosynthesis is quite an old one (Ciamician 1912). Among the main tasks, two very important ones are to determine the relevant degrees of freedom and to understand the driving forces involved in the processes. The overwhelming complexity of the real systems, however, demands that simplified artificial systems should be made available by molecular self-assembling or more traditional chemical synthesis.

Supramolecular assemblies are ideal playgrounds to investigate the relationships between structure and function. They contain simple units, which can be tailored

to perform a specific task, as it happens in natural photosynthetic complexes. The performance of a device based on supramolecular systems can be optimized both at the molecular and at the supramolecular level (Gust 2015). For example, the Förster energy transfer rate between nearby moieties strongly depends on the spectral overlap and strength of specific electronic transitions in the donor and in the acceptor and on the relative transition dipoles. These properties can be tailored by a wise choice of the molecular components and by organizing the moieties in a suitable spatial arrangement. Their mutual distances and orientations will also affect the intermolecular electronic coupling, which, in turn, controls triplet-triplet energy transfer via the Dexter mechanism. Marcus' theory (1993) highlights the importance of the overlap of electronic orbitals in the donor and the acceptor for electron transfer reactions. Such overlap depends not only on the distance and the orientation of the moieties but also on the nature of their linker groups and on the presence of covalent bonding.

Nano-sized compounds able to provide sufficient thermodynamic driving force to promote photoinduced electron transfer have been obtained by combining porphyrins and fullerenes into covalently linked structures (Imahori 2004). The rationale behind this strategy, as outlined in Sect. 1, is to design each individual moiety to perform a specific action. Furthermore such ensembles display self-assembling properties that can be exploited to ease the fabrication of nano-devices (Umeyama and Imahori 2006), also useful for dye-sensitized solar cells (Planells et al. 2011). The basic donor-acceptor skeleton can be further engineered by adding one or more bridge units to reduce the recombination rate (see below). Other modifications include functionalizing the component moieties to improve their solubility or self-assembling properties, joining together other molecules such as carotenoids to elongate the chain (Liddell et al. 1994), building superstructures to scale up toward the size of natural light-harvesting complexes, yet resorting on simpler components and higher symmetry (Astruc et al. 2010).

As a prototypical case, we examine first a carotene-porphyrin-fullerene triad (CPF) about which a body of photochemical and photophysical studies exists (Kodis et al. 2004, and references therein). The molecule has a donor-bridge-acceptor structure. The porphyrin bridge acts also as the main chromophore. The linear absorption of the CPF is well described as the superposition of the absorption of each moiety (Kuciauskas et al. 2000), which suggests that the moieties are nearly uncoupled in the ground state (Spallanzani et al. 2009). Picosecond transient absorption experiments, however, reveal that upon visible light absorption, the excitation is transferred from the porphyrin to the fullerene. After about 120 ps, a strong new absorption band appears corresponding to the formation of the carotenoid radical cation of the final charge-separated state (i.e., a state having a hole localized on the carotenoid and an electron on the C_{60}).

Here, we focus on the early onset of the charge separation: in particular, we show that mixed quantum-classical simulations can reveal coherent wavelike electronic-charge motions which are driven by nuclear vibrations. The system can be reduced to a smaller molecule with 207 atoms (see Fig. 2a), and its dynamics is studied in the sub-ps regime. Previous studies about the CPF geometry in liquid solvents
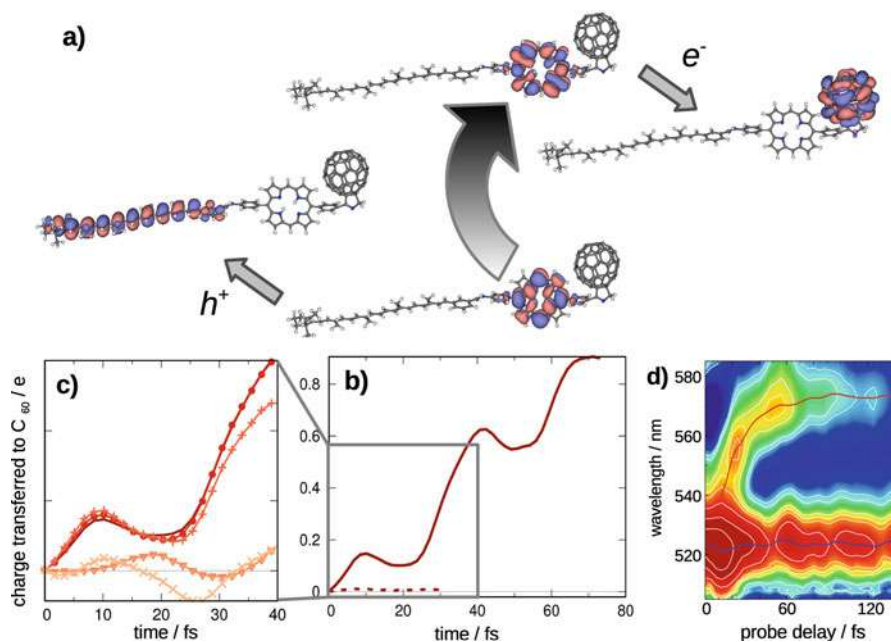
Fig. 2 (**a**) Kohn-Sham states involved in the porphyrin excitation (middle) and the final charge-separated states (sides). (**b**) Net charge flow from the carotene-porphyrin to the C_{60} in the ultrafast regime: no charge is transferred when all of the nuclei are clamped (dashed line); wavelike charge accumulation appears when the nuclei are free to move and their dynamics is coupled to the electronic one (solid line). (**c**) Lines from darker to lighter colors correspond to lower charge-transfer rates in this order: clamped carotene (circles almost on top of the free nuclei line); clamped C_{60} (plus sign); clamped porphyrin (triangles); clamped aryl ring connecting the porphyrin to the C_{60}. (**d**) Experimental result: differential transmission following impulsive excitation at 550 nm (the scale from blue to red corresponds to a signal from -0.05 to 0.12). The marked feature at 520 nm corresponds to ground-state photobleaching. An increased transmission band drifts toward lower energy. This is the footprint of a transient state, which has the same temporal behavior as the simulated charge-transfer signal. (Data from Rozzi et al. 2013)

have shown that the triad is structurally flexible under ambient conditions. Baruah and Pederson (2009), by combining DFT at the generalized gradient approximation level and kinetic Monte Carlo, have shown that the solvent polarization contributes to the reordering the excited-state energies and have identified a stable bent configuration in vacuo. Su et al. (2012) have investigated the impact of thermal fluctuation on both the triad and solvent. They have probed the conformational distribution of the CPF ground state and have found that the energy landscape of an ensemble of the triad is characterized by a very shallow barrier separating local minima ($<1k_BT$). Manna et al. (2015) employed linear response TDDFT with the BNL range-separated hybrid functional (Baer and Neuhauser 2005; Livshits and Baer 2007) and a polarizable continuum model for the solvent to calculate the energy of solvated charge-transfer states. They also estimate the kinetics of electron transfer in various configurations by means of molecular

dynamics and conclude that a linear configuration of the molecule is the most favorable for the charge separation process. In the linear configuration, the states involved in the photon absorption are strongly localized on the porphyrin, with virtually no overlap with donor and acceptor states (see Fig. 2a). In this case the electron dynamics with no nuclear motion is unable to explain the observed charge separation. RT-TDDFT and EMD together provide, instead, valuable insight into the process. Under the condition that the atoms are free to move, simulations show that the excited electronic wave packet undergoes an oscillatory dynamics, while at the same time charge is being accumulated on the acceptor. These facts reveal that the coupled motion of electrons and nuclei drives the charge separation in the ultrafast window. The simulated charge transfer depicted in Fig. 2b provides two quantitative predictions: (1) that the electron charge oscillations occur with a period of about 30 fs; (2) one electron can be transferred from the carotene-porphyrin to the fullerene moiety in about 70 fs. These results (Rozzi et al. 2013) are in remarkable agreement with high time-resolution observations of the transient absorption spectrum, which shows oscillations in the intensity of the porphyrin Q-band absorption at 523 nm, and also a peculiar stimulated emission band, whose intensity *and resonant wavelength* oscillate at the same frequency as the ground-state one, but in phase opposition with it. The latter band is strongly red-shifted and it disappears after about 70 fs (see Fig. 2d). These features make it impossible to associate the stimulated emission signal with the ballistic motion of the excited wave packet from the donor to the acceptor, and they are compatible with the scenario of coherent charge oscillations devised in the simulations.

It is also possible to perform numerical experiments by clamping different groups of atoms and therefore predict different charge-transfer scenarios. The simulated charge signals in Fig. 2c confirm that while the C_{60} moiety and the carotenoid have little or no influence on the ultrafast regime, impairing the movement of the porphyrin or the linker group between the porphyrin and C_{60} can completely quench the transfer while still leaving the vibrational sloshing-type motion of the wave packet active (see Fig. 2c). These insights would not be easily available without running numerical simulations and can be exploited to manipulate the vibrational properties of the molecule. Vibronic couplings can affect the electron transfer rates since they can transform individual donor-acceptor electronic states into new manifolds of states, which can reduce the transfer energy barrier or open up alternative transfer paths through resonance. Delor et al. (2014) provide a specific example of vibronic manipulation. They show that the efficiency of electron transfer in a covalently bonded donor-bridge-acceptor triad can be controlled by means of narrow-band infrared laser pulses affecting the vibrations of the bridge unit. Moreover, when the system reorganization energy is small, vibronic states become delocalized throughout the moieties. Delocalization, in turn, provides a robust protection against environmental fluctuations due to noise or disorder, and it makes extended coherent states available. The observation of long-lasting coherent oscillations in 2D electronic spectroscopy (Engel et al. 2007) has stimulated an extensive debate about the possible role of quantum coherence in charge and energy

transfer phenomena. A more extensive overview of these aspects can be found in Brédas et al. (2016) and Scholes et al. (2017).

3.1   From Triads to Dyads

The rationale behind building triads instead of dyads is that having a longer charge-separating chain is a way to effectively decrease the coupling between the electron and the hole collectors, thereby reducing the exciton recombination rate. Moreover, the presence of a long sequence of conjugated bonds in the bridge connecting donor and acceptor can enhance the spectral absorption of the complex at longer wavelengths – a beneficial effect. In the case of the CPF triad, transient absorption experiments reveal pronounced long-lived oscillatory modulations in the range between 510 and 580 nm due to the wave packet oscillations in the ground state of the carotene. All this action, however, is not relevant to the interesting part of the ultrafast photoinduced dynamics. Besides, the synthesis of triads is more complex than the synthesis of dyads.

One way to simplify both synthesis and analysis is to build a dyad by using a component that can be easily attached to a variety of other molecules. For example, the diaminoterephthalate unit – which was originally designed for combinatorial chemistry (see Fig. 3a) – can be easily attached to a C_{60} acceptor by a pyrrolidine linker (Freimuth et al. 2015). The resulting system (simply named as dyad hereafter) is much smaller than the CPF, and yet it may support similar charge separation mechanisms.

This dyad is significantly less flexible than the CPF triad, but it may undergo free rotations around the C-C bond connecting the chromophore to the benzyl group and around the contiguous C-O single bond. This produces *syn* and *anti* conformers energetically interchangeable at room temperature (see Fig. 3a). As we will shortly see, these conformers imply different absorption properties.

Although the overall shape of the absorption spectra for the two conformers is expected to be very similar, linear response-level calculations show that there are indeed significant differences (see Fig. 3b). The *syn* conformer displays a weak absorption band at very low energy having charge-transfer character. The donor and acceptor are closely facing in this conformer, and there is a non-zero tunneling probability for an electron to move between the two. An interesting region of the spectrum is the green-yellow band around 2.5 eV. In this spectral region, a chromophore excitation is optically active in both conformers. However, while the excitation is very well localized on the diaminoterephtalate in the *anti* configuration, in the *syn* configuration, the transition has a more delocalized character as it involves also the functionalized C_{60} acceptor. As a consequence, it is expected that the *syn* conformer should more favorably support charge displacements along the molecule than the *anti* conformer. However, the absorption in the visible is not particularly strong.

The aforementioned trends suggested by the inspection of the absorption cross section can be confirmed by performing RT propagations of the coupled electron-
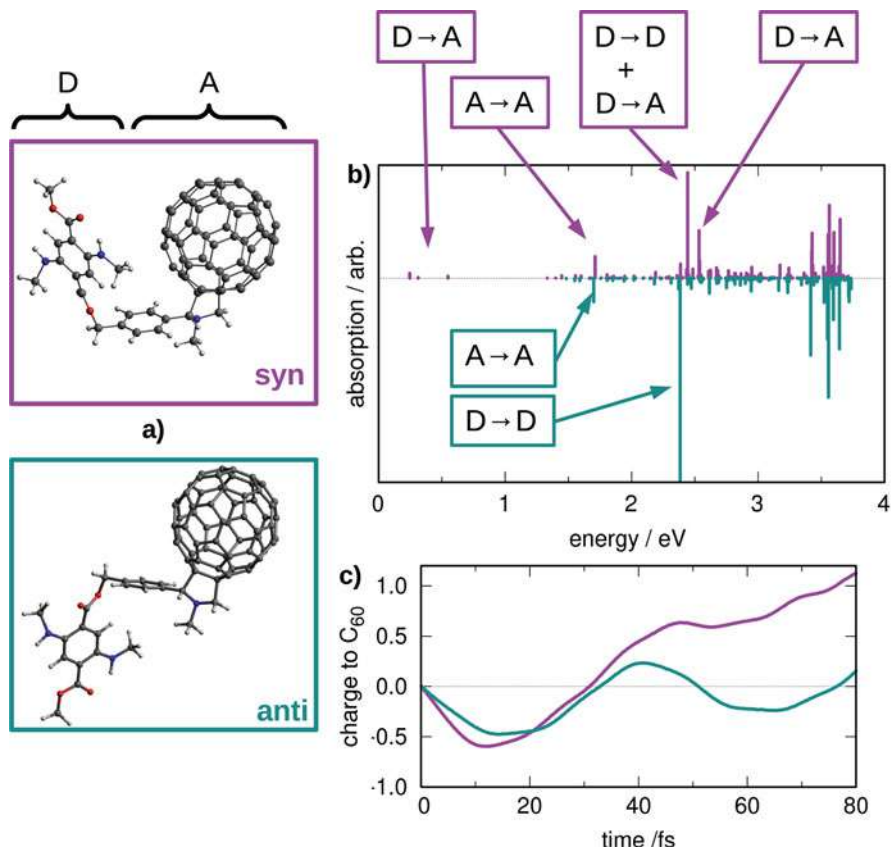
Fig. 3 (**a**) Energetically quasi degenerate conformers of the diaminoterephtalate-C_{60} molecule. D and A indicate the donor and the acceptor, respectively. (**b**) Linear response UV-Vis absorption: the upper part refers to the *syn* conformer and the lower part to the *anti* conformer. (**c**) Real-time simulation of the charge flow from the chromophore to the C_{60}. (Copyright Wiley 2015, adapted with permission)

nuclei system using TDDFT and EMD. As an initial state, we may take the single-particle excitation from the chromophore HOMO to the chromophore LUMO. While this choice is realistic for the *anti* conformer, it is less appropriate for the *syn* conformer as it overestimates the localization of the initial excitation. For a comparison, however, it is useful to start with similar initial conditions. The RT propagations yield two clearly distinct scenarios (Pittalis et al. 2015). By tracking the charge motion by numerical integration of the particle density in a volume including the acceptor moiety, it is evident that the *syn* conformer has a substantially larger charge separation power than the *anti* conformer (see Fig. 3c). In other words, the dynamics is not robust with respect to the choice of the initial nuclear configuration. As a common feature with the CPF case, it is observed that the

charge oscillations disappear by suppressing the electron-nuclei coupling, thereby confirming the prominent role of intramolecular vibrations.

These simulations help identify the possible channels which drive the path of the photoexcited charge in the dyad. While in the *syn* conformer both through-bonds and through-space transfer mechanisms could be active (Pettersson et al. 2006), in the *anti* conformer the direct hopping probability of an electron through space tunneling from the chromophore to the acceptor is strongly suppressed. Due to the competition of different transmission channels and considering that a solution should contain both conformers in equal parts at room temperature, it is not expected that a coherent signal can be easily observable in pump-probe experiments on this compound. We remark that the simulations shown in Fig. 3c are from single trajectory run – not from a statistical sampling.

In summary, the considered dyad exhibits some desired features. But to become useful in applications, the system needs to be further engineered, improving the light-harvesting efficiency and enhancing more favorable charge separation pathways.

4   Ultrafast Charge Separation in Bulk Heterojunctions: The Case of P3HT:PCBM

In 1992, Sariciftci et al. reported that when a semiconducting polymer is mixed with an electron acceptor (C_{60}), a quenching of the polymer photoluminescence is observed. This is a clear evidence for the charge-separating properties of the mixture. The luminescence decreases in time because the polymer's S1 excited state, initially populated by solar light, is being depleted, as the electrons transfer to the acceptor before they recombine back to the S0 ground state. This observation provides an upper bound of the order of 1 ps for the charge separation time scale in this material (Lee et al. 1993). Radiative and non-radiative recombination channels are two orders of magnitude slower, which makes quantum efficiency close to unity. The same happens in many other combinations of organic electron donors and acceptors. Time-resolved photoluminescence however has a rather low time resolution, and the estimate of the upper boundary of the time scale does not disclose the details of the electron dynamics in the material.

Using transient absorption spectroscopy with high time resolution (about 10 fs), Brabec et al. (2001) reported that some features in the differential transmission spectra change on a much shorter time scale. In particular, the polymer stimulated emission band turns into a photoinduced absorption band in about 100 fs after pumping the sample with a broadband laser pulse. The photoinduced absorption band is a property of the excited state assigned to the polymer cation radical. This observation was later confirmed by many groups in similar blends, such as P3HT:PCBM (Hwang et al. 2007; Cook et al. 2009; Piris et al. 2009; Howard et al. 2010; Kaake et al. 2012).

These experiments highlight the existence of a sub-100 fs dynamics connected to charge transfer, despite the fact that the production of *free* charges is a much

slower process (which goes from few ps to hundreds of ps) (Hwang et al. 2008). One might argue that the direct excitation of charge-transfer states – which is possible, albeit, with a small cross section – might explain early time dynamics. However, by comparing photoluminescence to photocurrent, it turns out that the contribution of the below-gap radiation is negligible with respect to that of the above-gap exciton (Drori et al. 2010). This clearly indicates that direct charge transfer is not a charge generation mechanism of relevance in these materials. Moreover, Gelinas et al. (2014) reported that the Coulomb binding attraction can be overcome on a time scale as short as 40 fs, suggesting that delocalization may be a crucial factor affecting the system dynamics.

This is a case in which RT-TDDFT simulations, can help clarify the situation and provide information complementing high time-resolution pump-probe spectroscopic data. In order to make the system tractable, suitable simplifications should be admitted. If the photoinduced charge dynamics is an interfacial phenomenon, it can be studied by reducing the P3HT:PCBM to its smallest unit: a single polymer chain (including four thiol rings per unit cell) facing a single C_{60} molecule. The structure optimization of this cell yields an equilibrium distance between the donor and the acceptor of 3.2 Å, corresponding to $\pi - \pi$ stacking of fullerene and polymer rings. Also suitable initial conditions must be set on the electronic wave packet. Real-time simulations can be performed by including explicitly a classical pulsed laser field. A simpler starting point consists in assuming that the first excited state corresponds to a sudden vertical excitation of the polymer in the Franck-Condon framework. Adiabatic (semi-)local functionals describe this transition by predominantly involving the HOMO and LUMO of the polymer. Therefore, the initial state in our subsequent RT propagation can be set by modifying the occupations of these two levels at time zero; i.e., by depopulating the polymer HOMO while populating the polymer LUMO.

Free RT propagation of the electronic wave packet from the aforementioned initial condition reveals the crossover of the singlet exciton to the charge-transfer state (Falke et al. 2014). This can be rationalized by examining the spatial localization of the HOMO of the polymer and the LUMO of the fullerene. A periodic crossover between the corresponding KS energy levels emerges when nonadiabatic coupling is enabled in the system Hamiltonian (see Fig. 4b). The excited state described by the HOMO of the polymer is free to delocalize over one of the fullerene unoccupied states as soon as the two become isoenergetic. The change in localization provides a net current flowing from P3HT to C_{60}. Driven by the vibrations of the polymer backbone, a wavelike charge movement occurs. Because of the nuclear motion, it progressively gets stabilized to a self-trapping charge-separated state. Looking at (a) and (b) in Fig. 4, it is easy to match the crossing points of the excited donor and the unoccupied acceptor states to times at which the charge-transfer instantaneous rate is maximum. The dynamics clearly proceeds on three different time scales. At times earlier than 23 fs, the dynamics is purely electronic, and it is driven by a fast stabilization of the polymer excited state. After about 23 fs, the first crossing point is reached and the oscillatory motion kicks in. Its frequency is compatible with the polymer C=C stretch frequency. This motion, in combination with a slower
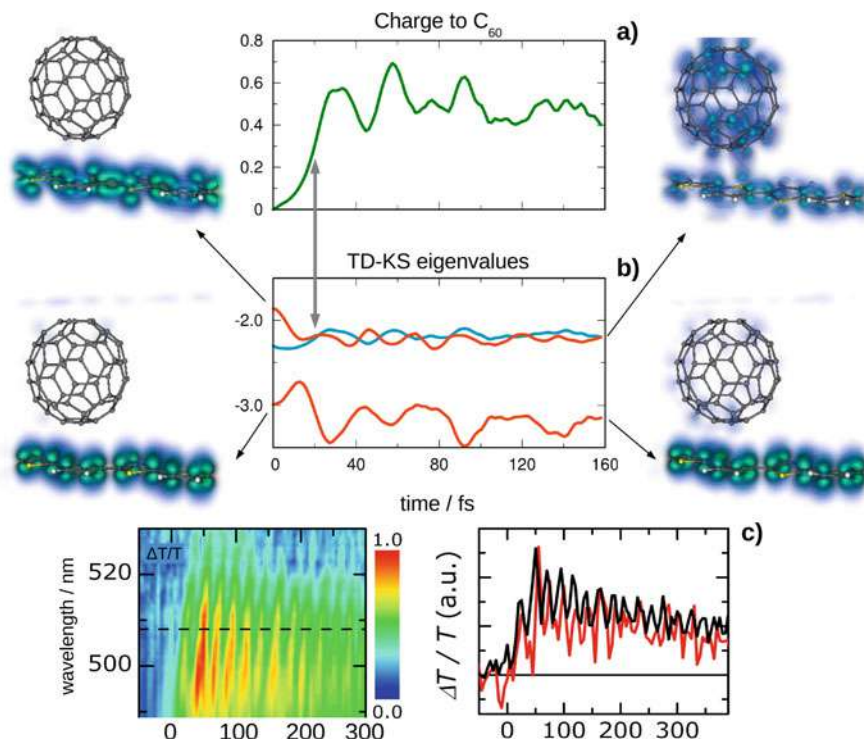
Fig. 4 Snapshots of the simulated joint electronic and nuclear dynamics of a P3HT:PCBM blend. (**a**) Probability of charge transfer from the polymer chain to the fullerene cage as a function of time. (**b**) Kohn-Sham energy levels mainly involved in the charge displacement. Red line: the time-evolved KS eigenvalues starting from the polymer HOMO and LUMO. Blue line: the average of the three lowest unoccupied levels on the C_{60} (almost degenerate in the ground state). When the red and blue lines cross, the flow of the current from the P3HT to the C_{60} is maximal. Left and right panels show the localization of the corresponding Kohn-Sham single-particle orbitals soon after the photon absorption (time 0) and after about 100 fs. During this interval, the highest occupied state in the system delocalizes from the donor to the acceptor. (**c**) Left: differential transmission ($\Delta T/T$) map, which shows coherent oscillations. Right: a section corresponding to the dashed line in the $\Delta T/T$ map (black line) compared to the signal in the pristine polymer (red line). (Data from Falke et al. 2014)

adjustment in the nuclear configuration, leads to a maximum charge transfer of about 0.7 e after approximately 50 fs. Up to this point both the computed vibrational frequency and the electron transfer time constant are in quantitative agreement with the observed ones. At the same time, however, the combination of the adiabatic LDA functional and EMD becomes inaccurate. The excited donor is eventually stabilized to a maximally delocalized state as higher-frequency vibrational modes are progressively engaged and the charge accumulation falls back to about 0.5 e. In reality, inelastic scattering channels (especially vibrational ones) damp the oscillatory motion at later times. Simulating such relaxation processes, however,

would entail a substantial extension of the theoretical approach described in this chapter.

In summary, we have seen that important features of the coherent charge motion following photoexcitation in the P3HT:PCBM bulk heterojunction can be accurately described by employing RT-TDDFT and Ehrenfest-path molecular dynamics. In order to capture more phenomenology, additional capabilities should be added to the employed computational tools. For example, it was recently shown by adopting a non-equilibrium Green's function approach that space-time non-local correlations may explain the decay of the charge-transfer exciton into a charge-separated state in terms of correlation-induced decoherence, provided that the corresponding electronic levels are suitably aligned, nicely complementing our picture (Boström et al. 2018).

5   Conclusions

The challenging quest for developing efficient, reliable, and cheap organic photovoltaic materials also with the help of computational methods based on first principles is undergoing a fast evolution (Rozzi et al. 2018). Prototyping simple functional systems is an important step toward the understanding of the fundamental physics underlying ultrafast phenomena in photovoltaics. This line of research will stimulate the development of more advanced theoretical and computational methods.

In this chapter, we have examined a few paradigmatic cases for which even simple density-functional approximations lead to a description of the ultrafast, coherent components of photoinduced charge dynamics in quantitative agreement with high-resolution transient spectroscopy experiments. For the considered cases, a pivotal role was played by the dynamical coupling of the nuclei with the electrons. Modeling the nuclear dynamics semiclassically on the Ehrenfest mean-field path permits simulations up to about 100 fs, when the actual quantum nature of the nuclei can be neglected. Hence, we have seen that the information gained from time-resolved experiments and RT-TDDFT simulations, when combined, provide more useful information than the sum of their parts taken independently.

References

Alvarado SF, Seidler PF, Lidzey DG, Bradley DDC (1998) Direct determination of the exciton binding energy of conjugated polymers using a scanning tunneling microscope. Phys Rev Lett 81(5):1082–1085. https://doi.org/10.1103/PhysRevLett.81.1082

Andrade X, Castro A, Zueco D, Alonso JL, Echenique P, Falceto F, Rubio A (2009) Modified Ehrenfest formalism for efficient large-scale ab initio molecular dynamics. J Chem Theory Comput 5(4):728–742. https://doi.org/10.1021/ct800518j

Andrade X, Strubbe D, De Giovannini U, Larsen AH, Oliveira MJT, Alberdi-Rodriguez J, Varas A, Theophilou I, Helbig N, Verstraete MJ, Stella L, Nogueira F, Aspuru-Guzik A, Castro A, Marques MAL, Rubio A (2015) Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems. Phys Chem Chem Phys 17(47):31371–31396. https://doi.org/10.1039/C5CP00351B

Astruc D, Boisselier E, Ornelas C (2010) Dendrimers designed for functions: from physical, photophysical, and supramolecular properties to applications in sensing, catalysis, molecular electronics, photonics, and nanomedicine. Chem Rev 110(4):1857–1959. https://doi.org/10.1021/cr900327d

Baer R, Neuhauser D (2005) Density functional theory with correct long-range asymptotic behavior. Phys Rev Lett 94:043002. https://doi.org/10.1103/PhysRevLett.94.043002

Baruah T, Pederson MR (2009) DFT calculations on charge-transfer states of a carotenoid-porphyrin-C_{60} molecular triad. J Chem Theory Comput 5(4):834–843. https://doi.org/10.1021/ct900024f

Boström EV, Mikkelsen A, Verdozzi C, Perfetto E, Stefanucci G (2018) Charge separation in donor-C_{60} complexes with real-time Green functions: the importance of nonlocal correlations. Nano Lett 18(2):785–792. https://doi.org/10.1021/acs.nanolett.7b03995

Brabec CJ, Zerza G, Cerullo G, De Silvestri S, Luzzati S, Hummelen JC, Sariciftci NS (2001) Tracing photoinduced electron transfer process in conjugated polymer/fullerene bulk heterojunctions in real time. Chem Phys Lett 340(3–4):232–236. https://doi.org/10.1016/S0009-2614(01)00431-6

Brédas JL, Sargent EH, Scholes GD (2016) Photovoltaic concepts inspired by coherence effects in photosynthetic systems. Nat Mater 16(1):35–44. https://doi.org/10.1038/nmat4767

Casida M (1995) Time-dependent density functional response theory for molecules. In Recent Advances in Computational Chemistry: Volume 1 Recent Advances in Density Functional Methods (Part I) Ed. DP Chong, World Scientific, pp 155–192. https://doi.org/10.1142/9789812830586_0005

Castro A, Marques MAL, Rubio A (2004) Propagators for the time-dependent Kohn-Sham equations. J Chem Phys 121(8):3425–33. https://doi.org/10.1063/1.1774980, http://www.ncbi.nlm.nih.gov/pubmed/15303905

Castro A, Appel H, Oliveira MJT, Rozzi CA, Andrade X, Lorenzen F, Marques MAL, Gross EKU, Rubio A (2006) Octopus: a tool for the application of time-dependent density functional theory. Phys Status Solidi 243(11):2465–2488. https://doi.org/10.1002/pssb.200642067

Ciamician G (1912) The photochemistry of the future. Science 36(926):385–394. https://doi.org/10.1126/science.36.926.385

Cook S, Katoh R, Furube A (2009) Ultrafast studies of charge generation in PCBM:P3HT blend films following excitation of the fullerene PCBM. J Phys Chem C 113(6):2547–2552. https://doi.org/10.1021/jp8050774

Delor M, Scattergood PA, Sazanovich IV, Parker AW, Greetham GM, Meijer AJHM, Towrie M, Weinstein JA (2014) Toward control of electron transfer in donor-acceptor molecules by bond-specific infrared excitation. Science 346(6216):1492–1495. https://doi.org/10.1126/science.1259995

Dreuw A, Head-Gordon M (2004) Failure of time-dependent density functional theory for long-range charge-transfer excited states: the zincbacteriochlorin–bacteriochlorin and bacteriochlorophyll–spheroidene complexes. J Am Chem Soc 126(12):4007–4016. https://doi.org/10.1021/ja039556n

Drori T, Holt J, Vardeny Z (2010) Optical studies of the charge transfer complex in polythiophene/fullerene blends for organic photovoltaic applications. Phys Rev B 82(7):1–8. https://doi.org/10.1103/PhysRevB.82.075207

Dunlap WC, Watters RL (1953) Direct measurement of the dielectric constants of silicon and germanium. Phys Rev 92:1396–1397. https://doi.org/10.1103/PhysRev.92.1396

Engel GS, Calhoun TR, Read EL, Ahn TK, Mančal T, Cheng YC, Blankenship RE, Fleming GR (2007) Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. Nature 446(7137):782–786. https://doi.org/10.1038/nature05678

Falke SM, Rozzi CA, Brida D, Maiuri M, Amato M, Sommer E, De Sio A, Rubio A, Cerullo G, Molinari E, Lienau C (2014) Coherent ultrafast charge transfer in an organic photovoltaic blend. Science 344(6187):1001–1005. https://doi.org/10.1126/science.1249771

Freimuth L, Rozzi CA, Lienau C, Christoffers J (2015) A Diaminoterephthalate-C60 Dyad: a new material for optoelectronic applications. Synthesis (Stuttg) 47(09):1325–1328. https://doi.org/10.1055/s-0034-1379963

Gaigeot MP, Lopez-Tarifa P, Martin F, Alcami M, Vuilleumier R, Tavernelli I, Hervé du Penhoat MA, Politis MF (2010) Theoretical investigation of the ultrafast dissociation of ionised biomolecules immersed in water: direct and indirect effects. Mutat Res Rev Mutat Res 704(1–3):45–53. https://doi.org/10.1016/j.mrrev.2010.01.004

Gelinas S, Rao A, Kumar A, Smith SL, Chin AW, Clark J, van der Poll TS, Bazan GC, Friend RH (2014) Ultrafast long-range charge separation in organic semiconductor photovoltaic diodes. Science 343(6170):512–516. https://doi.org/10.1126/science.1246249

Green MA (2013) Improved value for the silicon free exciton binding energy. Adv 3(11):112104. https://doi.org/10.1063/1.4828730

Gust D (2015) Supramolecular photochemistry applied to artificial photosynthesis and molecular logic devices. Faraday Discuss 185:9–35. https://doi.org/10.1039/C5FD00142K

Hieringer W, Görling A (2006) Failure of time-dependent density functional methods for excitations in spatially separated systems. Chem Phys Lett 419(4):557–562. https://doi.org/10.1016/j.cplett.2005.11.112

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871. https://doi.org/10.1103/PhysRev.136.B864

Howard IA, Mauer R, Meister M, Laquai F (2010) Effect of morphology on ultrafast free carrier generation in polythiophene: fullerene organic solar cells. J Am Chem Soc 132(7):14866–14876. https://doi.org/10.1021/ja105260d

Hwang IW, Soci C, Moses D, Zhu Z, Waller D, Gaudiana R, Brabec CJ, Heeger AJ (2007) Ultrafast electron transfer and decay dynamics in a small band gap bulk heterojunction material. Adv Mater 19(17):2307–2312. https://doi.org/10.1002/adma.200602437

Hwang IW, Moses D, Heeger AJ (2008) Photoinduced carrier generation in P3HT/PCBM bulk heterojunction materials. J Phys Chem C 112(11):4350–4354. https://doi.org/10.1021/jp075565x

Imahori H (2004) Porphyrin-fullerene linked systems as artificial photosynthetic mimics. Org Biomol Chem 2(10):1425–1433. https://doi.org/10.1039/B403024A

Jakowetz AC, Böhm ML, Zhang J, Sadhanala A, Huettner S, Bakulin AA, Rao A, Friend RH (2016) What controls the rate of ultrafast charge transfer and charge separation efficiency in organic photovoltaic blends. J Am Chem Soc 138(36):11672–11679. https://doi.org/10.1021/jacs.6b05131

Jonas DM (2003) Two-dimensional femtosecond spectroscopy. Ann Rev Phys Chem 54(1):425–463. https://doi.org/10.1146/annurev.physchem.54.011002.103907

Kaake LG, Jasieniak JJ, Bakus RC, Welch GC, Moses D, Bazan GC, Heeger AJ (2012) Photoinduced charge generation in a molecular bulk heterojunction material. J Am Chem Soc 134(48):19828–19838. https://doi.org/10.1021/ja308949m

Kodis G, Liddell PA, Moore AL, Moore TA, Gust D (2004) Synthesis and photochemistry of a carotene–porphyrin–fullerene model photosynthetic reaction center. J Phys Org Chem 17(9):724–734. https://doi.org/10.1002/poc.787

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133

Kuciauskas D, Liddell Pa, Lin S, Stone SG, Moore AL, Moore Ta, Gust D (2000) Photoinduced electron transfer in carotenoporphyrin−fullerene triads: temperature and solvent effects. J Phys Chem B 104(18):4307–4321. https://doi.org/10.1021/jp9935135

Lee C, Yu G, Moses D, Pakbaz K, Zhang C, Sariciftci NS, Heeger AJ, Wudl F (1993) Sensitization of the photoconductivity of conducting polymers by C60: photoinduced electron transfer. Phys Rev B 48(20):15425–15433. https://doi.org/10.1103/PhysRevB.48.15425

Livshits E, Baer R (2007) A well-tempered density functional theory of electrons in molecules. Phys Chem Chem Phys 9:2932–2941. https://doi.org/10.1039/B617919C

Li G, Zhu R, Yang Y (2012) Polymer solar cells. Nat Photon 6(3):153–161. https://doi.org/10.1038/nphoton.2012.11

Liddell PA, Sumida JP, Macpherson AN, Noss L, Seely GR, Clark KN, Moore AL, Moore TA, Gust D (1994) Preparation and photophysical studies of porphyrin-C_{60} dyads. Photochem Photobiol 60(6):537–541. https://doi.org/10.1111/j.1751-1097.1994.tb05145.x

Maitra NT (2017) Charge transfer in time-dependent density functional theory. J Phys Condens Matter 29(42):423001. https://doi.org/10.1088/1361-648X/aa836e

Manna AK, Balamurugan D, Cheung MS, Dunietz BD (2015) Unraveling the mechanism of photoinduced charge transfer in carotenoid-porphyrin-C 60 molecular triad. J Phys Chem Lett 6(7):1231–1237. https://doi.org/10.1021/acs.jpclett.5b00074

Marcus RA (1993) Electron transfer reactions in chemistry. Theory and experiment. Rev Mod Phys 65(3):599–610. https://doi.org/10.1103/RevModPhys.65.599

Marques MAL, Maitra NT, Nogueira FMS, Gross EKU, Rubio A (eds) (2012) Fundamentals of time-dependent density functional theory, vol 837. Springer, Berlin/Heidelberg. https://doi.org/10.1007/978-3-642-23518-4

Nogueira F, Castro A, Marques M (2003) A tutorial on density functional theory. Springer, Berlin/Heidelberg, pp 218–256. https://doi.org/10.1007/3-540-37072-2_6

Perdew JP, Schmidt K (2001) Jacob's ladder of density functional approximations for the exchange-correlation energy. AIP Conf Proc 577(1):1–20. https://doi.org/10.1063/1.1390175

Pettersson K, Kyrychenko A, Rönnow E, Ljungdahl T, Mårtensson J, Albinsson B (2006) Singlet energy transfer in porphyrin-based donor-bridge-acceptor systems: interaction between bridge length and bridge energy. J Phys Chem A 110(1):310–8. https://doi.org/10.1021/jp053819d

Piris J, Dykstra TE, Bakulin AA, Van Loosdrecht PHM, Knulst W, Trinh MT, Schins JM, Siebbeles LDA (2009) Photogeneration and ultrafast dynamics of excitons and charges in P3HT/PCBM blends. J Phys Chem C 113(32):14500–14506. https://doi.org/10.1021/jp904229q

Pittalis S, Delgado A, Robin J, Freimuth L, Christoffers J, Lienau C, Rozzi CA (2015) Charge separation dynamics and Opto-electronic properties of a diaminoterephthalate-C60 dyad. Adv Funct Mater 25(13):2047–2053. https://doi.org/10.1002/adfm.201402316

Planells M, Pellejà L, Ballester P, Palomares E (2011) Utilization of a heterosupramolecular self-assembled trisporphyrin complex in dye-sensitised solar cells. Energy Environ Sci 4(2):528–534. https://doi.org/10.1039/C0EE00406E

Rozzi C, Falke S, Spallanzani N, Rubio A, Molinari E, Brida D, Maiuri M, Cerullo G, Schramm H, Christoffers J, Lienau C (2013) Quantum coherence controls the charge separation in a prototypical artificial light-harvesting system. Nat Commun 4:1602. https://doi.org/10.1038/ncomms2603

Rozzi CA, Troiani F, Tavernelli I (2018) Quantum modeling of ultrafast photoinduced charge separation. J Phys Condens Matter 30(1):013002. https://doi.org/10.1088/1361-648X/aa948a

Runge E, Gross EKU (1984) Density-functional theory for time-dependent systems. Phys Rev Lett 52:997–1000. https://doi.org/10.1103/PhysRevLett.52.997

Sariciftci NS, Smilowitz L, Heeger AJ, Wudl F (1992) Photoinduced electron transfer from a conducting polymer to buckminsterfullerene. Science 258(5087):1474–1476. https://doi.org/10.1126/science.258.5087.1474

Scholes GD, Fleming GR, Chen LX, Aspuru-Guzik A, Buchleitner A, Coker DF, Engel GS, van Grondelle R, Ishizaki A, Jonas DM, Lundeen JS, McCusker JK, Mukamel S, Ogilvie JP, Olaya-Castro A, Ratner MA, Spano FC, Whaley KB, Zhu X (2017) Using coherence to enhance

function in chemical and biophysical systems. Nature 543(7647):647–656. https://doi.org/10.1038/nature21425

Spallanzani N, Rozzi CA, Varsano D, Baruah T, Pederson MR, Manghi F, Rubio A (2009) Photoexcitation of a light-harvesting supramolecular triad: a time-dependent DFT study. J Phys Chem B 113(16):5345–5349. https://doi.org/10.1021/jp900820q

Su G, Czader A, Homouz D, Bernardes G, Mateen S, Cheung MS (2012) Multiscale simulation on a light-harvesting molecular triad. J Phys Chem B 116(29):8460–8473. https://doi.org/10.1021/jp212273n

Tavernelli I, Röhrig UF, Rothlisberger U (2005) Molecular dynamics in electronically excited states using time-dependent density functional theory. Mol Phys 103(6–8):963–981. https://doi.org/10.1080/00268970512331339378

Ullrich C (2012) Time-dependent density-functional theory: concepts and applications. Oxford graduate texts. Oxford University Press, Oxford

Umeyama T, Imahori H (2006) Self-organization of porphyrins and fullerenes for molecular photoelectrochemical devices. Photosynth Res 87(1):63–71. https://doi.org/10.1007/s11120-005-4632-z

Yabana K, Nakatsukasa T, Iwata JI, Bertsch GF (2006) Real-time, real-space implementation of the linear response time-dependent density-functional theory. Physica Status Solidi B 243(5):1121–1138. https://doi.org/10.1002/pssb.200642005

Zhong C, Bartelt JA, McGehee MD, Cao D, Huang F, Cao Y, Heeger AJ (2015) Influence of intermixed donor and acceptor domains on the ultrafast charge generation in bulk heterojunction materials. J Phys Chem C 119(48):26889–26894. https://doi.org/10.1021/acs.jpcc.5b09572

Part IV
Modeling Applications to Magnetism, Magnetic Materials, and Spintronics

Applications of Materials Modeling to Magnetism, Magnetic Materials, and Spintronics: Overview

16

Stefano Sanvito

Contents

Abstract

Magnetism underpins many applications in modern technology, where either its static or dynamical properties are exploited. This chapter discusses some of these applications, in particular focussing on three main aspects. Firstly, it will review how modern electronic structure theory can contribute to the discovery and design of novel magnetic materials with multiple functionalities. Secondly, it will address the phenomena related to magnetization dynamics for both spintronics and data-storage applications. Finally, it will look at the interplay between electrical currents and spins. Throughout the chapter special attention will be devoted to present both the necessary theoretical background and discuss how theory may underpin the development of applications.

1 Open issues in Magnetic Materials and Spintronics

The love affair between mankind and magnetic materials has been continuing for 2,000 years and does not show any sign of losing steam. From the invention of the compass, which enabled the eleventh-century Chinese dynasty of Song to embark

S. Sanvito (✉)
Center for Materials Genomics, Duke University, Durham, NC, USA

School of Physics and CRANN Institute, Trinity College, Dublin, Ireland
e-mail: sanvitos@tcd.ie

on long-distance navigation, magnetic materials have been key for developments in human technology. Today magnets are everywhere in our life. The hard disk of our computer is composed by billions of tiny magnets and it is read by a sensor, whose main component is a complex magnetic nanostructure. Wind turbines are made of strong permanent magnets, as are the electrical motors in our cars, in our kitchen blenders, and in our lawn mowers.

Yet, nature is stingy when it comes to magnets. The first report of a magnetic material dates back to 79 AD, and it is contained in what was probably the first scientific encyclopedia ever, the *Naturalis Historia* (Natural History) from the Roman natural philosopher Plinius the Elder. In the Historia, Plinius, who perished in the eruption of the Vesuvio in 79 AD, reports the legend of "Magnes the shepherd," who discovered two metallic stones attracting each other. These were likely made of magnetite, Fe_3O_4, a mineral abundant in the earth crust and one of the main iron ores. It took about 2,000 years to understand the microscopic origin of magnetism, which was possible only after the formulation of quantum mechanics. During all this time, we have discovered about 4,000 materials, which behave as magnets. In other words, our discovery speed is two magnets per year. Interestingly, only a handful of them are used in mainstream applications.

This section reviews the most recent applications of magnetism and the theoretical methods most suitable for describing them. It does not pretend to be a comprehensive review but to cover some of the most modern and fascinating aspects. In particular it will touch upon three main aspects, namely, the discovery and design of new magnets, the description of macroscopic magnetization dynamics for both spintronics and data-storage applications, and the interplay between electrical currents and spins at the atomic level.

From the materials point of view, magnetism is a relatively rare occurrence in the materials world. In fact, it requires the presence of magnetic ions (the formation of local moments) and a mechanism to make them interacting to each other (the exchange interaction). Thus, among the 100,000 inorganic materials documented to date, only about 4,000 are magnetic. An accelerated strategy to find new magnets is then reviewed and discussed by Sanvito and coworkers (▶ Chap. 17, "Machine Learning and High-Throughput Approaches to Magnetism"). This involves a combination of high-throughput electronic structure theory and machine learning techniques. The methods have been proved to be successful at least for relatively simple magnets (metals) and magnetic orders (ferromagnetic and antiferromagnetic). When the materials properties become more complex large-scale materials prediction becomes more difficult, and one has to use all the tools available for a limited number of cases. This is the situation with multiferroic compounds and Rashba semiconductors. In the first case, the compounds present subtle interplays between magnetic and structural degrees of freedom, and in the second, the spin-orbit interaction produces a band structure with spin-dependent features, even in absence of the moment formation. These classes are reviewed by Picozzi (▶ Chap. 18, "Multiferroic and Ferroelectric Rashba Semiconductors").

Magnetic data storage, the hard-disk drive industry, and spintronics, the use of spins for logic operations, are two main drivers behind the development of modern magnetism. In both cases, magnetization dynamics is crucial. This is an

intrinsic multi-scale problem, where the microscopic interactions (Hund's coupling, exchange interaction, and magneto-crystalline anisotropy) determine the macroscopic behavior. Such multi-scale approach is described here by Hertel (▶ Chap. 19, "Applications of Multi-Scale Modeling to Spin Dynamics in Spintronics Devices"), with a particular focus on how the various levels of theory are connected to each other. A crucial aspect of the multi-scale description concerns the mechanisms with which the energy is distributed between spins and vibrations. This provides an important channel for dissipation, allowing the description of relaxation effects and long-time dynamics. A microscopic approach to spin dynamics is provided in the chapter of Evans (▶ Chap. 20, "Atomistic Spin Dynamics"), while Greaves (▶ Chap. 21, "Ultra-fast Dynamics for Heat-Assisted Magnetic Recording") overviews the computation approach to investigate heat-assisted magnetic recording, where heat fluxes are used to manipulate the local magnetization of a thin film.

A second channel for energy dissipation is provided by the interaction between spins and electrical currents. This occurs mainly through exchange interaction and so it pertains fast dynamics. At the atomic scale, current-to-spin interaction can be used to investigate elementary magnetic excitations and the very nature of the magnetic interaction. The main experimental tool here is scanning electron microscopy, while theoretically a master equation approach, sometime informed by electronic structure theory, is the method of choice. Choi and Lorente (▶ Chap. 22, "Magnetic Impurities on Surfaces: Kondo and Inelastic Scattering") provide a complete view of these aspects. Finally, when a current is driven through a nano- and/or meso-scaled magnet, macroscopic magnetization dynamics can be induced. This can be generated by passing the current through a ferromagnet in a vertical geometry (spin-transfer torque) or through materials with strong spin-orbit interaction in a planar configuration (spin-orbit torque). These aspects and their microscopic description are overviewed by Nikolic and coworkers (▶ Chap. 23, "First-Principles Quantum Transport Modeling of Spin-Transfer and Spin-Orbit Torques in Magnetic Multilayers").

2   Conclusion

Magnetism underpins many technologies, which support our modern way of life. These range from magnetic-data storage devices to electrical motors and turbines to nonvolatile memories. Intriguingly, the different technologies exploit different aspects of magnetism, such as the dynamics of the magnetization or the interaction between electrical currents and the spins. The description of these requires a range of theoretical tools spanning over length and time scales and constructed at different level of approximation. This chapter reviews some of the theoretical attempts to understand the magnetic order and provides an overview of how theory and modelling can facilitate, and sometime drive, the development of high-end applications.

Machine Learning and High-Throughput Approaches to Magnetism

17

Stefano Sanvito, M. Žic, J. Nelson, T. Archer, C. Oses,
and S. Curtarolo

Contents

Abstract

Magnetic materials have underpinned human civilization for at least one millennium and now find applications in the most diverse technologies, ranging

S. Sanvito (✉)
Center for Materials Genomics, Duke University, Durham, NC, USA

School of Physics and CRANN Institute, Trinity College, Dublin, Ireland
e-mail: sanvitos@tcd.ie

M. Žic · J. Nelson · T. Archer
School of Physics and CRANN Institute, Trinity College, Dublin, Ireland
e-mail: zicm@tcd.ie; archert@tcd.ie

C. Oses · S. Curtarolo
Center for Materials Genomics, Duke University, Durham, NC, USA

Departments of Mechanical Engineering and Materials Science, Physics, and Chemistry, Duke
University, Durham, NC, USA
e-mail: stefano@duke.edu

from data storage, to energy production and delivery, to sensing. Such great diversity, associated to the fact that only a limited number of elements can sustain a magnetic order, makes magnetism rare and fascinating. The discovery of a new high-performance magnet is often a complex process, where serendipity plays an important role. Here we present a range of novel approaches to the discovery and design of new magnetic materials, which is rooted in high-throughput electronic structure theory and machine learning models. Such combination of methods has already demonstrated the ability of discovering ferromagnets with high Curie temperature at an unprecedented speed.

1   Why New Magnets?

Magnetism is one of the most fascinating macroscopic orders of matter, and it is deeply rooted in quantum mechanics. The Hund's coupling is responsible for the formation of local magnetic moments, m, often localized close to the atomic nuclei, while the exchange interaction, J, makes the moments interacting to each other. This is the m-J paradigm that, together with spin-orbit interaction, generates a multitude of magnetic orders and phenomena.

The formation of the magnetic moment in the solid state is a relatively rare event among the elements of the periodic table. It is possible in $3d$ transition metals and in $4f$ rare-earths, both with an open shell configuration. It is also found, although more rarely, in some $4d$ ions. Light elements, presenting $2p$ valence orbitals, form close shell compounds. Therefore, despite presenting large Hund's coupling, they can sustain a moment only in their radical form, and the possibility of long-range magnetic order remains associated to extremely defective compounds, and it is surrounded by skepticism and debate (Magda et al. 2014). Finally, for heavier elements the valence shell is too delocalized, and m does not form (Janak 1977), with the exception of some low-dimensional structures (Moruzzi and Marcus 1989; Requist et al. 2016).

Once the moments are formed, there are many possible mechanisms to couple them to each other, depending on the details of any given material and in particular on its position with respect to the metal/insulator boundary (Coey 2009). Magnetic moments in metals usually couple by mean of interactions mediated by itinerant electrons, such as the RKKY or the double-exchange ones. These lead to a ferromagnetic order. At the same time, there are mechanisms, such as magnetic super-exchange, which are mostly active in the insulating regime and may produce an antiferromagnetic order. Such rules, however, are not general and the actual magnetic coupling is very sensitive to details. The local chemical environment, the crystal structure, the density of magnetic ions, etc. all may play a critical role so that it is not uncommon to find also ferromagnetic insulators (Wohlfarth 1980) and antiferromagnetic metals (Wadley et al. 2013). In general, however, it is a fact that magnetism is relatively rare in the materials world, with only about 4,000 inorganic compounds among the 100,000 known to date (ICSD 2018) presenting a magnetic order of some kind.

There are several good technological reasons to extend the available menu of magnetic materials. Firstly, there is a growing number of applications that rely on high-performing magnets for which the ideal compound is yet to be discovered or needs to be "rediscovered." Room-temperature applications, of any kind, require a magnetic ordering temperature in excess of 300 °C, a value that ensures stable operation in the temperature range $-50°/+120$ °C. Only about 400 magnets meet such criterion. Furthermore, specific applications dictate several other conditions. For instance, permanent magnets should maximize the magneto-crystalline anisotropy, K_1, and the total magnetization, M_S, so that energy product is large. At the same time, they should be made of elements abundant in the earth crust. In contrast, element criticality is less problematic for the magnets employed by the data storage industry (e.g., Pt is used), but these need to be grown on desired substrates; their grain structure needs to be highly optimized and so does their magnetic hysteresis. Furthermore, changes in technology may generate novel requirements. For instance, it is likely that the advent of heat-assisted magnetic recording may add constrains on the thermal properties of the magnetic recording media.

Sometimes the demand for a new magnet is not dictated by a new technological need. For instance, $Nd_2Fe_{14}B$-based magnets currently cover about 90% of the permanent magnets market, with the remaining share also being largely occupied by rare-earth-containing compounds (e.g., Sm-Co alloys). The average price of a kilogram of neodymium metal was US$8 in 2006, US$56 at the end of 2010, US$450 in 2011, and then fell to US$125 in September 2012. Price volatility is a significant issue for all those applications requiring large volumes of raw elements (an electric vehicle requires about 3 kg of Nd-Fe-B, a direct-drive wind generator about 250 kg), where steady supply is key. Importantly, the rare-earth prize volatility is not associated to fluctuations in the mining production but to commercial strategies between producers and retailers. Such economical geopolitical constraint is the main driver toward a worldwide effort in developing rare-earth-free permanent magnets.

Finally, in addition of being rare and technologically important, magnetism is complex and often defies our physical/chemical intuition. A good example is given by $SrTcO_3$ (Rodriguez et al. 2011). This is a distorted cubic perovskite displaying a G-type antiferromagnetic order with the remarkably high Néel temperature of $T_N = 750$ °C. $SrTcO_3$ is unusual from several points of view. Tc is the magnetic element, one of the few displaying a local moment among the $4d$ transition metal series and highly radioactive. The Néel temperature is exceptionally high with only a handful antiferromagnets having T_N above 500 °C. Finally, the chemical substitution of Tc with elements in its neighborhood in the periodic table produces only low-temperature antiferromagnets ($SrCrO_3$, $SrMnO_3$ and $SrFeO_3$), one low-temperature ferromagnet ($SrRuO_3$), and a diamagnetic compound ($SrMoO_3$). As such, the strong antiferromagnetism of $SrTcO_3$, which is due to a subtle interplay between p-d hybridization and Jahn-Teller distortion (Franchini et al. 2011), appears as a singularity in our magnetic materials landscape.

Given magnetism complexity, the variety of microscopic mechanisms at play, and the subtle sensitivity to details, it is not surprising that the "traditional" trial and error approach to the discovery/design of new magnets has a pretty low throughput. This

has been recently challenged by two completely new approaches to materials discovery, namely, high-throughput electronic structure theory (HTEST) and machine learning (ML) methods. In this new framework, one aims at accelerating the discovery process by performing an extremely large volume of ab initio calculations, which are directed by astute ways for navigating large datasets and identifying patterns among the data. Here we will review such new materials discovery strategy with a particular emphasis on the problems and requirements associated with the magnetic interaction. We will take the discovery of novel magnetic Heusler alloys as prototype, since this is the materials class where most of the work has focused to date (Sanvito et al. 2017). A large part of the methodology described here is based on the AFLOW project, which is discussed in detail in a chapter of the MTM set (volume 3) of this handbook (Chap. 82, "The AFLOW Fleet for Materials Discovery").

2 The High-Throughput Approach to Materials Discovery

2.1 General Principles

In general there are three steps common to all the attempts at HTEST (Curtarolo et al. 2013): (1) the creation of large dataset of computed materials properties, (2) the organization of such calculations in easy-to-access databases, and (3) the screening of the databases in the search of new materials with desired properties or of new rules for designing novel compounds. Let us discuss these three steps in some detail.

Creating large datasets of computed materials properties imposes a number of choices, which somehow determine the further capability of the entire scheme. One strategy consists in producing highly standardized and curated data. Here one decides the level of theory to use and no data are accepted unless they satisfy some stringent criteria. A rather popular choice is that of using density functional theory (DFT) in either the local density or the generalized gradient approximation (LDA and GGA). These ensure enough accuracy, in particular in terms of total energies and elementary linear-response properties, and most importantly the possibility of a high throughput. The convergence parameters (cutoff energies, k-point sampling, force tolerance, etc.) are also highly standardized, so that total energies can be computed across different chemical compositions, structures, and unit cell sizes. Almost all the most popular databases are produced in this way. These include AFLOW.org (Curtarolo et al. 2012b), Materials Project (Jain et al. 2013), OQMD (Kirklin et al. 2015), and 2D (Rasmussen and Thygesen 2015). A second strategy consists in accepting all possible data, regardless of the source (the particular electronic structure theory code used) and the level of theory (e.g., different exchange correlation functionals). The convergence criteria are usually monitored, but stringent tests making different computational platforms compatible with each other are still under development. Typically the data are much more difficult to compare, but the database volume can grow significantly larger than in the case of highly curated data, since it benefits from a large user base. Examples of

this strategy are the Harvard Clean Energy (Hachmann et al. 2011) and the NoMaD projects (Ghiringhelli et al. 2017).

Common to all the strategies is the necessity to deal with large volumes of calculations produced by several users, often with a rather diverse geographical distribution, and the necessity to store the data in easy-to-access databases. This requires efficient managers to run the calculations, check for convergence, create the appropriate entries for the database, and analyze the data. Again one can adopt different strategies for performing this task. On the one hand, there are examples of a single software platform that performs all these tasks, as, for instance, the AFLOW code (Curtarolo et al. 2012a). On the other hand, one can construct a suit of tools (often written in python), which handle the different tasks, as, for instance, the Aiida platform (Pizzi et al. 2016). The main advantage of the first strategy is the efficiency and the high level of curation, while that of the second is the flexibility of the platform. In both cases the databases are usually accessible through a graphical web interface and a back end suitable for data processing.

Creating large datasets of computational data strictly speaking is not a necessity. In principle one can replace those with experimental data and navigate them with a variety of machine learning tools (see the second part of this chapter). However, although the materials science community produces annually an enormous volume of data, these are communicated via scientific publications in a rather narrative form. As such the experimental data are extremely fragmented and difficult to harvest. There is also a multitude of curated databases, but most of them are proprietary and typically focused only on a limited number of properties (e.g., crystal structure, thermochemistry, etc.). Crystallographic databases are probably the only exception, and projects like ICSD (2018) and the crystallographic open database (Grazulis et al. 2009) currently enjoy a great popularity.

Finally, one needs to develop tools to navigate the databases. Here the central role is played by the *descriptors*. A descriptor is an easy-to-calculate property that defines a material. This can be a direct observable (e.g., the quasiparticle gap, the bulk modulus, etc.) or a quantity that acts as a proxy for a given property, such as the effective mass as a proxy for the mobility of a semiconductor. Descriptors are constructed directly from the elementary electronic properties computed with the chosen electronic structure theory. Their definition usually reflects the understanding of a property from a given community. Examples of descriptors are the *high-throughput topological insulator robustness* for the discovery topological insulators (Yang et al. 2012), the *GIBBS* method for thermal conductivity (Toher et al. 2014), the *spectroscopic limited maximum efficiency* of light-harvesting materials (Yu and Zunger 2012), etc.

2.2 Constructing Magnetic Libraries: Heusler Compounds

In short the HTEST strategy to the discovery of new materials consists in generating large libraries of hypothetical compounds and then in evaluating their properties, including the likelihood of these compounds to be made. Again, the criteria used to generate such initial libraries vary significantly depending on the targeted
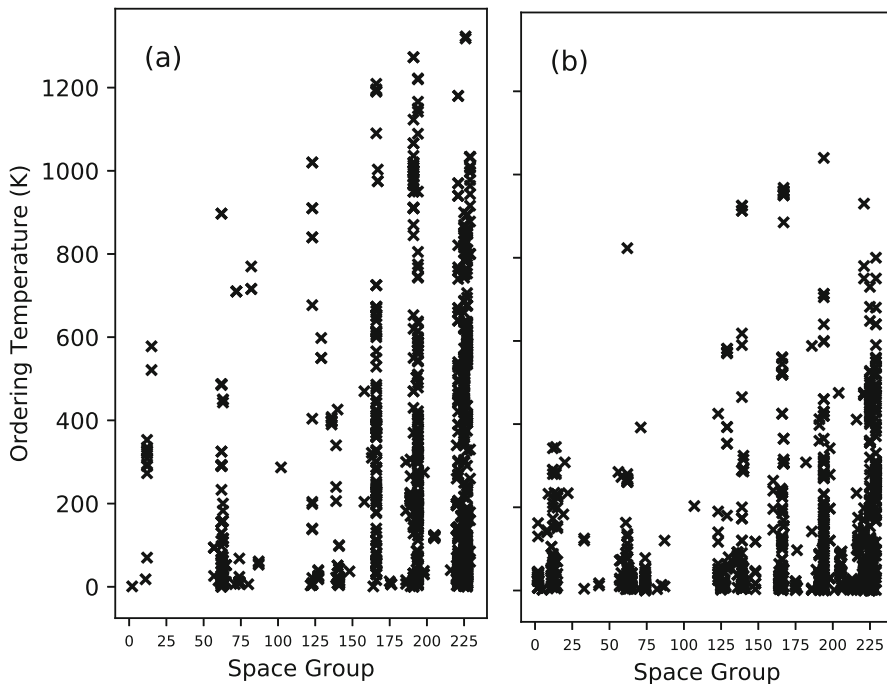
Fig. 1 Critical ordering temperature, either Curie or Néel, for known (**a**) ferromagnets and (**b**) antiferromagnets as a function of the materials space group. Note that magnetic order seems to be more favorable for high space group numbers, namely, for high-symmetry cubic crystals. This is the region in the space group space where we find the magnets with the highest critical temperatures

properties. For instance, one can select a specific crystal structure and try all possible chemical compositions compatible with that structure, sometimes imposing additional constraints such as the total valence electron count (Yan et al. 2015). Magnetic compounds are distributed over a broad range of space groups (see Fig. 1) and, in addition to the limited number of magnetic ions, contain a large variety of nonmagnetic elements. As such establishing a stringent criterion for generating a library of magnetic prototypes is not straightforward. One possible criterion is to look at classes of compounds where the incidence of magnetism is high, as, for instance, in transition metals/rare-earth intermetallics (Dam et al. 2017).

Here we will describe a less selective criterion, namely, a library of Heusler alloys (Sanvito et al. 2017), a prototypical family of ternary compounds populated with several high-performance magnets (Graf et al. 2011). In their most common form, Heusler alloys are cubic compounds consisting of four interpenetrating *fcc* lattices. The so-called *regular* Heusler alloy, X_2YZ (Cu$_2$MnAl-type), crystallizes in the $Fm\bar{3}m$ cubic space group (No. 225), with the X atoms occupying the 8c Wyckoff position (1/4, 1/4, 1/4) and the Y and Z atoms being, respectively, at 4a (0, 0, 0) and 4b (1/2, 1/2, 1/2). The Y and Z elements form an octahedral-coordinated rock-salt structure, while the X atoms occupy the tetrahedral voids
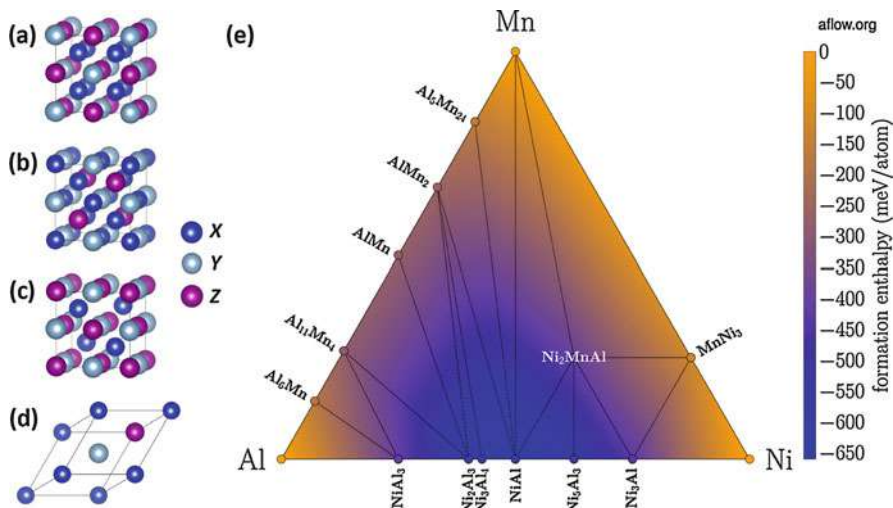
Fig. 2 Possible Heusler alloys: (**a**) regular Heusler, (**b**) inverse Heusler, and (**c**) half Heusler. In panel (**d**) we show the tetrahedral $F\bar{4}3m$ cell used to construct the electronic structure database. (**e**) Ternary convex hull diagram for Al-Mn-Ni. Note the presence of the stable HA, Ni$_2$MnAl. (Figure from Sanvito et al. 2017)

(see Fig. 2a). Alternatively, in the *inverse* Heusler structure, $(XY)XZ$ (Hg$_2$CuTi-type), now X and Z form the rock-salt lattice, while the remaining X and Y atoms fill the tetrahedral sites (Fig. 2b), so that one X atom presents sixfold octahedral coordination, while the other has fourfold tetrahedral coordination. Finally, one can also have the so-called *half* Heusler alloy, XYZ (MgCuSb-type). This is obtained by removing one of the X atoms from the regular type, thus leaving a vacancy at one of the tetrahedral sites (Fig. 2c). The Wyckoff positions are now $4a$ (0, 0, 0), $4b$ (1/2, 1/2, 1/2), and $4c$ (1/4, 1/4, 1/4), respectively, for X, Y, and Z. The minimal unit cell describing all three types can be constructed as a tetrahedral $F\bar{4}3m$ cell, containing four (three for the case of the half Heusler) atoms (Fig. 2d). Such a cell allows for a ferromagnetic spin configuration and for a limited number of antiferromagnetic and ferrimagnetic ones.

Note that it is not uncommon for Heusler alloys to have a significant amount of site-occupation disorder. In this case there is a special classification depending on the particular sites where the disorder takes place, as illustrated in Table 1 for the complete Heusler alloys and in Table 2 for the inverse ones. Modeling site-occupation disorder is a notoriously challenging problem in ab initio studies due to computational costs, although recent developments look promising for future investigations (Yang et al. 2016). Finally, it is worth mentioning that there exist a number of Heusler alloys presenting significant tetragonal distortion. These are usually associated to Mn-containing compounds, in particular with the Mn$_2YZ$ composition (Faleev et al. 2017), and they are interesting since they can potentially display large magneto-crystalline anisotropy.

Table 1 Site occupancy and general formula for different atomic orders of Heusler compounds. The notations according to the Inorganic Crystal Structure Database (ICSD), the Strukturberichte (SB), the Pearson database, as well the space group are given. (Table adapted from Graf et al. 2011)

Site occupancy	Formula	Type (ICSD)	SB	Pearson	Space group
X, X', Y, Z	$XX'YZ$	LiMgPdSn	Y	cF16	$F\bar{4}3m$ (No. 216)
$X = X', Y, Z$	X_2YZ	Cu_2MnAl	$L2_1$	cF16	$Fm\bar{3}m$ (No. 225)
$X, X' = Y, Z$	$XX_2'Z$	$CuHg_2Ti$	X	cF16	$F\bar{4}3m$ (No. 216)
$X = X' = Y, Z$	X_3Z	BiF_3	DO_3	cF16	$Fm\bar{3}m$ (No. 225)
$X = X', Y = Z$	X_2Y_2	CsCl	B2	cP2	$Pm\bar{3}m$ (No. 221)
$X = Y, X' = Z$	X_2X_2'	NaTl	B32a	cF16	$Fd\bar{3}m$ (No. 227)
$X = X' = Y = Z$	X_4	W	A2	cI2	$Im\bar{3}m$ (No. 229)

Table 2 Site occupancy and general formula for differently ordered half-Heusler compounds. The notations according to the Inorganic Crystal Structure Database (ICSD), the Strukturberichte (SB), the Pearson database, as well the space group are given. Wyckoff position 4d (3/4, 3/4, 3/4) denotes the second tetrahedral lattice site, which is void in ordered materials. (Table adapted from Graf et al. 2011)

Site occupancy	Formula	Type (ICSD)	SB	Pearson	Space group
$4a, 4b, 4c$	XYZ	LiAlSi	$C1_b$	cF16	$F\bar{4}3m$ (No. 216)
$4a = 4b, 4c$	XZ_2	CaF_2	C1	cF12	$Fm\bar{3}m$ (No. 225)
$4a, 4b, 4c = 4d$	X_2YZ	Cu_2MnAl	$L2_1$	cF16	$F\bar{4}3m$ (No. 216)
$4a = 4b, 4c = 4d$	XY	CsCl	B2	cP2	$Pm\bar{3}m$ (No. 221)
$4a = 4c, 4b = 4d$	YZ	NaTl	B32a	cF16	$Fd\bar{3}m$ (No. 227)
$4a = 4b = 4c = 4d$	X	W	A2	cI2	$Im\bar{3}m$ (No. 229)

We have constructed a library of potential Heusler alloys by considering all possible combinations of three elements chosen from the $3d$, $4d$, and $5d$ periods and from the groups III, IV, V, and VI. These include Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Cl, Co, Cr, Cu, Fe, Ga, Ge, Hf, Hg, In, Ir, K, La, Li, Mg, Mn, Mo, Na, Nb, Ni, Os, P, Pb, Pd, Pt, Re, Rh, Ru, Sb, Sc, Se, Si, Sn, Sr, Ta, Tc, Te, Ti, Tl, V, W, Y, Zn, and Zr. We have thus avoided rare-earth elements, since in the last decade their price has been volatile on the market, so that the search for rare-earth-free magnets has become a sought-after target. If one considers all the possible stoichiometry compatible with the elements chosen and all the available crystal structures (regular, inverse, and half Heusler), a total number of 236,115 prototypes will be reached. Their electronic structure has been computed at the AFLOW standard, namely, by DFT with the GGA parameterized by Perdew et al. (1996). The DFT platform is VASP (Kresse and Furthmuller 1996), and the convergence criteria is quite stringent, namely, the convergence tolerance is 1 meV/atom and the typical Brillouin zone sampling is over a dense grid of 6,000–10,000 k-points per reciprocal atom (Calderon et al. 2015). The structures are fully relaxed, and in general the electronic structure is initialized in such a way to have the maximum possible spin. This gives a unit-cell ferromagnetic initial configuration whenever

the unit cell contains more than one magnetic ion. The calculations are performed without spin-orbit interaction. More details about the computational method are reported in Calderon et al. (2015).

2.3   The Descriptors

2.3.1   Energy-Related Descriptors

The very first property that one usually wishes to monitor in the HTEST protocol is the likelihood that a particular compound can be grown. The appropriate descriptor is then the Gibbs free energy, $G = H - ST$, where H is the enthalpy, S the entropy, and T the temperature. If G can be calculated, then all thermodynamically stable compounds can be identified, leaving out only those possible in the form of long-living metastable phases. However, the calculation of the Gibbs free energy is too demanding to be carried out in a high-throughput framework, so that in general one chooses to evaluate quantities related only to the DFT total energy.

A very crude but yet selective criterion is given by the enthalpy of formation. For the Heusler X_2YZ one has to check that its enthalpy, H_{X_2YZ}, is lower than the sum of the enthalpies of formation of its elementary constituents, namely, $\Delta H = H_{X_2YZ} - (2H_X + H_Y + H_Z) < 0$. For instance, for Ni$_2$MnAl, a now well-established magnetic shape memory alloy (Ziebeckt and Webster 1975), one has to check that $H_{\text{Ni}_2\text{MnAl}}$ is lower than the sum of the enthalpies of formation of *fcc* Ni, Mn, and Al, their energy-lowest structures. As a further approximation, the enthalpy of formation is replaced by the DFT total energy.

A second, much more stringent, stability criterion is obtained by constructing for every three elements the convex hull diagram of the associated ternary compounds (Lukas et al. 2007), *Descriptor 1*. This essentially consists in establishing whether the enthalpy of formation of X_2YZ is lower than that of all the possible binary and elementary decompositions, for instance, $XY + XZ$, $X_2Y + Z$, $XYZ + X$, etc. Again by taking the example of Ni$_2$MnAl, we find that it is stable against all possible decompositions along the Al-Mn-Ni convex hull diagram, which is presented in Fig. 2e. The most stable structure is the regular Heusler with a formation energy of -404 meV/atom, but there are also three unstable ones with $\Delta H < 0$, namely, Mn$_2$NiAl ($\Delta H = -209$, 121 meV/atom above tie-plane), NiMnAl ($\Delta H = -39$, 400 meV/atom above tie-plane), and Al$_2$MnNi ($\Delta H = -379$, 100 meV/atom above tie-plane). Such three unstable structure, all with $\Delta H < 0$, clearly prove that the enthalpy of formation alone says little about the stability of a given compound. It is also worth to mention that even the convex hull, as calculated so far, provides only an incomplete descriptor for the thermodynamical stability. In fact, in principle one has also to check possible decompositions involving alternative ternary compounds. Furthermore, once a ternary material is predicted stable, there is still the possibility that the ground state crystal structure is not a Heusler type. As such, once a stoichiometry has been predicted stable against decomposition, one should perform further crystal structure prediction calculations. Possible methods

include random sampling (Pickard and Needs 2011), genetic algorithms (Oganov and Glass 2006), or targeted sampling (d'Avezac et al. 2012).

One has to consider that the probability of finding new ground states per alloy reduces with increasing the number of species. This can be seen by performing a phase stability experiment in the thermodynamic data available in the AFLOW.org repository (Curtarolo et al. 2012b), which is performed by calculating all the available binary, ternary, quaternary, etc. convex hulls (Lukas et al. 2007). The enthalpies can then be analyzed, and the average gain in enthalpy with respect to the number of species N, $\delta H(N) \equiv H(N) \rightarrow (N + 1)$, decreases with N. This quantity competes against the disorder state, unavoidably promoted by configurational entropy going as $\approx \log(N)$. If follows that, in our case ternary case, the existence/absence of a ternary compound depends *mostly* on decomposition in binaries than in competition by other ternaries. As such, a first-order stability analysis, based on simply checking binaries, will be an appropriate approach capable to weed out most of the uninteresting compositions.

Additional energy-related descriptors can be defined, depending on the properties and materials class that one wants to investigate. In general, it is useful to have an idea on how robust is the thermodynamical stability of a compounds. This can be estimated by calculating the enthalpy of formation relative to that of the most stable balanced decomposition. If such difference is around $k_B T$ at room temperature (k_B is the Boltzmann constant, $k_B T_{RT} = 25$ meV), then one usually defines the material as robustly stable (Sanvito et al. 2017). This descriptor, *Descriptor 2*, is denoted as Δ^{kT}.

Finally, considering that Heusler alloys are often prone to site-occupation misplacement, it is useful to define a descriptor, *Descriptor 3*, which identifies the tendency to disorder. This is the entropic temperature (Curtarolo et al. 2013; Hart et al. 2013; Yong et al. 2014), T_S, which is defined for a binary XY alloy as

$$T_S = \max_i \left[\frac{\Delta H(X_{x_i} Y_{1-x_i})}{k_B [x_i \log x_i + (1 - x_i) \log(1 - x_i)]} \right], \tag{1}$$

where i counts all the stable compounds in the XY binary system. In practice T_S measures the ability of an ordered phase to resist deterioration into a temperature-driven, entropically promoted, disordered mixture. This is quantified by the concentration-maximized formation enthalpy weighted by the inverse of its ideal entropic contribution (random alloy). By convention we choose the sign of T_S so that a positive temperature is needed for competing against the compound stability ($T_S < 0$ if $\Delta H > 0$), and one expects $T_S \rightarrow 0$ for a compound spontaneously decomposing into a disordered mixture.

In Eq. (1), the curvature of the iso-max ideal-latent heat locus can be seen as a descriptor of the (ideal) entropy that a system can absorb upon nucleation in a single phase and internal reorganization containing ideal disorder. As such, it is a viable descriptor to tackle the unavoidable entropy stabilization happening in half Heuslers (entropy of vacancies) or in full Heuslers (entropy of species permutations). In these

systems, disorder might overcome enthalpy stabilization and lead to stable solid solutions unobtainable with prediction algorithms mentioned above.

2.3.2   Magnetic Descriptors

A magnet can be defined as "high-performing," when it presents a range of electronic and magnetic properties particularly suitable for a given application. Thus, for instance, a high-performing permanent magnet will need to score high against descriptors different from those of a magnet designed for magnetocaloric applications (Coey 2009). Some magnetic descriptors are observables and can be extracted directly from the DFT calculations. This is, for instance, the case of the magnetic moment, *m* (*Descriptor 4*), which can be directly compared with the experimental saturation magnetization. In the case the magnet is screened for spintronics applications (for magneto-transport), it is important to assess its spin polarization at the Fermi level. This is defined as (*Descriptor 5*)

$$P_{\mathrm{F}}^{\alpha} = \frac{n_{\mathrm{F}}^{\uparrow}(v_{\mathrm{F}}^{\uparrow})^{\alpha} - n_{\mathrm{F}}^{\downarrow}(v_{\mathrm{F}}^{\downarrow})^{\alpha}}{n_{\mathrm{F}}^{\uparrow}(v_{\mathrm{F}}^{\uparrow})^{\alpha} + n_{\mathrm{F}}^{\downarrow}(v_{\mathrm{F}}^{\downarrow})^{\alpha}} \ , \tag{2}$$

where n_{F}^{σ} and v_{F}^{σ} are, respectively, the density of states (DOS) at the Fermi level and the Fermi velocity for the spin-σ sub-band. In general, P_{F}^{α} describes the degree of spin polarization of the current that the material can sustain in different transport conditions (Mazin 1999). Here $P = 1$ indicates that only one spin species is responsible for the electron transport (Coey and Sanvito 2004). Note that different values of α characterize different transport experiments. For instance, $\alpha = 0$ is for tunneling, $\alpha = 1$ for ballistic transport, and $\alpha = 2$ for diffusive. Note also that P_{F}^{α} is a direct observable only for $\alpha = 0$ (e.g., with a spectroscopical measurement).

There are also magnetic descriptors, whose definition requires additional DFT calculations, possibly combined with other theories. This is, for instance, the case of the magneto-crystalline anisotropy (MCA – *Descriptor 6*), which measures the ability of a magnet to maintain the magnetization direction fixed in space. The MCA contributes to the hysteresis of a magnet, but it is difficult to isolate from other contributions to the total anisotropy (e.g., shape anisotropy). Computing the MCA is in general involved as it requires highly accurate total energy or perturbation theory calculations including spin-orbit interaction (Błoński and Hafner 2009).

Finally, common to all magnets is the critical ordering temperature, T_{C} (Curie or Néel), *Descriptor 7*. This requires knowledge of the compound magnetic excitation spectrum, namely, of the susceptibility, χ. There are several ways to determine such quantity and all require significant additional work. One of such possibilities is that of using time-dependent DFT (Savrasov 1998), which is capable of describing both Stoner excitations and spin waves. In the case the excitation spectrum is dominated by spin waves, a popular choice is that of performing a series of DFT total energy calculations for different magnetic configurations and mapping these onto a Heisenberg model of some kind. There are several different approaches to perform both the mapping and the total energy calculations, including magnetic

force theorem (Oswald et al. 1985), spin spirals (Sandratskii 1986), and collinear multi-configurations (Archer et al. 2011). Finally, usually T_C is computed either by using mean-field theory or with Monte Carlo simulations.

2.4   Analysis

Out of the 236,115 prototypes contained in our Heusler alloys library, only 35,602 have a DFT energy lower than the sum of that of their elementary phases, namely, they satisfy the condition $\Delta H < 0$. These 35,602 include 6,778 compounds with a magnetic ground state. Such an estimate can be slightly conservative. In fact we are able to describe only those magnetic phases, where the magnetic cell is the same as the primitive one. This clearly excludes many potential magnetic solutions, in particular complex antiferromagnetic and ferrimagnetic phases, which are often found in some of the magnetic Heuslers (Rode et al. 2013). Furthermore, we may expect that the ferromagnetic DFT initialization may force final ferromagnetic solutions to otherwise antiferromagnetic ground states, i.e., the DFT run may converge to a local minimum. In any case 6,778 potentially new magnets is an extraordinary large number, since only 4,000 magnets are known to date.

The application of *Descriptor 1*, however, reduces drastically such number. Since the construction of the ternary convex hull diagrams requires a significantly large number of calculations, we have initially limited our analysis to intermetallic prototypes, namely, to Heusler structures made of $3d$, $4d$, and $5d$ elements only. These are 36,540 in total, but only 248 are on the convex hull; hence only 248 are predicted stable. Most importantly only 20 have a magnetic ground state, meaning that the incidence of stable magnetism in our Heusler library has to be estimated to be around 0.05%. A close look at the newly predicted magnetic Heusler alloys reveals that, with the exception of Rh_2FeZn, only three families are found, namely, Co_2YZ, Mn_2YZ and X_2MnZ.

The Co_2YZ family is perhaps the most interesting; it contains already 25 known compounds all lying on the Slater-Pauling curve (Graf et al. 2011) and houses several half-metals. Our search found four new stable alloys, namely, Co_2VZn, Co_2NbZn, Co_2TaZn, and Co_2MnTi. The first three have the low valence electron count of 25, while for Co_2MnTi this is large, 29. The numerical regression used to evaluate T_C correctly places these four on the Slater-Pauling curve, as shown in Fig. 3, and predicts Co_2MnTi to have the remarkable T_C of 940 K. The other three new compounds have all a T_C around 200 K, but two of them become nonmagnetic upon tetragonal distortion leaving only Co_2VZn magnetic (predicted $T_C \sim 228$ K).

Five Mn_2YZ compounds populate the second family, namely, Mn_2PtRh, Mn_2PtCo, Mn_2PtPd, Mn_2PtV, and Mn_2CoCr. In general Mn_2YZ alloys are found to crystallize in the regular Heusler structure if the atomic number of the Y ion is smaller than that of Mn, $Z(Y) < Z(Mn)$, and the inverse one for $Z(Y) > Z(Mn)$. This, however, is valid when one element from the main groups occupies the $4c$ position. In our case of intermetallic Heuslers, we find that the regular $Fm\bar{3}m$ structure is always the ground state, regardless of the chemical composition. Furthermore, all the new compounds present some degree of antiferromagnetic
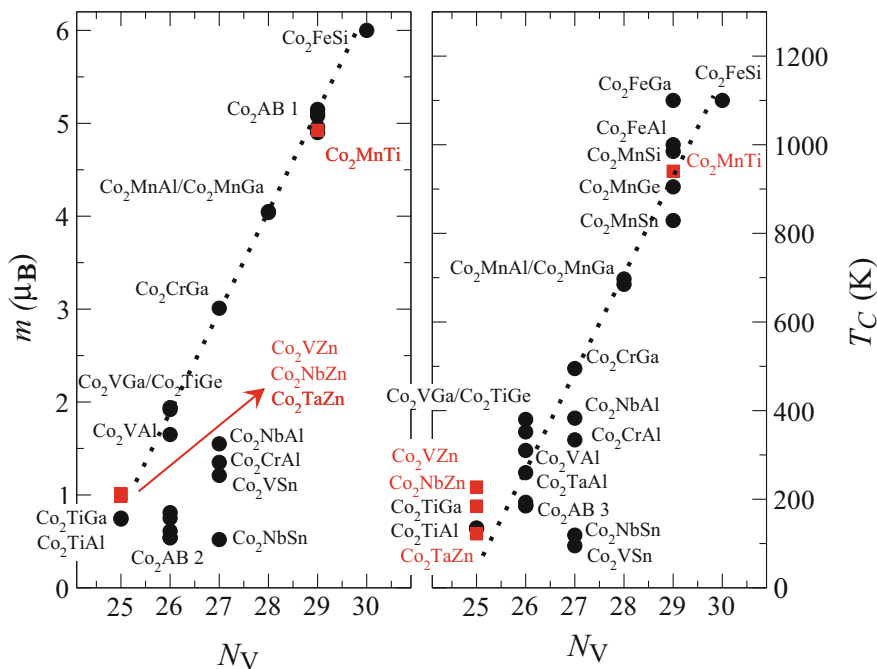
Fig. 3 Slater-Pauling curve for magnetic Heusler alloys of the form Co_2YZ. The magnetic moment per formula unit, m, is plotted against the number of valence electron, N_V, in the left panel, while T_C is displayed on the right. Red symbols correspond to predicted HAs while the black ones to existing materials. For the sake of clarity, several compounds have been named collectively on the picture. Co_2AB 1 (Co_2FeGa, Co_2FeAl, Co_2MnSi, Co_2MnGe, Co_2MnSn), Co_2AB 2 (Co_2TaAl, Co_2ZrAl, Co_2HfAl, Co_2HfGa, Co_2TaGa), Co_2AB 3 (Co_2ZrAl, Co_2HfAl, Co_2HfGa, Co_2TaGa). (Figure from Sanvito et al. 2017)

coupling, which results in either a zero-moment ground state when Mn is the only magnetic ion, and in a ferrimagnetic configuration when other magnetic ions are present. No estimate of T_C is possible in this case since there are no data for regular Mn_2YZ alloys to run the regression against.

Finally, we have found several X_2MnZ compounds (13), most of them including a $4d$ ion (Ru, Rh, and Pd) in the tetrahedral X position. In general, these compounds have a magnetic moment per formula unit ranging between 4 and 5 μ_B, consistent with the nominal 2+ valence of Mn in octahedral coordination. The regression, run against 18 existing compounds of which 13 are with X=Ru, Rh, or Pd, establishes a correlation between the Mn-Mn nearest neighbor distance, d_{Mn-Mn}, and T_C. This effectively reproduces the empirical Castelliz-Kanomata curves (Castelliz 1955; Kanomata et al. 1987) and returns us the prediction that none of the X_2MnZ alloys will display a T_C exceeding 500 K.

The final and ultimate validation of the high-throughput approach must come from an experimental lab. With this in mind, we have attempted the growth of four newly predicted alloys, namely, Co_2MnTi, Mn_2PtPd, Mn_2PtCo, and

Mn$_2$PtV. Co$_2$MnTi is chosen because of its high Curie temperature, while among the Mn$_2$-based alloys, we have selected two presenting a ferrimagnetic ground state (Mn$_2$PtCo and Mn$_2$PtV) and one meeting the stringent robustness criterion (Mn$_2$PtPd). Details of the growth and magnetic characterization can be found in Sanvito et al. (2017).

Two of the four Heusler alloys have been successfully synthesized, Co$_2$MnTi and Mn$_2$PtPd, while the other two, Mn$_2$PtCo and Mn$_2$PtV, decompose into binary compounds (notably in both cases, one of the products of the decomposition was an intermetallic with a very large unit cell, and it was not included in our database). Co$_2$MnTi was found to crystallize in the regular $Fm\bar{3}m$ Heusler structure with a lattice parameter, $a = 5.89$ Å in close agreement to the predicted one, $a = 5.84$ Å. Most strikingly the T_C extrapolated from the zero-field cooled magnetization curve in a field of 1 T is found to be 938 K, essentially identical to that predicted by our regression, 940 K. This is a remarkable result, since to our knowledge it is the first time that a new high-temperature ferromagnet has been discovered by high-throughput means.

Also Mn$_2$PtPd was found to be a single phase without evidence of decomposition. X-ray diffraction reveals that the structure is that of a tetragonally distorted regular Heusler (*I4/mmm* – TiAl$_3$-type) with lattice parameters $a = 4.03$ Å and $c = 7.24$ Å. Magnetization curves at room temperature and 4 K show no hysteresis or spontaneous magnetization indicating that the compound is antiferromagnetic at low temperature.

3   Machine Learning for Materials Discovery

The general concept underpinning machine learning methods for materials science is relatively simple. One wishes to establish relations between known microscopic features, including chemical and perhaps structural information, with a number of physical/chemical properties. In practice the machine learning model, f_{ML}, defined as

$$f_{ML} : \vec{v} = (\{Z_i\}, N_V, \ldots) \rightarrow p, \tag{3}$$

associates to a *feature vector*, $\vec{v}$, the property p. Then, $\vec{v}$ contains information such as the atomic numbers, $\{Z_i\}$, of the elements in a compound; the number of valence electrons, N_V; etc. The machine learning models are purely numerical and do not require any particular physical knowledge of the property under investigation. There is a multitude of algorithms available, going from simple linear regression to complex neural network schemes (Shalev-Shwartz and Ben-David 2014; Hastie et al. 2013). These can deliver a classification, namely, they can sort materials in classes or the continuous value of a given property (e.g., the Curie temperature of a ferromagnet). The standard procedure in creating a machine learning model is to train the model over a given subset of data, the *training set*, and then to evaluate the quality of the learning by predicting the property of data never seen before, the *test set*.

There are a number of issues related to machine learning applied to materials science, which are relatively general to any algorithm and property. The first concerns the many possible ways to represent the feature vector, which should satisfy a number of criteria. In particular the feature vector should be continuous for small variations in the features, with two materials having the same representation only if their target properties are identical. This criterion essentially establishes a notion of similarity across the materials space. For instance, if one seeks at describing a property that strongly depends on the number of valence electrons, elements are represented by their group in the periodic table better then by their atomic number. In fact, Si and Ge will have the same representation when using the group (both in group IV), while they will appear as rather distant if the atomic number is considered (14 for Si and 32 for Ge). The representation of the crystal structure poses particular problems. The space group, for instance, is difficult to use, since tiny distortions from a high-symmetry structure (high space group number) may result in a space group with a rather low number. Furthermore, whatever representation one chooses, it should be both translational and rotational invariant and should be insensitive on the choice of the unit cell (for solids). Note that these criteria associated to the feature vector are not stringent conditions in general. In principle a solid machine learning model based on an extremely large number of data should be able to establish patterns among the data even if the distance between different feature vectors does not follow an intuitive physical/chemical metric. Unfortunately the materials space is not large, so that using feature vectors with the properties described above is highly desirable.

The second issue is related to the so-called curse of dimensionality, which establishes that as the dimension of the feature space gets larger, the number of possible distinct configurations increases exponentially. In other words the density of points needed to extrapolate the data in a multidimensional space grows exponentially with the dimension of the space itself. This essentially means that in order to obtain the same quality of learning, the number of data needed grows exponentially with the dimension of the feature vector, i.e., adding more information to the feature vector may decrease the change to establish patterns in the data. In materials science the curse of dimensionality is a serious issue for two main reasons. Firstly, the chemical materials space is large but not huge. Secondly, the availability of experimental data is limited by the fragmentation of the various databases, and the fact that results are communicated in scientific publications, which are little data oriented. This has two main consequences. Firstly, most of the machine learning studies to date train on theoretical data (as the ones contained in AFLOWLIB.org) (Isayev et al. 2017) and not experimental ones, though predictions show to be consistent with the experimental results. Secondly, one has to design machine learning algorithms, where the dimension of the feature vector remains small (Ghiringhelli et al. 2015).

When applied to magnetism, the use of machine learning models remains very limited. Here we will present two illustrative examples, looking, respectively, at predicting the magnetic moment of Fe-containing Heusler alloys and at sorting tetragonally distorted ones into hard and soft magnets. Other examples of machine learning approaches to magnetism include the identification of novel magnetic
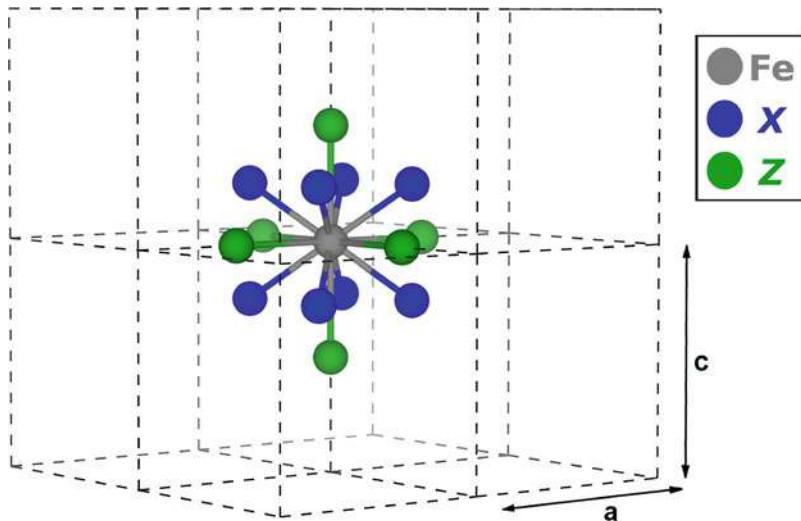
Fig. 4 The local coordination of the atomic sites in a Heusler alloy. The neighbors of the central Fe atom form two shells of different symmetry. Atoms belonging to the nearest neighbor shell, shown in blue, coordinate the central atom tetrahedrally. The next nearest neighbor shell is made out of six (green) atoms and has octahedral symmetry. (Figure adapted from Žic et al. 2017)

phases with enhanced magnetic anisotropy (Kusne et al. 2014), the discovery of intermetallic compounds (Oliynyk and Mar 2018), and the prediction of the Curie temperature of transition-metal rare-earth compounds (Dam et al. 2017).

3.1   Magnetic Moment Predictions

We will now present a machine learning strategy to predict the magnetic moment of $L2_1$ Fe-containing Heusler alloys (Žic et al. 2017) of the $X_2\mathrm{Fe}Z$ type. In general the magnetic moment of $3d$ transition metals is highly localized and can be described as atomic-like. This may get modified by the surrounding ions, which should be then included into the definition of a suitable feature vector. We represent the local structure by constructing clusters centered around the Fe ions as illustrated in Fig. 4. The center of the cluster is occupied by Fe, the first coordination shell is populated by ions X with atomic number Z_1, while the second nearest neighbors are ions Z of atomic number Z_2.

A feature vector can then be defined as the following:

$$\vec{v} = (\{Z_i\}, \{r_{0i}\}, \{N_i\}, a_{\mathrm{lat}}), \tag{4}$$

where $\{Z_i\}$ are the atomic number of the ions in the cluster ($i = 1, 2$), $\{N_i\}$ is their valence, r_{0i} is the distance between the central Fe ion and the i-th one, while a_{lat} is the lattice constant of the relative Heusler compound.
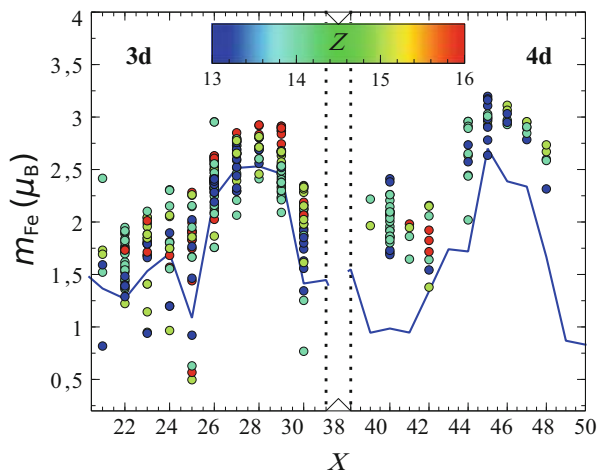
Fig. 5 *Left panel* – Magnetic moment of Fe, m_{Fe} (in μ_B), for a wide range of nearest neighbors at a constant Wigner-Seitz volume ($R_{WS} = 2.7$ bohr). The atomic number of the next nearest neighbor, Z_3, is color coded, while here we plot data as a function of the first nearest neighbor atomic number, Z_1. We can notice a linear increase of the magnetic moment across the transition metal series which does not depend on Z_3. The symbols are the DFT data, while the corresponding machine learning trend is shown with the blue line. *Right panel* – a data sample containing a wider range of main group elements. The data elucidates the origin of the oscillation in the machine learning trend throughout the main group series. (Figure adapted from Žic et al. 2017)

We have then used such feature vector in a random forest regression algorithm trained over a large number of Heusler alloys, whose electronic structure was computed at the level of GGA-DFT (Žic et al. 2017). In fact, we have constructed a number of machine learning models using feature vectors similar to that presented here and capable of predicting the magnetic moment of an atom at the center of the cluster of any type (not just Fe). After having noticed that the Fe magnetic moment is little affected by the choice of Z_2, as long as this crystal position is occupied by an element belonging to one of the main groups, we have focused our attention on the dependence of the Fe moment, m_{Fe}, over the Z_1. In general we have discovered that the valence of X determines m_{Fe}. In fact, if this is smaller than 8, then the moment will decrease; otherwise it will increase. This trend is true for all the elements in the $3d$ and $4d$ periods, as shown in Fig. 5, so that the largest Fe moment is found when $X = $ Ni and Pd.

3.2   Anisotropy Analysis: Saving Computational Time

As a second example of machine learning methods applied to magnetism, we describe here the construction of a classification scheme to sort out hard from soft magnets (Žic 2017). The magnetic hardness of a compound is determined, at

the microscopic level, by the magneto-crystalline anisotropy (Błoński and Hafner 2009). This is a well-defined quantity, which depends on the spin-orbit coupling strength, the crystal structure, and the magnetic ion responsible for the magnetic moment. In general it can be computed either by using perturbation theory or by total energy differences. In both cases the calculation is delicate as it required an extremely accurate evaluation of the Fermi surface. In fact, the anisotropy is very sensitive to the occupation of the various d orbitals so that one needs to determine the position of the Fermi energy at a high degree of accuracy. As a result sampling the k-points over a very fine mesh is mandatory, and the calculations end up to be quite time-consuming. Clearly, if one can construct a machine learning strategy to the calculation of the magneto-crystalline anisotropy, the high-throughput analysis of magnetic materials will become more feasible.

Our dataset was constructed from a Heusler alloys library, from which we have selected about 300 compounds presenting a magnetization exceeding $0.5\,\mu_B$/fu and tetragonal distortion $c/a \neq 1$. In this case the thermodynamical stability was not considered so that the dataset includes both stable and unstable structures. The magnetic anisotropy was then calculated at the level of DFT-GGA using magnetic force theorem. These initial calculations have returned us magneto-crystalline anisotropy values ranging between 0 and $15\,\mathrm{MJm}^{-3}$, although for most of the compounds, the anisotropy is smaller than $8\,\mathrm{MJm}^{-3}$.

We have then constructed a machine learning model based on a ridge classifier (Pedregosa et al. 2011) and different choices of feature vector. In particular the best model uses the following vector:

$$\vec{v} = (\{Z_i\}, \{N_i\}, \{\zeta_i\}, \{M_i\}, V), \tag{5}$$

where Z_i, N_i, and M_i are the atomic number, the number of valence electron, and the local magnetic moment of the different species in X_2YZ, respectively. In addition V is the unit cell volume and ζ_i is a quantity associated to the spin-orbit coupling strength (Dunn 1961). Importantly, the magnetic moments are estimated using a generalization of the machine learning model described in the previous section (the central atom does no longer need to be Fe) (Žic et al. 2017), so that the feature vector does not comprise any quantities computed ab initio. We have then set the magneto-crystalline anisotropy value of $0.8\,\mathrm{MJm}^{-3}$ as the boundary between hard and soft magnets. This choice ensures that the population of hard and soft magnets in our dataset is identical.

Although hard magnets are our target, we have decided to train the machine learning model so to identify the soft ones. The reason behind such decision is primarily related to the fact that hard magnets are scarce and the cost of unintentionally dismissing a viable candidate is high. This means that a very selective machine learning model will be able to predict reliably hard magnets but will likely discharge many of them. We therefore choose a strategy where we maximize the number of correctly identified soft magnets, namely, the true positive rate (TPR), under the constraint that no hard magnet is mistaken for a soft magnet, i.e., we aim to keep the false positive rate (FPR) as small as possible. This will select the soft compounds and discharge both soft and hard.

Fig. 6 Receiver operating curve (ROC) for the two hard magnet classifiers discussed in the text (left). The area under the curve (auc) is indicated in the brackets. The machine learning model described in the text is clf 2, while clf 1 represents a second model with a different choice of feature vector. The dashed red line is for a random classifier
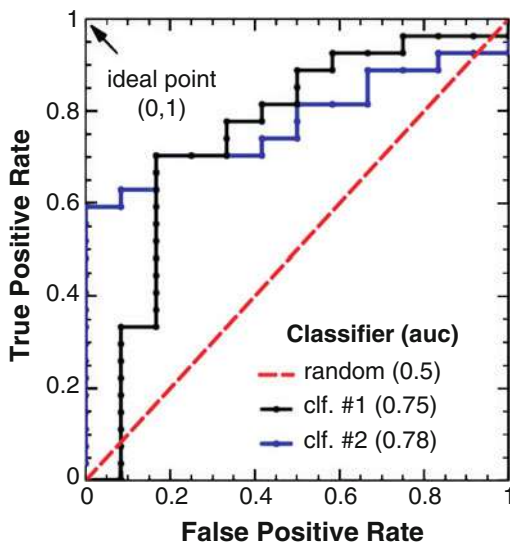
In Fig. 6 we present the receiver operating curve (ROC) for the machine learning model described in the text (clf 2), for a second one constructed with a different feature vector (clf 1) and for a random classifier (dashed red line). One can move along the ROC by changing the decision value cutoff. Ideally one wishes the model to be at the top left corner of the ROC, namely, to have a true positive rate of 100% (all compounds with the desired property are selected) and a false positive rate of 0 (no compounds are misclassified). In a realistic machine learning model, this is not possible and one has to take a compromise. In our case the best solution is that of having a false positive rate of 0 and a true positive rate of around 60%. When applied to soft magnets, this means that 60% of the soft magnets will be found without an error, meaning that 30% of the total compounds in the dataset will be correctly classified as soft (we consider a sample where the soft/hard ratio is 1). For those one will not need to perform any further calculation. Clearly such result becomes important if the dataset is very extended. Let us imagine, for the sake of the argument, that one wishes to compute the magneto-crystalline anisotropy for all the 236,000 Heusler alloys contained in the AFLOWLIB.org database. Our machine learning scheme will then save about 79,000 calculations.

4   Conclusion

In this chapter we have described how high-throughput electronic structure theory and machine learning strategies can accelerate the discovery and the design of new magnets. This indeed is an area of magnetism, which is still in its infancy, but that has already demonstrated a strong potential. At a first look, it appears that the efficiency of the approach is not high and the most skeptical reader may argue that

a more conventional trial and error approach can be more fruitful. This is true only in part. There are, in fact, a few key points to be noted.

Firstly, one has to expect that as the databases grow, the need for further calculations will be reduced. The mapping of the Heusler alloys thermodynamical stability has certainly been a challenging computational undertaking. However, we were able to conduct a first rapid screening of the intermetallic subset simply because the data for binary compounds were already in the AFLOWLIB.org database. It is particularly important to remark that many of these data were generated for research projects completely different from the one described here, meaning that data sharing is a unique asset of any high-throughput initiative.

The opposite argument is also valid, namely, our data have been useful not only to discover new magnets but also for screening materials according to other properties. For instance, the half Heusler library was investigated in the search of compounds presenting low thermal conductivity (Carrete et al. 2014). This concept of "recycling data" is indeed an important one, and certainly the success of the method needs to be evaluated over the entire range of discoveries that a dataset enables.

Finally, even data that do not contribute to a direct discovery (materials that are not thermodynamically stable) still can be used to construct machine learning algorithms to rapidly navigate the data themselves. Most importantly such machine learning models will help us in directing further calculations, so to improve the discovery throughput and identify gaps and opportunities in the materials space.

References

Archer T, Pemmaraju C, Sanvito S, Franchini C, He J, Filippetti A, Delugas P, Puggioni D, Fiorentini V, Tiwari R, Majumdar P (2011) Exchange interactions and magnetic phases of transition metal oxides: benchmarking advanced ab initio methods. Phys Rev B 84:115114

Błoński P, Hafner J (2009) Density-functional theory of the magnetic anisotropy of nanostructures: an assessment of different approximations. J Phys Condens Matter 21:426001

Calderon C, Plata J, Toher C, Oses C, Levy O, Fornari M, Natan A, Mehl M, Hart G, Nardelli M, Curtarolo S (2015) The AFLOW standard for high-throughput materials science calculations diagrams. Comput Mat Sci 108:233–238

Carrete J, Li W, Mingo N, Wang S, Curtarolo S (2014) Finding unprecedentedly low-thermal-conductivity half-heusler semiconductors via high-throughput materials modeling. Phys Rev X 4:011019

Castelliz L (1955) Beitrag zum Ferromagnetismus von Legierungen der Ubergangsmetalle mit Elementen der B-Gruppe. Z Metallk 46:198–203

Coey J (2009) Magnetism and magnetic materials. Oxford University Press, Oxford

Coey J, Sanvito S (2004) Magnetic semiconductors and half-metals. J Phys D Appl Phys 37: 988–993

Curtarolo S, Setyawan W, Hart G, Jahnatek M, Chepulskii R, Taylor R, Wang S, Xue J, Yang K, Levy O, Mehl M, Morgan D (2012a) AFLOW: an automatic framework for high-throughput materials discovery. Comput Mat Sci 58:218–226

Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor R, Nelson L, Hart G, Sanvito S, Nardelli M, Mingo N, Levy O (2012b) AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. Comput Mat Sci 58:227–235

Curtarolo S, Hart G, Nardelli M, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12:191–201

Dam HC, Nguyen VC, Pham TL, Nguyen AT, Kino H, Terakura K, Miyake T (2017) A regression-based feature selection study of the curie temperature of transition-metal rare-earth compounds: prediction and understanding. https://arxiv.org/abs/1705.00978

d'Avezac M, Luo JW, Chanier T, Zunger A (2012) Genetic-algorithm discovery of a direct-gap and optically allowed superstructure from indirect-gap Si and Ge semiconductors. Phys Rev Lett 108:027401

Dunn TM (1961) Spin-orbit coupling in the first and second transition series. Trans Farad Soc 57:1441

Faleev SV, Ferrante Y, Jeong J, Samant MG, Jones B, Parkin SS (2017) Origin of the tetragonal ground state of Heusler compounds. Phys Rev Appl 7:034022

Franchini C, Archer T, He J, Chen XQ, Filippetti A, Sanvito S (2011) Exceptionally strong magnetism in the $4d$ perovskites RTcO$_3$ (R = Ca, Sr, Ba). Phys Rev B 83:220402

Ghiringhelli L, Vybiral J, Levchenko S, Draxl C, Scheffler M (2015) Big data of materials science: critical role of the descriptor. Phys Rev Lett 114:105503

Ghiringhelli L, Carbogno C, Levchenko S, Mohamed F, Huhs G, Lueders M, Oliveira M, Scheffler M (2017) Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. NPJ Comput Mater 3:46

Graf T, Felser C, Parkin S (2011) Simple rules for the understanding of Heusler compounds. Prog Solid State Chem 39:1–50

Grazulis S, Chateigner D, Downs RT, Yokochi AT, Quiros M, Lutterotti L, Manakova E, Butkus J, Moeck P, Le Bail A (2009) Crystallography open database – an open-access collection of crystal structures. J Appl Crystallogr 42:726–729

Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Snchez-Carrera RS, Gold-Parker A, Vogt L, Brockway AM, Aspuru-Guzik A (2011) The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. J Phys Chem Lett 2(17):2241–2251

Hart G, Curtarolo S, Massalski T, Levy O (2013) Comprehensive search for new phases and compounds in binary alloy systems based on platinum-group metals, using a computational first-principles approach. Phys Rev X 3:041035

Hastie T, Tibshirani R, Friedman J (2013) The elements of statistical learning: data mining, inference, and prediction. Springer, New York

ICSD (2018) FIZ Karlsruhe and NIST, inorganic crystal structure database. https://www.fiz-karlsruhe.de/en/produkte-und-dienstleistungen/inorganic-crystal-structure-database-icsd

Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A (2017) Universal fragment descriptors for predicting properties of inorganic crystals. Nat Comm 8:15679

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater 1(1):011002

Janak J (1977) Uniform susceptibilities of metallic elements. Phys Rev B 16:255–262

Kanomata T, Shirakawa K, Kaneko T (1987) Effect of hydrostatic pressure on the Curie temperature of the Heusler alloys Ni$_2$MnZ (Z = Al, Ga, In, Sn and Sb). J Magn Magn Mater 65:76

Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, Aykol M, Rühl S, Wolverton C (2015) The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. npj Comput Mat 1:15010

Kresse G, Furthmuller J (1996) Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15–50

Kusne AG, Gao T, Mehta A, Ke L, Nguyen MC, Ho KM, Antropov V, Wang CZ, Kramer MJ, Long C, Takeuchi I (2014) On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. Sci Rep 4:6367

Lukas H, Fries S, Sundman B (2007) Computational thermodynamics, the Calphad method. Cambridge University Press, Cambridge

Magda G, Jin X, Hagymási I, Vancsó P, Osváth Z, Nemes-Incze P, Hwang C, Biró L, Tapasztó L (2014) Room-temperature magnetic order on zigzag edges of narrow graphene nanoribbons. Nature 514:608–611

Mazin I (1999) How to define and calculate the degree of spin polarization in ferromagnets. Phys Rev Lett 83:1427–1430

Moruzzi VL, Marcus PM (1989) Magnetism in FCC rhodium and palladium. Phys Rev B 39: 471–474

Oganov A, Glass C (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. J Chem Phys 124:244704

Oliynyk AO, Mar A (2018) Discovery of intermetallic compounds from traditional to machine-learning approaches. Acc Chem Res 51:59–68

Oswald A, Zeller R, Braspenning P, Dederichs P (1985) Interaction of magnetic impurities in Cu and Ag. J Phys F 15:193

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825

Perdew J, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Pickard CJ, Needs R (2011) Ab initio random structure searching. J Phys Condens Matter 23:053201

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) Aiida: automated interactive infrastructure and database for computational science. Comput Mat Sci 111:218–230

Rasmussen FA, Thygesen KS (2015) Computational 2D materials database: electronic structure of transition-metal dichalcogenides and oxides. J Phys Chem C 119(23):13169–13183

Requist R, Baruselli P, Smogunov A, Fabrizio M, Modesti S, Tosatti E (2016) Metallic, magnetic and molecular nanocontacts. Nat Nanotech 11:499–508

Rode K, Baadji N, Betto D, Lau YC, Kurt H, Venkatesan M, Stamenov P, Sanvito S, Coey J, Fonda E, Otero E, Choueikani F, Ohresser P, Porcher F, André G (2013) Site-specific order and magnetism in tetragonal Mn_3Ga thin films. Phys Rev B 87:184429

Rodriguez E, Poineau F, Llobet A, Kennedy B, Avdeev M, Thorogood G, Carter M, Seshadri R, Singh D, Cheetham A (2011) High temperature magnetic ordering in the $4d$ perovskite $SrTcO_3$. Phys Rev Lett 106:067201

Sandratskii L (1986) Energy band structure calculations for crystals with spiral magnetic structure. Phys Status Solidi B 136:167

Sanvito S, Oses C, Xue J, Tiwari A, Zic M, Archer T, Tozman P, Venkatesan M, Coey M, Curtarolo S (2017) Accelerated discovery of new magnets in the Heusler alloy family. Sci Adv 3:e1602241

Savrasov S (1998) Linear response calculations of spin fluctuations. Phys Rev Lett 81:2570–2573

Shalev-Shwartz S, Ben-David S (2014) Understanding machine learning. Cambridge University Press, Cambridge

Toher C, Plata J, Levy O, de Jong M, Asta M, Nardelli MB, Curtarolo S (2014) High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasi-harmonic Debye model. Phys Rev B 90:174107

Žic M (2017) Towards data-driven magnetic materials discovery. Ph.D Thesis, Trinity College Dublin, pp 1785–1812

Žic M, Archer T, Sanvito S (2017) Designing magnetism in Fe-based Heusler alloys: a machine learning approach. https://arxiv.org/pdf/1706.01840.pdf

Wadley P, Novák V, Campion R, Rinaldi C, Martí X, Reichlová H, Železný J, Gazquez J, Roldan M, Varela M, Khalyavin D, Langridge S, Kriegner D, Máca F, Mašek J, Bertacco R, Holý V, Rushforth A, Edmonds K, Gallagher B, Foxon C, Wunderlich J, Jungwirth T (2013) Tetragonal phase of epitaxial room-temperature antiferromagnet cumnas. Nat Commun 4:2322

Wohlfarth EP (1980) Ferromagnetic materials: a handbook on the properties of magnetically ordered substances. Elsevier, New York

Yan F, Zhang X, Yu Y, Yu L, Nagaraja A, Mason T, Zunger A (2015) Design and discovery of a novel half-Heusler transparent hole conductor made of all-metallic heavy elements. Nat Commun 6:7308

Yang K, Setyawan W, Wang S, Nardelli MB, Curtarolo S (2012) A search model for topological insulators with high-throughput robustness descriptors. Nat Mater 11:614–619

Yang K, Oses C, Curtarolo S (2016) Modeling off-stoichiometry materials with a high-throughput ab-initio approach. Chem Mater 28:6484–6492

Yong J, Jiang Y, Usanmaz D, Curtarolo S, Zhang X, Shin J, Li L, Pan X, Tachuchi I, Greene R (2014) Composition-spread growth and the robust topological surface state of Kondo insulator SmB_6 thin films. Appl Phys Lett 105:222403

Yu L, Zunger A (2012) Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. Phys Rev Lett 108:068701

Ziebeckt K, Webster P (1975) Helical magnetic order in Ni_2MnAl. J Phys F Met Phys 5:1756–1766

Multiferroic and Ferroelectric Rashba Semiconductors

18

Silvia Picozzi

Contents

Abstract

The coupling of different phenomena, such as ferroelectric or magnetic order, often mediated by spin-orbit interaction, offers a rich playground for emergent phenomena in multifunctional materials. In this chapter, we focus on multiferroics and ferroelectric semiconductors showing a strong Rashba effect, we explain the motivations behind their study, and we discuss a few prototypical highlights (from BiFeO$_3$ to spin-spiral TbMnO$_3$, from E-type manganites to GeTe). Our perspective is focused on the key role that first-principles calculations

S. Picozzi (✉)
Consiglio Nazionale delle Ricerche, CNR-SPIN, c/o Univ. Chieti-Pescara "G. D'Annunzio", Chieti, Italy
e-mail: silvia.picozzi@spin.cnr.it

played in this field in recent years, in terms of understanding microscopic
mechanisms, quantifying relevant quantities, and guiding the interpretation of
experimental results, in view of possible materials optimization for technological
purposes.

1   Introduction

One of the most common keywords in modern materials science is "multifunc-
tionality," i.e., a material's peculiarity to show at the same time more than one
active degree of freedom (charge, spin, dipole, etc. – possibly coupled) and,
correspondingly, its potential use when exploiting more than one "functionality."
Indeed, several breakthroughs in the last decades were achieved when merging
different subfields of materials science. This was, for example, the case in early 2000
with diluted magnetic semiconductors (Ohno 1999; Jungwirth et al. 2006), aimed
at bridging semiconducting behavior with magnetism, to combine logic and storage
functionalities. Unfortunately, their low operating temperature ($\leq$200 K) hindered a
real impact on technology. Within the race for multifunctionality, one of the grand
challenges in nowadays spintronics is represented by the electrical control of spin
degrees of freedom. Several approaches are being pursued, but natural candidates
appear to be those materials where – below a specified phase-transition critical
temperature – spontaneous dipolar electric and magnetic orders coexist (Schmid
1994), i.e., multiferroics (MF, see Fig. 1). Even more important and intriguing is
the study of compounds where the two relevant order parameters (i.e., ferroelectric
polarization and magnetization) not only coexist but are intimately connected via
the so-called "magnetoelectric" (ME) coupling: one can then induce/control/switch
ferroelectric properties via a magnetic field and, vice versa, magnetic properties
via an electric field. In the same context, an alternative class of emerging materials
is represented by (nonmagnetic) ferroelectrics with strong Rashba effects, the so-
called ferroelectric Rashba semiconductors (FERSC, see Fig. 1) (Picozzi 2014).
In those materials, the spin degrees of freedom become active, due to spin-orbit
coupling (SOC) and not to long-range magnetic order (as straightforward and
conventional in magnetic materials). As detailed below, in FERSC, SOC removes
the spin degeneracy of the band structure and creates an electrically controllable
k-dependent spin polarization in the reciprocal space.

The research on these two materials classes implied, implies, and will imply
progresses in various aspects, from the fundamental physics point of view to mate-
rials science issues to new technological paradigms that could revolutionize future
information technology. Both materials classes are relatively new: multiferroics
have experienced a hype since 2003 (Kimura et al. 2003; Wang et al. 2003), and
ferroelectric Rashba semiconductors were first proposed in 2013 (Di Sante et al.
2013). Materials modeling has therefore played a crucial role in recent years:
computational materials scientists heavily contributed to explain experimental data,
to identify new microscopic mechanisms, and to design new MF or FERSC. A
concerted effort from various modeling approaches was carried out in these fields,
ranging from phenomenological approaches (e.g., the Landau theory of second-
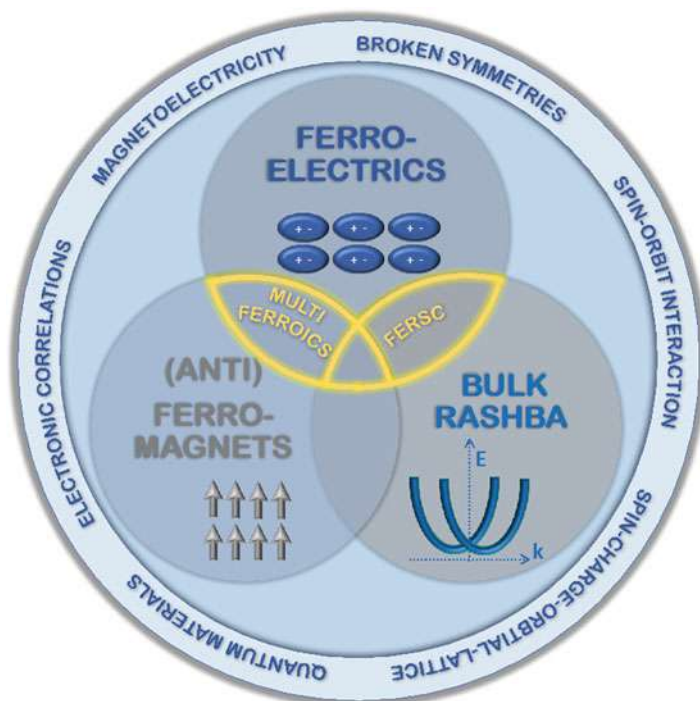
Fig. 1 Pictorial view of the different classes of multifunctional materials discussed in this chapter. Multiferroics (as the overlap between ferroelectrics and (anti-)ferromagnets) and ferroelectric Rashba semiconductors (as the overlap between ferroelectrics and materials showing a bulk Rashba effect). Relevant keywords are reported in the outer circle

order phase transitions (Harris 2007), with symmetry considerations playing a crucial role) to model-Hamiltonian approaches, ranging from many-body to – mostly classical – spin-based Hamiltonians, etc. In this chapter, however, we will mostly concentrate on the results obtained from materials modeling based on density functional theory (DFT). Moreover, since the research efforts on MF were massive on a worldwide and longer time scale, we will mainly focus on multiferroicity and magnetoelectricity, leaving a smaller part at the end of the chapter to the FERSC.

2   Basic Ingredients: Ferroelectrics, Symmetries, and Spin-Orbit Coupling

For the sake of clarity, several ingredients – most of them of relevance for both MF and FERSC – need to be preliminarily introduced, before moving to the main part of the chapter. We don't introduce (anti-)ferromagnets, since this chapter is part of a section fully devoted to magnetism, and we assume the reader to be already familiar with this topic.

Ferroelectrics (FE) (Rabe 2007; Scott 2000): materials where a spontaneous dipolar ordering occurs below a Curie temperature, leading to ferroelectric polarization P switchable via an electric field. Due to the presence of a polar axis, they intrinsically lack inversion symmetry. The phase transition occurs at a critical Curie temperature, T_C, above which the system is "paraelectric." Below T_C, even when removing the external electric field, the FE order is kept. As such, their nonvolatile behavior is at the heart of a worldwide multibillionaire market based on FE sensors, memories, transducers, etc.

Symmetries: in the framework of FE and/or (anti-)ferromagnetic phase transitions, symmetry considerations are undoubtedly helpful. In MF and FERSC, the main relevant symmetries to be considered are space-inversion symmetry (SIS) and time-reversal symmetry (TRS). An electric charge dipole (or the corresponding order parameter, i.e., polarization characterizing ferroelectrics) always breaks SIS, as it includes in its definition the vector position r; however, it preserves TRS. On the other hand, a spin moment (or the corresponding order parameter, i.e., magnetization characterizing ferromagnets) always breaks TRS but keeps SIS (as evident from the analogy with a current loop). The materials modeling of both MF and FERSC greatly benefits from symmetry considerations, in terms of space groups, point groups, time-reversal, mode decompositions, etc. Symmetry represents a precious tool to highlight the mechanism behind ferroelectricity and its coupling to other electronic degrees of freedom (mainly spin-related, but also linked to charge-ordering, orbital-ordering, etc.). Analogously, symmetry dictates allowed terms in k•p Hamiltonian (Yu 2005; Harrison 1989) around relevant high symmetry points in the Brillouin zone. To this end, crystallography software, available, i.e., at the Bilbao Crystallographic Server (Orobengoa et al. 2009) or via the Isotropy Software Suite (Stokes n.d.), is particularly helpful for materials modeling.

Spin-orbit coupling (SOC): relativistic interaction linking the electron spin with its orbital motion in space. For long time neglected due to its generally tiny magnitude for most of the technologically relevant materials, it is nowadays regarded as rich source of exotic physics, manifesting in topological insulators (Hasan and Kane 2010), weak ferromagnetism, spin-induced electrical polarization, etc. Thanks to the "locking" between spin and momentum k, SOC has recently become "functional" (Manchon et al. 2015), i.e., exploited to achieve new functionalities. SOC can be thought of as the source of Rashba effect, as explained in detail in Sect. 4.

3   Bulk Multiferroics

The recent research activity on MF has spanned many aspects, from bulk to heterostructures, from studies on magnetic/ferroelectric domains to domain wall engineering. Moreover, it has focused on many different materials (from oxides to chalcogenides, from inorganic to organic and hybrid metal-organic frameworks, etc.) with different microscopic mechanisms at play. For the sake of brevity, in this chapter we will only discuss three highlights on prototypical MF in the family of transition metal oxides (TMO) with perovskite crystal structure: $BiFeO_3$, $TbMnO_3$, and $(Ho/Y)MnO_3$. For each of them, the main experimental results will be briefly

recalled, along with some of the most representative results obtained via first-principles simulations. Furthermore, we will address the microscopic mechanisms and the related switching behavior. We however caution the reader that a huge literature on different aspects of MF exists, and we refer the reader to several excellent reviews for a deeper understanding (Cheong and Mostovoy 2007; Kimura 2007; Ramesh and Spaldin 2007; Picozzi and Ederer 2009; Dong et al. 2015; Catalan and Scott 2009).

The hype on MF started from an apparent problem: most of the prototypical ferroelectrics are perovskite TMO (such as $BaTiO_3$) where the polar distortion mainly arises from the B-site cation showing an empty d-shell (i.e., Ti is formally d^0). The latter electronic configuration can lead to some degree of covalency when bonding with surrounding oxygens during the FE distortion; this situation tends to off-center the B-site cation toward one (or three) of the octahedral oxygens at the expense of weakening the bonds with other nearby oxygens. Being an empty d-shell clearly incompatible with any magnetism (requiring a partially filled d-shell), there arose the famous question posed by Nicola A. Spaldin, one of the leading figures in the field: "*Why there are so few magnetic ferroelectrics?*" (Hill 2000). People tried therefore to search for alternative routes to reconcile magnetism and ferroelectricity, although the initial "exclusion rule" was later revisited (Bhattacharjee et al. 2009) (i.e., for $CaMnO_3$) (Barone et al. 2011). The three prototypes below describe successful ways of overcoming the obstacle: in $BiFeO_3$, the B-site cation (Fe) gives rise to magnetism and the A-site (Bi) gives rise to (displacive) ferroelectricity, whereas in $TbMnO_3$ and $HoMnO_3$, it is the spin ordering that gives rise to ferroelectricity. Within this context, an important concept established in MF during the last 15 years was the emergence of "improper" ferroelectricity (Cheong and Mostovoy 2007): a primary order parameter of "electronic" origin (such as spin, charge, or orbital order) is responsible for symmetry lowering by breaking SIS. As such, it paves the way to ferroelectric polarization, appearing therefore as a secondary order parameter obtained as a by-product of the electronic transition. The concept of electronic ferroelectricity (Picozzi and Ederer 2009) had a large impact, primarily for two technologically appealing features: (i) since polarization involves electronic degrees of freedom, its FE switching is expected to be much faster than that in conventional polarization switching of displacive ferroelectrics, where ionic degrees of freedom are involved; (ii) especially in the case of spin-driven ferroelectricity (such as $TbMnO_3$ and $HoMnO_3$), the ME coupling is expected to be exceedingly strong, since it is the frustrated magnetic order that induces the polar behavior, magnetism, and ferroelectricity, therefore sharing the same microscopic origin.

3.1   Prototype for Lone-Pair-Driven Polarization: BiFeO₃

$BiFeO_3$ (Wang et al. 2003) is often considered the "holy grail" of multiferroics and has been thoroughly studied due to the persistence of its multiferroic behavior up to large temperatures. In particular, the so-called "G-type" antiferromagnetic ordering (i.e., AFM coupling of an iron cation with all its nearest neighbors or,
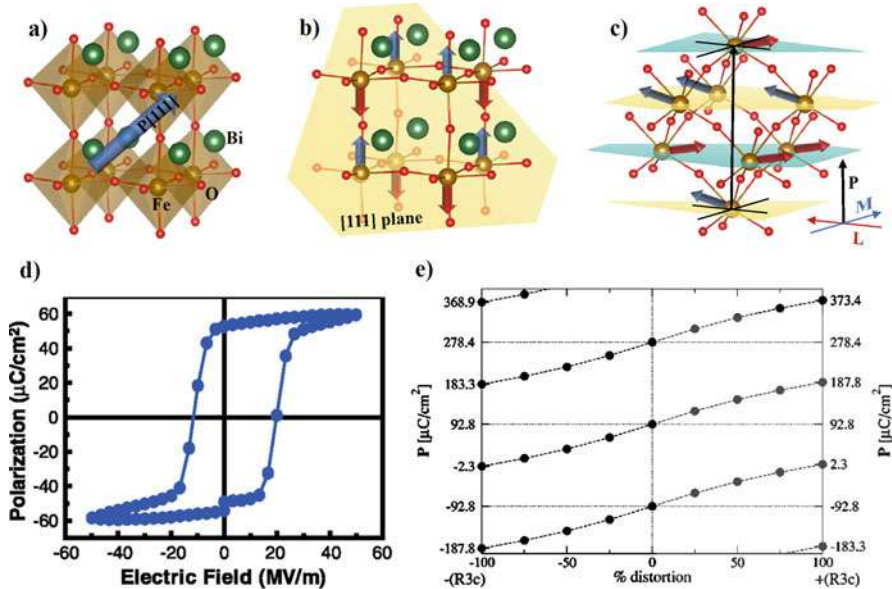
Fig. 2 Multiferroicity in BiFeO$_3$: (**a**) Rhombohedral structure (*R3c* space group), with FE
polarization pointing along the [111] pseudo-cubic axis. FeO$_6$ octahedral cages are highlighted
in brown, whereas Bi, Fe, and O ions are shown with green, brown, and red spheres, respectively.
(**b**) AFM-G-type magnetic ordering, seen as a sequence of FM planes (shown as yellow plane)
antiferromagnetically coupled along the [111] axis. (**c**) Spin canting of the (collinear) AFM-G
magnetic moments on Fe cations (represented by blue and red arrows) in the planes perpendicular
to the polarization direction (see yellow and light blue planes) and giving rise to a small net in-plane
magnetization. The three perpendicular vectors (weak magnetization M, antiferromagnetic order
parameter L, and ferroelectric polarization P) are shown in the lower right part. (**d**) P-E hysteresis
curve for the BiFeO$_3$ FE polarization measured (at room temperature) along the pseudo-cubic
[001] axis (i.e., the observed values are therefore a projection of the polarization vector which is
aligned along the [111] axis). (Reprinted from Catalan and Scott 2009). (**e**) DFT-calculated values
(within LSDA+U method, using $U = 2$ eV) of FE polarization along the FE switching path, going
from the –P state (R3c structure, left side of x-axis) to the centrosymmetric cubic structure to the
+P state (R3c structure, right side of x-axis). We recall that that the spontaneous polarization is
defined as half the difference between two opposite polarization states and that only polarization
differences are physically observable: the calculated values of polarization are therefore shown in
terms of several polarization "branches," differing by the polarization quantum. (Reprinted from
Neaton et al. 2005)

alternatively, seen as a stacking of FM planes antiferromagnetically coupled along
the [111] axis) shows a Néel ordering temperature of 673 K (cfr Fig. 2b). Moreover,
the FE behavior, consistent with a strong rhombohedral distortion of the pseudo-
cubic unit cell with polarization pointing along the [111] direction (cfr Fig. 2a), has
a Curie temperature of 1103 K. Ferroelectricity is driven by the presence of highly
polarizable Bi *6s* "lone-pair" electrons (Seshadri and Hill 2001), which hybridize
with surrounding oxygens in the Fe octahedral cages, therefore leading to strong
displacements of Bi cations along the [111] pseudo-cubic direction and to a related

huge FE polarization of about 90 μC/cm^2 (Wang et al. 2003) (cfr Fig. 2d). The antiferromagnetism is, on the other hand, originated by Fe ions; due to the presence of octahedral counterrotations of the oxygen cage, a non-negligible Dzyaloshinskii-Moriya (DM) interaction is also observed, which leads to a small canting of the magnetic moments in thin films and, via a generalized DM interaction, to a spin-cycloid in bulk samples (Lebeugle et al. 2008; Johnson et al. 2013).

DFT contributed exhaustively from the very beginning to the BiFeO$_3$ understanding, starting from the evaluation (Neaton et al. 2005) within the Berry phase theory of the ferroelectric polarization (King-Smith and Vanderbilt 1993; Resta 1994) (cfr Fig. 2e), directed along the pseudo-cubic [111] axis and of which experiments observe the out-of-plane projection (see Fig. 2d). Another important DFT contribution was the small deviation of the spin moments on Fe ions with respect to the collinear G-type AFM ordering (Ederer and Spaldin 2005), due to spin-orbit coupling and DM interaction (see Fig. 2c): the magnetic easy axis is found to be perpendicular to polarization, and the canting of the spin moments, induced by the (nonpolar) octahedral counterrotation of the oxygen cages, can tilt the iron spin moments. As a result, the net magnetization M, the AFM order parameter L, and the FE polarization P form a Cartesian set. Since the polar distortion is driven by the Bi off-centering and the weak FM is driven by the anti-ferrodistortive mode of oxygens, it follows that P and M are rather decoupled, so that the small net magnetization cannot be controlled via an electric field. Other relevant issues in BiFeO$_3$, both from the modeling and experimental point of view, are (i) the observation of a super-tetragonal phase obtained in BiFeO$_3$ under significant strain (Zeches 2009); (ii) the study of domain walls (Farokhipoor and Noheda 2011), particularly complex due to the rhombohedral setting which gives rise to eight possible directions of domain walls, but with the appealing perspective of a larger conductivity of domain walls with respect to domains; and (iii) the exchange bias observed at the interface with standard ferromagnets, leading, for example, to the electric control of ferromagnetism at the BiFeO$_3$/CoFe interface (Bea et al. 2008; Heron and Trassin 2011).

3.2   Prototype for Spin-Spiral-Driven Polarization: TbMnO₃

Orthorhombic TbMnO$_3$ is undoubtedly one of the multiferroics with the richest physics phenomenology, the efforts devoted to its understanding leading beyond question to significant fundamental progresses. From the experimental point of view (Kimura et al. 2003), the Mn spins order first at 42 K in a collinear structure with spins along the b-direction to form a longitudinal spin density wave. Moreover, at 28 K, they lock into a non-collinear incommensurate spin arrangement forming a spiral running in the bc plane (the spiral modulation vector depends on the temperature, but it is around $\mathbf{q} = (0,0.28,0)$. At lower temperatures (around 7 K), even the Tb spins order. In a landmark publication (Kimura et al. 2003), Kimura et al. reported a ferroelectric polarization along the c-axis (reaching
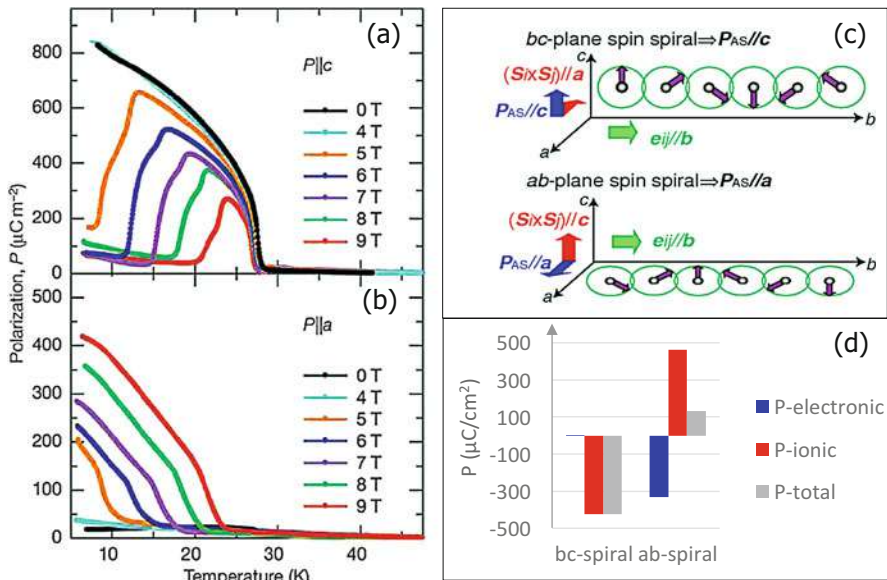
Fig. 3 Multiferroicity in orthorhombic TbMnO$_3$ with spiral-spin order. (**a**, **b**) Measured FE polarization along the c- and a-axis (in the *Pbnm* setting), respectively, as a function of temperature and of the magnetic field applied along the b-axis. (Reprinted from Kimura et al. 2003). (**c**) Relationship between the direction of polarization and the spin-spiral plane according to the spin-current model. The magnetic ordering in the upper part shows the ground state and gives polarization P_c (explaining experimental results shown in panel **a**), whereas the magnetic ordering in the lower part shows the spiral reorientation obtained under magnetic field along the b-axis and gives polarization P_a (explaining the results shown in panel **b**). (Reprinted from Mochizuki 2011). (**d**) Results from DFT (values from Malashevic and Vanderbilt 2008; Xiang et al. 2008): polarization (in μC/cm^2) calculated for TbMnO$_3$ (in two different spin-spiral configurations, confined in the *bc* plane and a*b* plane). The total polarization (P-total) is broken down into a purely electronic contribution (P-electronic) for ions fixed in centrosymmetric positions and "ionic" contribution obtained for "relaxed" atomic positions as P-ionic = P-total-P-electronic

a value of 0.08 μC/cm^2) emerging below the spin-spiral ordering temperature, therefore highlighting the intimate coupling between magnetic and dipolar orders (cfr Fig. 3a). Even more intriguingly, it was shown that the electric polarization can be controlled by a magnetic field: upon applying a magnetic field along the b-axis, the spiral flops into the *ab* plane, and, correspondingly, the polarization direction changes into the a-axis (cfr Fig. 3b).

The microscopic origin of polarization will be discussed in Sect. 3.4, but we anticipate here that the spin-spiral breaks SIS and the consequent magnetically driven polarization is explained in terms of the spin-current model (Katsura et al. 2007) (cfr Fig. 3c). From the first-principles point of view, two papers (Xiang et al. 2008; Malashevic and Vanderbilt 2008) (appeared one after the other in Phys. Rev. Lett. 2008) gave important contributions, reaching similar conclusions.

In Malashevic and Vanderbilt (2008) and Xiang et al. (2008), the spiral was modeled within DFT+U by imposing a commensurate magnetic order with rotating spins using a unit cell three times larger (along the *b*-axis) than the *Pbnm* chemical unit cell. In Fig. 3d), we show the results for FE polarization, broken down into different contributions, coming from (1) a "purely electronic" term, i.e., given by the breaking of SIS imposed by the spin spiral, with ions fixed in their centrosymmetric "paramagnetic" positions, and (2) an "ionic" term, arising from exchange-strictive mechanisms, i.e., polar displacements of ions following the spin-driven electronic polarization. While the careful discussion of the microscopic sources of polarization will be deferred to Sect. 3.4, we remark that (i) the sign and the order of magnitude of polarization estimated from DFT are in excellent agreement with experiments, a remarkable achievement for such a complex computational task; (ii) the contribution of ionic vs electronic terms can significantly vary from *bc* to *ab* spirals, being also strongly affected by octahedral distortions (i.e., the purely electronic term is sizeable in the *ab*-plane spiral, but negligible for the *bc* spiral). Moreover, the sign is sensitive to fine magnetic and electronic details: for the *ab* spiral, the ionic and electronic contributions show an opposite sign with comparable magnitude. On the other hand, in the *bc* spiral, the ionic contribution is much greater in magnitude than the electronic term.

3.3   Prototype for Heisenberg-Exchange-Driven Polarization: $RMnO_3$ (R = Ho, Y)

One of the remarkable examples in MF in which materials modeling significantly contributed and helped experiments (Lorentz et al. 2007) was represented by some orthorhombic manganites (*Pbnm* space group) with small rare-earth ions (such as $HoMnO_3$, $YMnO_3$, etc.); these systems show a collinear AFM-E-type order (i.e., in-plane FM zigzag chains antiferromagnetically coupled to nearby chains and with an AFM out-of-plane coupling, or equivalently ↑↑↓↓ Mn chains) below ≈26 K. Sergienko et al. (2006) proposed for the first time a mechanism based on symmetric-exchange striction (or equivalently, spin-lattice coupling) to occur in E-type manganites, where a remarkable role is played by the Mn-O-Mn angle. In closer detail, based on the double-exchange model, hopping was forbidden between opposite $t2_g$ spins but allowed to depend on the Mn-O-Mn angle in the case of parallel spins. As such, the competing contributions given by elastic energy and hopping led to different Mn-O-Mn angles for parallel and antiparallel spins (see Fig. 4a) and, eventually, to a coherent displacement of the O anions with respect to the Mn sublattice, resulting in a net polarization along the a-axis (see results from Monte Carlo simulations shown in Fig. 4a). Although from Sergienko's model one can infer relevant information on the direction of the polarization and its manifest origin in the E-type magnetic configuration, it is clear that one cannot deduce any reliable estimate of polarization. To this end, soon after Sergienko's publication, Picozzi et al. (2007) performed DFT simulations, which also provided a deeper analysis of microscopic effects, in addition to the somewhat simplified
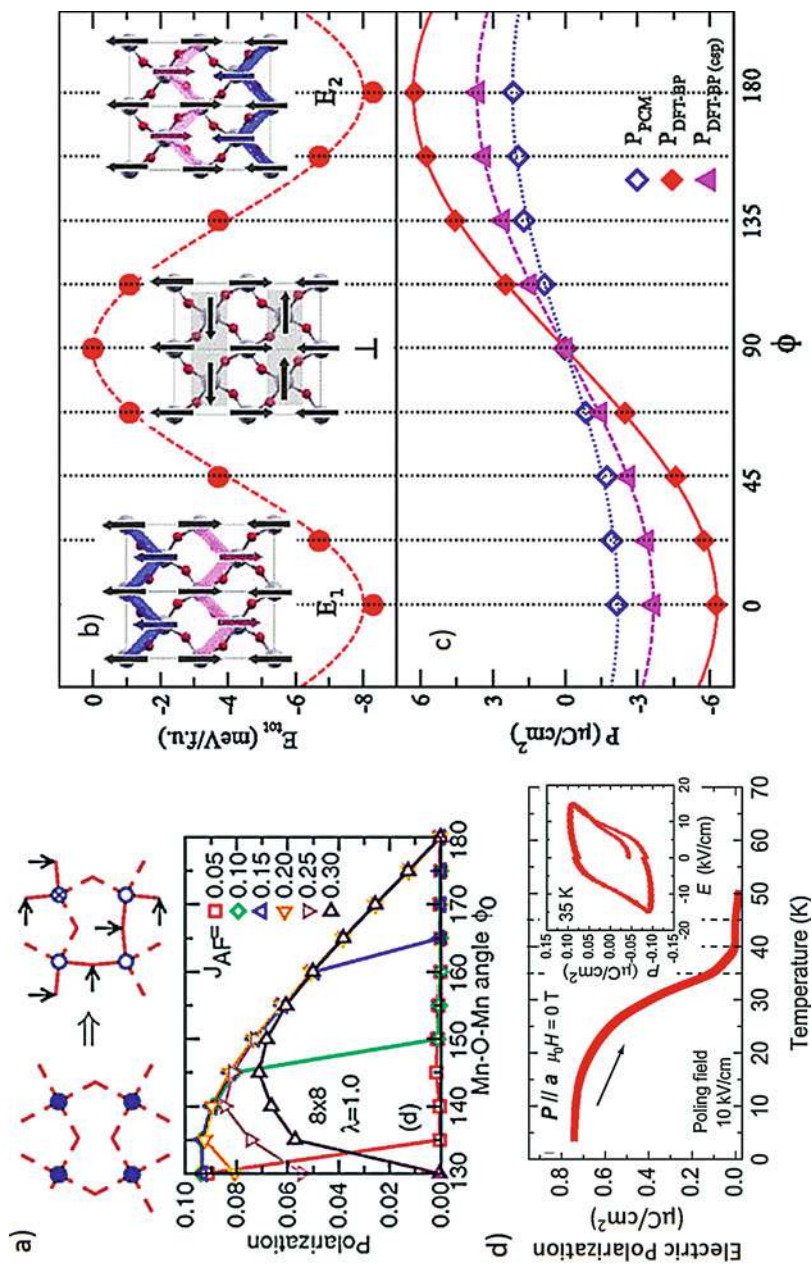
Fig. 4 (continued)

picture provided by double-exchange models. In particular, the "dual nature" of ferroelectric polarization was discussed in Picozzi et al. (2007) (also shown in Fig. 4c): the role of purely electronic contribution to P (i.e., with ionic degrees of freedom frozen in the centrosymmetric structure) compared to magnetostrictive contribution (i.e., allowing ions to move from the centrosymmetric configuration, when subject to the AFM-E spin ordering) was highlighted and found to be of more or less equal size. Moreover, polarization was found to be much larger than in noncollinear MF, as will be more clearly explained in Sect. 3.4. The confirmation from experiments took some time, primarily due to synthesis issues: the ground state for manganites with small rare-earth cations is the hexagonal crystal structure, $P6_3cm$ space group, and the stabilization of orthorhombic phase with *Pnma* space group can only be achieved under pressure, often leading to polycrystalline samples. We report in Fig. 4d the measured ferroelectric polarization for AFM-E $YMnO_3$ (Nakamura et al. 2011), along with a P-E hysteresis loop undoubtedly showing the symmetric-exchange-striction mechanism at play. As for the magnitude of polarization – expected to be much larger than any spin-orbit-driven polarization but still measured to be lower than what is theoretically predicted – the situation is still unclear and requires further studies.

3.4   Microscopic Mechanisms for Polarization

In this subsection we summarize the microscopic mechanisms that were found as most active in multiferroics, along with their corresponding materials modeling. Let us first recall that, as for the sources of ferroelectricity, one has to distinguish two cases (Cheong and Mostovoy 2007): (i) "proper" ferroelectricity, where polarization is the primary order parameter (Bousquet et al. 2008) of the FE transition, and (ii) "improper" ferroelectricity, where – as already mentioned above – polarization is a *secondary* order parameter, whose behavior is controlled by a symmetry-lowering

◄─────────────────────────────────────

Fig. 4 Multiferroicity in orthorhombic $RMnO_3$ (R=Ho, Y) with E-type AFM spin order (shown in the insets of panel **b**) with pink and blue stripes showing the zigzag chains. (**a**) Left: Mn-O arrangement with Mn-disordered spins (filled blue circles). Right: oxygen displacements (represented by black arrows) occurring with zigzag chains of Mn spins (open and crossed blue circles denote up and down spin, respectively) in the E-AFM spin ordering. Lower part: results (obtained from Monte Carlo method) for FE polarization as a function of the Mn-O-Mn angle for different values of the AFM superexchange constant J_{AF}. (Reprinted from Sergienko et al. 2006). (**b**) Results from DFT of total energy along the ferroelectric switching path. Here the switching path is obtained by changing the orientation of the central spins (cfr gray regions in the central configuration) with respect to the Mn spin in the origin of the unit cell. (**c**) Polarization calculated via the point charge model (empty blue diamonds), via Berry phases within DFT (filled red diamonds), and via Berry phases within DFT for centrosymmetric structure (pink triangles). Panels (**b**, **c**) are reprinted from Picozzi et al. 2007). (**d**) Temperature dependence of the FE polarization along the *a*-axis in orthorhombic YMnO3 in the AFM-E spin configuration. The inset shows the P-E hysteresis curve measured at 35 K. (Reprinted from Nakamura et al. 2011)

primary electronic transition (very often, though not always, of "electronic" origin). $BiFeO_3$ clearly belongs to the first class, its transition being induced by the Bi $6s$ "lone pair." As shown in Fig. 4a), Bi is surrounded by 12 O atoms, 6 of which approximatively lie in the same plane of Bi (with a marginal role for the FE transition), whereas the other 6 oxygens form a prismatic cage. Under FE distortion, the main FE mode involve Bi displacing toward three of them, forming therefore three short and three long bonds, consistently with the lone-pair distribution lying opposite to the short bonds. Since FE requires bistability, it is important to focus also on the configuration showing opposite FE polarization. In the $BiFeO_3$ case, this involves an opposite shift of Bi toward the other three O in the prismatic cage, as shown in Fig. 5a').

Among MF belonging to the second class, we mention that in both $TbMnO_3$ and $(Ho/Y)MnO_3$, ferroelectricity is spin-induced. At this point, we incidentally mention that another classification exists: Khomskii (2009) proposed to classify as "type II" all those MF where ferroelectricity has a magnetic origin and as "type I" all the others. We remark, however, that no matter which classification one follows, the origin of ferroelectricity constitutes the main ingredient to characterize a specific MF. Within spin-induced ferroelectricity (Dong et al. 2015; Tokura et al. 2014), one often distinguishes P as being originated by two kinds of exchange interactions: (A) symmetric (Heisenberg-like, i.e., of the form $P \propto S_i \cdot S_j$, pictorially represented in Fig. 5c, d) and (B) antisymmetric (Dzyaloshinskii-Moriya-like, i.e., of the form $P \propto e_{ij} \times (S_i \times S_j)$, pictorially represented in Fig. 5b). Since the Heisenberg exchange interaction in *3d*-TMO is usually much stronger than the DM exchange (based on relativistic spin-orbit coupling), the first class of MF usually shows a much larger polarization than the second class (typically, the order of magnitude of *P* are 0.1–1 $\mu C/cm^2$ and 0.01–0.1 $\mu C/cm^2$, for class (A) and (B), respectively). Common to both classes is the observation that polarization is the sum of a purely "electronic" term (i.e., not involving ionic relaxations with respect to ideal centrosymmetric positions) and of an "exchange-striction" term, given by the atomic relaxations according to exchange-energy minimization. It is helpful to discuss first the electronic mechanism and then the exchange-strictive one. In case (B), following Katsura, Nagaosa, and Balatsky (KNB) (Katsura et al. 2007), SOC may induce an electronic charge dipole for a pair of non-collinear spins. The KNB mechanism is also often addressed as driven by "spin current," as the cross product $S_i \times S_j$ is proportional to the spin current j_s, where the DM vector acts as its vector potential. The same mechanism is also often described as "inverse DM": a spin spiral generates a uniform bias of DM, therefore producing a uniform displacement of oxygens, in turn giving rise to P (see Fig. 5b, lower part). The latter mechanism represents therefore the "exchange-strictive" term and was first proposed by Sergienko and Dagotto (2006). On the other hand, in case (A), one deals with Heisenberg-driven polarization, where the latter can even occur for collinear spins. Its most common realization is the up-up-down-down spin chains, observed in AFM-E $HoMnO_3$, Ca_3MnCoO_6 (with spin chains running along the c-axis) (Wu et al. 2009), or even in rare-earth orthoferrites (Tokunaga et al. 2008) (where the two kinds of spins involved in the symmetry-lowering exchange interaction are
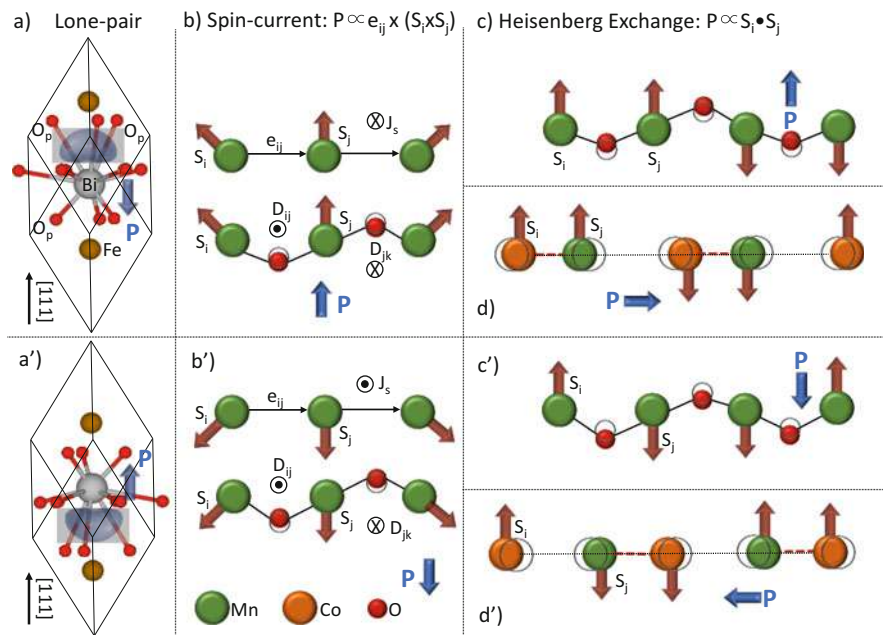
Fig. 5 Main mechanisms for multiferroicity. (**a**) Lone-pair driven: displacive ferroelectricity in BiFeO$_3$. Bi displaces toward the three lower oxygens, and correspondingly the lone-pair shifts toward the three upper oxygens. (**a'**) Opposite polarization state with respect to (**a**): Bi displaces toward the three upper oxygens and correspondingly the lone-pair shifts toward the three lower oxygens. (**b**) Spin-current-driven ferroelectricity (e.g., in spin-spiral TbMnO$_3$): polarization according to the KNB model (i.e., electronic term, upper part of the panel) and according to the Sergienko-Dagotto's model (i.e., ionic contribution guided by exchange striction, lower part of the panel). O atoms in paramagnetic (centrosymmetric) configuration are represented with empty black circles, consistent with (magnetic-independent) octahedral tilting. Upon spin-spiral ordering, the O atoms displace downward (see red spheres). (**b'**) Opposite polarization state with respect to (**b**): the spin-spiral rotation is reversed, leading to an opposite spin current (see upper part of the panel). As such, even the O displacements are reversed (O ions displace upward – see difference between black circles and red spheres), leading to an opposite exchange-strictive contribution (lower part of the panel). (**c, d**) Heisenberg-exchange-driven ferroelectricity in (**c**) AFM-E (Ho/Y)MnO$_3$ and (**d**) Ca$_3$CoMnO$_6$ – lower part. In ortho-manganites, with respect to the undistorted structure with centrosymmetric positions (empty black circles), O ions displace to reduce (increase) the Mn-O-Mn angle if the Mn spins are antiparallel (parallel) – see red spheres. In Ca$_3$CoMnO$_6$, with respect to the undistorted structure with centrosymmetric positions (empty black circles), Co and Mn get closer (farther) for parallel (antiparallel) spins. (**c', d'**) Opposite polarization state with half of spins switched. In this case, the exchange-striction mechanism is reversed (due to opposite couple of parallel and antiparallel spins). In all panels, the light blue arrows represent the axis and orientation of FE polarization

3d of iron and *4f* of the rare-earth ion). The purely electronic term contributing to polarization can be understood as follows: (a) in the case of ortho-manganites, by considering that the charge cloud around the O linking two Mn is different depending on whether the two spins are parallel or antiparallel, and (b) in the case

of Co-Mn, by considering that the charge cloud in between Co and Mn is different, depending on whether their spins are parallel or antiparallel. As for the exchange-strictive term in ortho-manganites, as briefly mentioned, the Mn-O-Mn angle is very sensitive to the spins on the two Mn ions: it tends to be closer to (smaller than) 180° for parallel (antiparallel) spins. Similarly, two different spins (say, Co and Mn as in Ca_3MnCoO_6) tend to be closer (farther) in the case of parallel (antiparallel) spins. In both situations (A) and (B), both the exchange-strictive and the purely electronic terms are relevant, and DFT reveals as a precious tool to quantify them (as previously discussed and shown in Figs. 3 and 4).

We remark that, due to lack of space, only few mechanisms were here selected and described in detail. However, several more mechanisms were proposed, and we mention those that were carefully addressed from the first-principles point of view:

(i) the so-called hybrid improper ferroelectricity (HIF) (Perez-Mato et al. 2004), involving a trilinear coupling between two unstable nonpolar modes coupled with a polar mode (which might itself be stable or "hard" when taken alone, but that is softened or "activated," thanks to the coupling with the other two modes). HIF was successfully proposed also for multiferroics and/or magnetoelectrics, for example, in the layered perovskite $Ca_3Mn_2O_7$ (Benedek and Fennie 2011) or in metal-organic frameworks (Stroppa et al. 2013);

(ii) polarization driven by charge-ordering or charge-spin ordering, i.e., polarization arising from a SIS-breaking charge-ordering transition or from a combined spin-charge transition (see, e.g., in magnetite (Alexe 2009; Yamauchi et al. 2009), half-doped manganites (Giovannetti et al. 2009a), or nickelates (Giovannetti et al. 2009b));

(iii) spin-dependent p-d hybridization, requiring relativistic SOC, as discussed in Jia et al. (2007), Xiang et al. (2013b), and Mochizuki (2011).

4   Bulk Ferroelectric Rashba Semiconductors

In order to understand the concept behind FERSC, it is necessary to first discuss where the Rashba effect originates from (Rashba 1960; Bihlmayer 2015), by considering the conduction band minimum (or equivalently, valence band maximum) of a conventional semiconductor. In a rough "free electron approximation" and in the presence of both SIS and TRS, the electronic bands are spin doubly degenerated for any $\mathbf{k}$-point in the Brillouin zone (cfr Fig. 6a). On the other hand, when considering the case of a two-dimensional electron gas (2DEG), for example, at the surface of a semiconductor, one can assume an electric field $\mathbf{E_0}$ perpendicular to the 2DEG plane (cfr Fig. 6b), so that SIS is broken. According to relativistic Lorenz transformations, an electron traveling with momentum $\mathbf{p}$ in the 2DEG plane will therefore experience an effective magnetic field $\mathbf{B_{eff}} \sim \mathbf{E_0} \times \mathbf{p}/mc^2$ in its rest frame, which the electron spin couples to via a Zeeman term (i.e., spin-orbit coupling). Carriers will therefore be subject to the so-called Rashba effect (RE), i.e., a $\mathbf{k}$-dependent spin splitting of
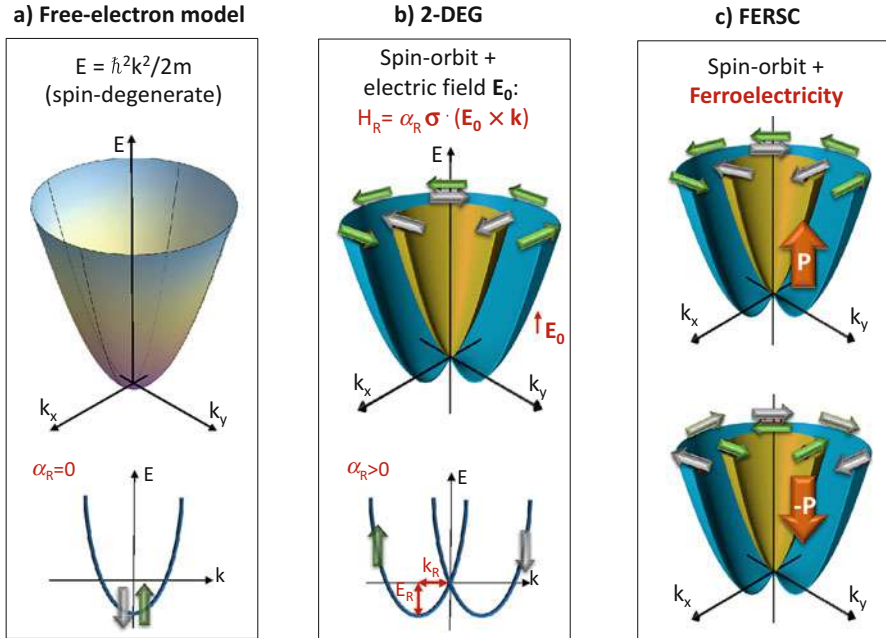
Fig. 6 From the free-electron model to the Rashba effect in 2DEG to FERSC. (**a**) Free-electron model, leading to parabolic spin-degenerate bands (both SIS and TRS are present). Upper and lower parts show a 3D view and 2D cut. (**b**) 2DEG model. In the presence of a surface or an interface, SOC + electric field (i.e., broken SIS) lead to the Rashba effect (i.e., a linear-in-k term is used in the Rashba Hamiltonian H_R with a coefficient α_R). Upper and lower parts show a 3D and 2D view. In the latter, the two conventional parameters to quantify the Rashba effect are defined: the difference in reciprocal space between the **k**-point where the minimum of the energy is located with respect to the high-symmetry **k**-point is labeled "k_R," whereas the difference in energies between the minimum of the conduction band and the energy at which the two spin-split bands cross – on the "diabolic" point – is labeled "E_R." The Rashba coefficient is defined as $\alpha_R = 2E_R/k_R$. (**c**) Rashba effect in FERSC (SIS broken): the spin texture is switched when polarization (see orange arrow) is switched. In all panels, green and gray arrows represent spins (i.e., spin-expectation value on the single eigenstate for a certain band index n and wave-vector **k**)

the electronic band structure [even of nonmagnetic compounds, see Fig. 6b). The Rashba Hamiltonian is usually written in terms of a linear-in-momentum term H_R (cfr Fig. 6b), leading to a spin polarization perpendicular to both $\mathbf{E_0}$ and momentum. Going beyond the usual picture of RE occurring at surfaces (or interfaces),there is conceptually nothing preventing the RE to occur in "bulk" materials. In particular, in ferroelectrics, the permanent and switchable polarization along a polar axis renders the compound intrinsically anisotropic, with an "effective Weiss-like" electric field able to generate momentum-dependent Rashba spin splitting in the plane (in reciprocal space) perpendicular to P. In this case, the spin-expectation value for any band with band index n and eigenvalue $\varepsilon_n(\mathbf{k})$ lies predominantly in the plane, i.e., perpendicular to the momentum and tangent to the energy isosurface $\varepsilon = \varepsilon_n(\mathbf{k})$ (see
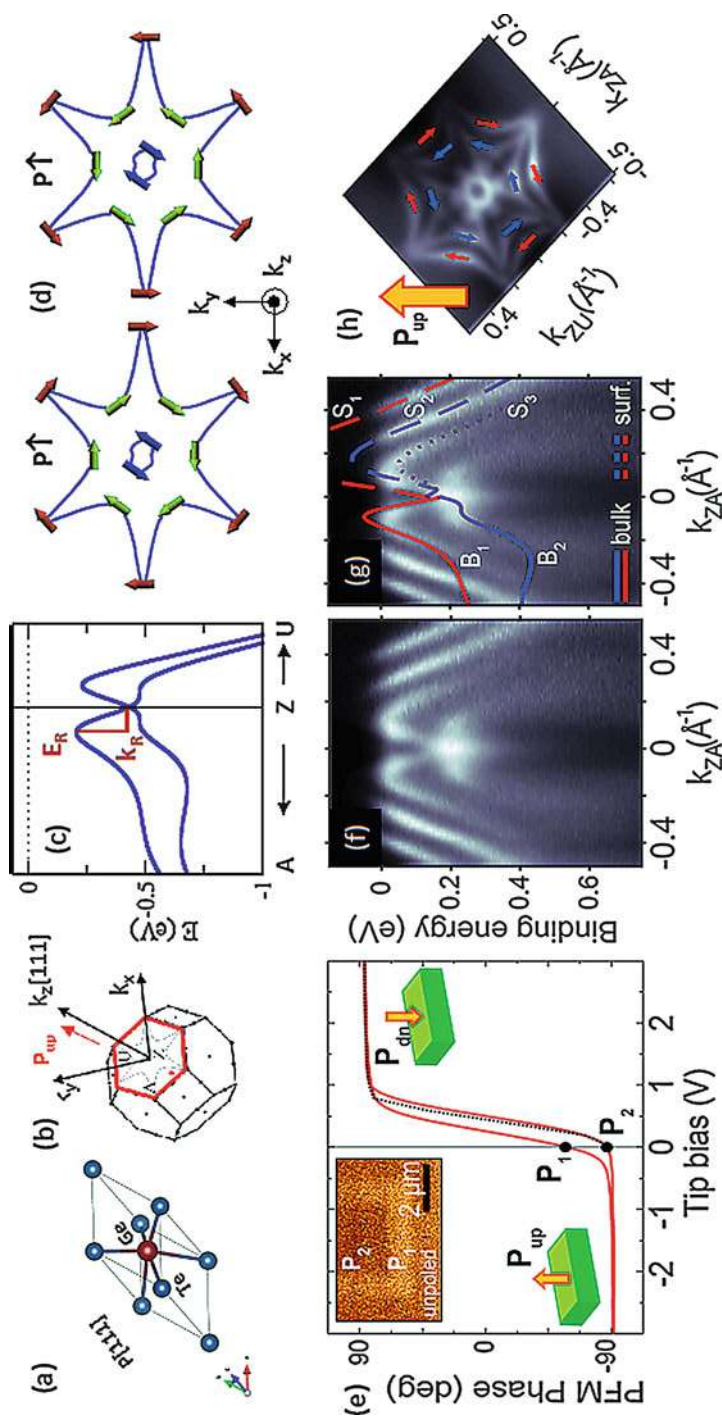
Fig. 7 (continued)

Fig. 6c). Moreover, an additional functionality brings up the main FERSC feature (Di Sante et al. 2013; Picozzi 2014): as shown in Fig. 6c (upper and lower parts), when switching the polarization direction, the Rashba field also becomes opposite, so that the spin texture is reversed. One therefore achieves a full control of the spin texture with an electric field in a nonvolatile way, i.e., due to bistability, each FE state is permanently kept even when the poling electric field used for switching is turned off.

4.1   The Prototypical FERSC: GeTe

The first example of RE in a "bulk" material was BiTeI, a polar layered material where ab initio predictions and angle-resolved photoemission spectroscopy (ARPES) reported a planar spin texture with huge RE (Ishizaka 2011; Crepaldi 2012). However, BiTeI – despite showing a polar axis – is not FE, i.e., the dipole moment cannot be switched. In 2013, Picozzi's group therefore proposed GeTe (Di Sante et al. 2013), known as a FE semiconductor with high Curie temperature ($T_C = 720$ K), rhombohedrally distorted rock salt structure (space group $R3m$), and ferroelectric P parallel to [111] (Fig. 7a). First-principles predictions showed not only that a gigantic RE is present (see Fig. 7c) but also that a switchable spin texture can be obtained upon polarization switching (i.e., when applying opposite electric fields, Fig. 7d). GeTe was therefore considered the first example of FERSC. While GeTe showed groundbreaking potential from ab initio predictions, its experimental confirmation posed many challenges, due to deviations from the simulated "ideal" crystal: GeTe is usually a highly defective material, with large concentration of Ge vacancies which makes it a degenerate p-type semiconductor. In order to probe the RE and the whole FERSC concept, epitaxial (111) GeTe films were grown on Si(111) using molecular beam epitaxy, and FE properties were confirmed via piezo-force microscopy (PFM): as first discussed in Liebmann (2016), Fig. 7e
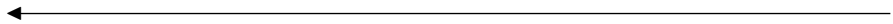
Fig. 7 Relevant properties of GeTe. (**a**) Rhombohedral unit cell of GeTe (polarization pointing along the pseudo-cubic [111] axis). (**b**) GeTe Brillouin zone. The direction corresponding to polarization, the related perpendicular plane, and the relevant A-Z-U directions are highlighted. (**c**) DFT-calculated (SOC included) valence maximum bands along ZA and ZU directions. The Rashba splitting is evident, with E_R ($\approx$230 meV) and k_R (≈ 0.09 Å^{-1}) Rashba parameters highlighted ($\alpha_R = 4.8$ eVÅ). (**d**) Theoretical CEM obtained by cutting the band structure in the ZAU plane at a binding energy of around -0.4 eV. The sixfold symmetry of the crystal is evident, while the circulation of the spins (see arrows for the inner and outer branches representing the spin-expectation value in the ZAU plane) is consistent with the RE. A reversal of the FE polarization P completely revers the sense of circulation (cfr left and right part). (**e**) FE hysteresis loop of GeTe(111) measured by piezo-force microscopy, with P pointing outward with respect to the sample surface. Inset: FE domains written by an AFM tip. (**b**) ARPES measured along ZA direction (for photon energy $\mathbf{h}\nu = 20$ eV). (**c**) Comparison between experimental and DFT-calculated surface (S_i) and bulk bands (B_i). (**d**) Measured CEM close to the Fermi level shows the expected "starlike" shape, in excellent agreement with theoretical data. Panels (**a–d**) are adapted from Di Sante et al. (2013); panels (**e–h**) are adapted from Liebmann (2016)

shows a clear FE hysteresis loop along with the possibility to write stable FE domains by applying a voltage bias to the tip of an atomic force microscope. It was also shown that the as-grown surface displays a FE stable state with polarization pointing outward with respect to the sample surface. Spin-ARPES measurements were performed and showed clearly visible Rashba bands along the ZA direction (cfr Fig. 7f, g). When performing a theory-experiment comparison, a remarkably good agreement was achieved, allowing to disentangle surface states and bulk bands and to conclude that the material shows a giant Rashba splitting both for surface and bulk bands. Moreover, Fig. 7h shows a constant energy map (CEM) close to the Fermi level: by probing the spin texture, the spin direction was found to lie predominantly perpendicular to the **k**-vector with a sense of circulation consistent with DFT predictions for an outward polarization, confirming the overall picture of a gigantic RE.

4.2   Additional Results on FERSC and Discussion

Since the first proposal on GeTe in 2013, many different bulk FE materials have been theoretically explored in view of possible switchable RE, ranging from oxides (Davanse da Silveira et al. 2016; Tao et al. 2017), to chalcogenides to small-gap semiconductors (Di Sante et al. 2016; Narayan 2015) and from "fully 3D" compounds to layered systems (Bruyer et al. 2016). In closer detail, a class of materials which has attracted an incredible attention in view of groundbreaking performances for photovoltaics are hybrid organic-inorganic halides, such as methylammonium lead iodide. Due to the presence of heavy elements (like I, Pb, Sn, etc.), SOC effects are relevant. This class of materials was therefore put forward as possible FERSC, in the case of (highly debated (Beilsten-Edmands et al. 2011)) FE ordering of the polar molecular groups (Stroppa et al. 2014; Kepenekian et al. 2015), with an envisaged crucial role of the RE for the excellent photovoltaic performances of these materials (Etienne et al. 2016; Zheng et al. 2015). From the experimental point of view, the research activity of FERSC is much less advanced, and, to our knowledge, GeTe remains the only experimentally studied FERSC so far.

5   Remarks and Conclusions

5.1   Theoretical Considerations

The aim of this subsection is to discuss potentialities and limits and advantages and disadvantages of DFT simulations in the framework of multiferroics and FERSC.

Let us first focus on multiferroics. In addition to the groundbreaking technological perspectives and to experimental synthesis/characterization challenges, from the theoretical point of view, part of the excitement toward this field was driven by the need to find new microscopic mechanisms that could give rise to the coupling between dipolar and spin degrees of freedom. As such, DFT with respect to other

techniques for materials modeling (e.g., devoted to multiscale or time-/temperature-dependent approaches) was particularly important and adequate. With respect to microscopic mechanisms, materials modeling in multiferroics has followed two parallel and complementary approaches, based on the one side on model Hamiltonians and on the other side on first-principles simulations. Undoubtedly, model-Hamiltonian approaches can provide a lot of hints on the microscopic origin of multiferroicity/magnetoelectricity; however, for a quantitative estimate of the different contributions involved, one needs to perform a DFT analysis, the latter being the only approach that can take into account all the different quantum-mechanical degrees of freedom at the same time and within a similar accuracy.

We would like to point out that performing DFT simulations for multiferroics (and complex oxides in general) is extremely challenging and the demands that the physics of multiferroics imposes on theory are extremely high. First, one has to deal with structural complexity: multiferroic oxides often show complicated structures, in presence of Jahn-Teller distortions, octahedral tilting and rotation of the oxygen cages, ferroelectric distortions, etc. Second, simulations need to take into account magnetic complexity: multiferroics often show exotic spin structures, such as non-collinear spin arrangements, like spirals, cycloids, spin canting, etc. Third, complex electronic correlations lead to an intricate interplay between spin, charge, and orbital degrees of freedom. In that case, conventional exchange-correlation functionals (like LDA (Jones and Gunnarsson 1989) or GGA (Perdew et al. 1996)) are generally not accurate enough, so alternative approaches such as the Hubbard-like DFT+U formalism (Anisimov et al. 1997) or hybrid functionals (Heyd 2003) are usually adopted for a better description of d- or f-electrons. However, both techniques are not "fully first principles," as they depend on parameters (the Hubbard-U and exchange-J within DFT+U (Anisimov et al. 1997) and the percentage of local vs Hartree exchange in hybrid functionals), usually provided by the users as external quantities. Finally, delicate relativistic effects have to be properly treated; for example, SOC is the driving mechanism in many electronic multiferroics, and simulations need to include SOC to evaluate the FE polarization in $TbMnO_3$.

As for FERSC, the prediction of the energy band gap is often crucial for a quantitative estimate of the Rashba effect, since the coupling between occupied and unoccupied states is believed to influence the magnitude of the Rashba spin splitting (Bahramy 2011). As is well known, DFT within conventional LDA or GGA exchange-correlation functionals underestimates the size of the gap, due to inaccurate treatment of excited states. Therefore, more rigorous techniques such as GW (Hybertsen 1986) or hybrid functionals (Heyd 2003) need to be used. However, a significant increase in the computational cost arises and makes the simulations accessible only when the unit cells are relatively small (say, of the order of 100 atoms for hybrid functionals).

As a final note, a word of caution to the reader is in order. Quantitative differences can occasionally arise in specific cases when comparing theory and experiments for multiferroics and FERSC, as far as polarization, structural properties, exchange coupling constants, anisotropy axis, energy band gaps, etc. are concerned. We however remark that DFT is very seldom incorrect in predicting *trends* vs strain,

pressure, doping, chemical constituents, etc. As such, DFT remains the technique of first choice when performing systematic studies of classes of materials, thereby including multiferroics and FERSC.

5.2   Outlook

The field of MF, after considerable progress in fundamental understanding in the last 15 years, is now moving forward to applications, especially in view of low-power electronics where exploiting electric fields rather than magnetic fields should bring the highest advantage (Scott 2000; Mundy 2016). Another promising area for ferroelectrics and MF could be related to energy harvesting in terms of light-to-electricity conversion, for example, exploiting dissipationless "shift currents" due to bulk photovoltaic effects in FE materials (Grinberg 2013). Although some progresses have been made specifically from the first-principles point of view (Young and Rappe 2012), this emerging field still requires a lot of fundamental understanding both from phenomenological models and from ab initio approaches. From the modeling point of view (both at the DFT and more phenomenological level), certainly an important field is that of FE/magnetic/MF domain walls, where totally different properties with respect to their corresponding "bulk-like" domains could be expected: different conductivities, different magnetic orders, and different tunability upon applied fields, all at the nanoscale (Seidel 2009; Meier et al. 2012). However, little theoretical work has been done so far, due to severe challenges, ranging from length scales (i.e., very large for first-principles simulations) to choosing appropriate boundary conditions, combined with still too few experimental data to compare with. A certainly exciting topic is the (linear) dynamical ME response, i.e., extending the static response to different frequency ranges (including optical one): $H(\omega)=\alpha(\omega)E(\omega)$. This involves the study of electromagnons (Pimenov 2006; Takahashi et al. 2013), spin wave excitations which are electrically active, where modeling from the first-principles point of view poses several challenges, as for the difficult treatment of excited and correlated states. Furthermore, nowadays topology is more and more commonly accepted as an additional degree of freedom in condensed matter. Little has been done in this respect in the field of MF, even less from the modeling point of view, although some topological FE insulators were addressed from first principles (Di Sante et al. 2016; Plekhanov et al. 2014; Liu et al. 2016). The connection between topology and correlations, gaining an increasing importance, might see MF to play a relevant role in the coming years. Within the same topology-related areas, an intriguing field where MF and ME could play a relevant role is magnetic skyrmions (Skyrme 1961; Roessler et al. 2006). We recall that skyrmions are vortex-like arrangements of spins (with typical size of tens of nanometers) whose spin texture is topologically nontrivial in real space and, being topologically protected macroscopic objects, can be used as nanometer-sized information carriers for high-memory density devices with low driving currents. Experimentally, in the chiral lattice of Cu_2OSeO_3 ferromagnetic oxide (Seki et al. 2012), skyrmions were found to induce FE polarization and could be electrically

controlled. Skyrmions were also observed in MF lacunar spinels, such as GaV_4S_8 (Kézsmarki 2015). From the modeling point of view, an outstanding challenge is the accurate and quantitatively reliable estimate of the DM interaction (Xiang et al. 2013a; Weingart et al. 2012), at the basis of the skyrmion emergence. In correlated materials, tiny quantities such as relativistic DM might show a strong dependence on the (often unknown) Hubbard parameters U and J within the DFT+U formalism, which questions the predictive power of DFT in the field of skyrmions and calls for other approaches to be developed and tested.

As for FERSC, the field is so new that it is hard to envisage how it will evolve. First of all, progress should be made from the experimental point of view. Indeed, GeTe was found (Liebmann 2016) via (spin)-ARPES to show a giant Rashba spin splitting, but the main concept of FERSC, i.e., an opposite spin polarization corresponding to opposite FE polarization, could not be detected yet, due to difficulties in switching ferroelectricity caused by overly conductive samples. In general, FERSC might evolve along three different lines (modeling being crucial in all of them): (i) from the fundamental point of view, it is not clear yet how to optimize/maximize the Rashba spin splitting and which are the main ingredients that influence it. In fact, the situation is more complex than what can be naively thought: the Rashba splitting is not simply proportional to polarization or atomic SOC, but depends on hybridization, orbital character, symmetry constraints, etc. In this context, the role of orbital angular momentum was put forward, leading to an "orbital Rashba" effect as a counterpart of the conventional spin RE (Kim 2013). (ii) From the materials science point of view, there is certainly a large activity to be carried out, as many FE materials exist and could potentially show large RE. More (spin)-ARPES studies on different FE compounds should therefore be performed in order to probe the band structure. (iii) From the technological point of view, FERSC offer intriguing perspectives in semiconductor spintronics, offering the appealing possibility to combine computing/logic and memory functions in a single material and in view of achieving efficient spin-to-charge conversion (Manchon et al. 2015). Many of the devices based on FERSC still have to be designed, with only a first conceptual example of a modified Datta-Das spin transistor reported in Di Sante et al. (2013). The concept of FERSC could also easily branch into neighboring areas: (1) topology, as SOC plays a crucial role in both RE and band inversion leading to nontrivial topological properties. Indeed, SnTe – a prototypical topological crystalline insulator (TCI) – was theoretically explored along this direction (Plekhanov et al. 2014), with the outcome that FERSC and TCI behaviors are not mutually exclusive. (2) Phase change materials (PCM) (Raoux and Wuttig 2009) are compounds which exist in an amorphous phase and in a crystalline phase. By means of optical pulses or electrical Joule heating, PCM can be rapidly and repeatedly switched between the amorphous and crystalline states (with related significant changes of their optical and electronic properties), which eventually allow them to be used for data storage. Being GeTe one of the paradigmatic examples of PCM, the connection between FERSC and PCM physics might be explored in the near future. (3) A plethora of interesting spin-related effects, in addition to the RE, might arise when inversion symmetry is broken in

nonmagnetic compounds; most of them are mediated by SOC and range from spin-valley coupling (Xu et al. 2014; Yamauchi et al. 2015) to electronic nematic phases in correlated metals (Harter et al. 2017). We also mention the theoretically proposed concept of "hidden spin polarization" (HSP) (Zhang et al. 2014), starting from the observation that SOC-induced effects are caused by the *local* rather than by the *global* symmetry of a specific material. Therefore, although in a crystal with SIS, all the bands are enforced by symmetry to be spin degenerate, the electric field is not forced to vanish everywhere (e.g., there might be an inversion asymmetry around a specific atom). As such, it might locally induce a spin polarization in the degenerate bands. A key feature of HSP is that the contributions to such bands are spatially localized on different parts/atoms of the unit cell (i.e., "inversion partners") and show opposite spin textures, so that an overall compensation of the spin polarization is present over the bulk crystal (Slawinska et al. 2016).

We conclude by saying that materials modeling in both FERSC and MF has been, since their birth, absolutely crucial and will likely continue to be so, due to the complexity of the involved phenomena and to the presence of coexisting/competing effects. The latter can be hardly addressed, unless a first-principles approach – which takes into account all the degrees of freedom at the same time – is employed. In this respect, DFT will undoubtedly play a major role for the future development of FERSC and MF and for their prospective scientific and technological impact.

References

Alexe M, Ziese M, Hesse D, Esquinazi P, Yamauchi K, Fukushima K, Picozzi S, Goesele U (2009) Ferroelectric switching in multiferroic magnetite (Fe_3O_4) thin films. Adv Mater 21:4452–4455

Anisimov V, Aryasetiawan F, Lichtenstein A (1997) First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA + U method. J Phys Condens Matter 9:767

Bahramy SA (2011) Origin of giant bulk Rashba splitting: application to BiTeI. Phys Rev B 84:041202

Barone P, Kanungo S, Picozzi S, Saha-Dasgupta T (2011) Mechanism of ferroelectricity in d^3 perovskites: a model study. Phys Rev B 84:134101

Bea H, Gajek M, Bibes M, Barthélemy A (2008) Spintronics with multiferroics. J Phys Condens Matter 20:434221

Beilsten-Edmands J, Eperon GE, Johnson RD, Snaith HJ, Radaelli PG (2011) Non-ferroelectric nature of the conductance hysteresis in $CH_3NH_3PbI_3$ perovskite-based photovoltaic devices. Appl Phys Lett 631 106:173502

Benedek N, Fennie C (2011) Hybrid improper ferroelectricity: a mechanism for controllable polarization-magnetization coupling. Phys Rev Lett 106:107204

Bhattacharjee S, Bousquet E, Ghosez P (2009) Engineering multiferroism in $CaMnO_3$. Phys Rev Lett 102:117602

Bihlmayer G (2015) Focus on the Rashba effect. New J Phys 17:050202

Bousquet E, Dawber M, Stucki N, Lichtensteiger C, Harmet P, Gariglio S, Triscone JM, Ghosez P (2008) Improper ferroelectricity in perovskite oxide artificial superlattices. Nature 452:732

Bruyer E, Di Sante D, Barone P, Stroppa A, Wahngbo M, Picozzi S (2016) Possibility of combining ferroelectricity and Rashba-like spin splitting in monolayers of the 1T-type transition-metal dichalcogenides $MX_2(M=Mo,W; X=S,Se,Te)$. Phys Rev B 94:195402

Catalan G, Scott J (2009) Physics and applications of bismuth ferrite. Adv Mater 21:2463

Cheong S, Mostovoy M (2007) Multiferroics: a magnetic twist for ferroelectricity. Nat Mater 6:13–20

Crepaldi A, Moreschini L, Autès G, Tournier-Colletta C, Moser S, Virk N, Berger H, Bugnon Ph Chang YJ, Kern K, Bostwick A, Rotenberg E, Yazyev OV, Grioni M (2012) Giant ambipolar Rashba effect in the semiconductor BiTeI. Phys Rev Lett 109:096803

da Silveira L, Barone P, Picozzi S (2016) Rashba-Dresselhaus spin-splitting in the bulk ferroelectric oxide $BiAlO_3$. Phys Rev B 93:245159

Di Sante D, Barone P, Bertacco R, Picozzi S (2013) Electric control of the giant Rashba effect in bulk GeTe. Adv Mater 25:509

Di Sante D, Barone P, Stroppa A, Garrity K, Vanderbilt D, Picozzi S (2016) Intertwined Rashba, Dirac, and Weyl fermions in hexagonal hyperferroelectrics. Phys Rev Lett 646 117:07640

Dong S, Liu J, Cheong SR (2015) Multiferroic materials and magnetoelectric physics: symmetry, entanglement, excitation, and topology. Adv Physics 64:519–626

Ederer C, Spaldin N (2005) Weak ferromagnetism and magnetoelectric coupling in bismuth ferrite. Phys Rev B 71:060401(R)

Etienne T, Mosconi E, De Angelis F (2016) Dynamical origin of the Rashba effect in organohalide lead perovskites: a key to suppressed carrier recombination in perovskite solar cells? J Phys Chem Lett 7:1638–1645

Farokhipoor S, Noheda B (2011) Conduction through 71 domain walls in $BiFeO_3$ thin films. Phys Rev Lett 107:127601

Giovannetti G, Kumar S, van den Brink J, Picozzi S (2009a) Magnetically induced electronic ferroelectricity in half-doped manganites. Phys Rev Lett 103:037601

Giovannetti G, Kumar S, Khomskii D, Picozzi S, van den Brink J (2009b) Multiferroicity in rare-earth nickelates. $RNiO_3$ Phys Rev Lett 653 103:156401

Grinberg I, West DV, Torres M, Gou G, Stein DM, Wu L, Chen G, Gallo EM, Akbashev AR, Davies PK, Spanier JE, Rappe AM (2013) Substantial bulk photovoltaic effect enhancement via nanolayering. Nature 503:509–512

Harris A (2007) Landau analysis of the symmetry of the magnetic structure and magnetoelectric interaction in multiferroics. Phys Rev B 76:054447

Harrison W (1989) Electronic structure and the properties of solids. Dover Publications, New York

Harter J, Zhao Z, Yan J, Mandrus D, Hsieh D (2017) A parity-breaking electronic nematic phase transition in the spin-orbit coupled metal $Cd_2Re_2O_7$. Science 356:295–299

Hasan M, Kane C (2010) *Colloquium*: topological insulators. Rev Mod Phys 82:3045–3067

Heron JT, Trassin M, Ashraf K, Gajek M, He Q, Yang SY, Nikonov DE, Chu Y-H, Salahuddin S, Ramesh R (2011) Electric-field-induced magnetization reversal in a ferromagnet-multiferroic heterostructure. Phys Rev Lett 107:217202

Heyd JS (2003) Hybrid functionals based on a screened Coulomb potential. J Chem Phys 118:8207

Hill NA (2000) Why are there so few magnetic ferroelectrics? J Phys Chem B 104:6694–6709

Hybertsen ML, Louie S (1986) Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. Phys Rev B 34:5390

Ishizaka K, Bahramy MS, Murakawa H, Sakano M, Shimojima T, Sonobe T, Koizumi K, Shin S, Miyahara H, Kimura A, Miyamoto K, Okuda T, Namatame H, Taniguchi M, Arita R, Nagaosa N, Kobayashi K, Murakami Y, Kumai R, Kaneko Y, Onose Y, Tokura Y (2011) Giant Rashba-type spin splitting in bulk BiTeI. Nat Mater 10:521

Jia C, Onoda S, Nagaosa N, Han J (2007) Microscopic theory of spin-polarization coupling in multiferroic transition metal oxides. Phys Rev B 76:144424

Johnson R, Barone P, Bombardi A, Bean R, Picozzi S, Radaelli P, Oh YS, Cheong S-W, Chapon LC (2013) X-ray imaging and multiferroic coupling of cycloidal magnetic domains in ferroelectric monodomain $BiFeO_3$. Phys Rev Lett 110:217206

Jones R, Gunnarsson O (1989) The density functional formalism, its applications and prospects. Rev Mod Phys 61:689

Jungwirth T, Sinova J, Masek J, Kucera J, Mac Donald AH (2006) Theory of ferromagnetic (III,Mn)V semiconductors. Rev Mod Phys 78:809

Katsura H, Balatsky A, Nagaosa N (2007) Dynamical magnetoelectric coupling in helical magnets. Phys Rev Lett 98:027203

Kepenekian M, Robles R, Katan C, Sapori D, Pedessau L, Even J (2015) Rashba and Dresselhaus effects in hybrid organic–inorganic perovskites: from basics to devices. ACS Nano 9: 11557–11567

Kézsmarki I, Bordács S, Milde P, Neuber E, Eng LM, White JS, Rønnow HM, Dewhurst CD, Mochizuki M, Yanai K, Nakamura H, Ehlers D, Tsurkan V, Loidl A (2015) Néel-type skyrmion lattice with confined orientation in the polar magnetic semiconductor GaV_4S_8. Nat Mater 14:1116

Khomskii D (2009) Trend: classifying multiferroics: mechanisms and effects. Physics 2:20

Kim B, Kim P, Jung W, Kim Y, Koh Y, Kyung W, Park J, Matsunami M, Kimura S, Kim JS, Han JH, Kim C (2013) Microscopic mechanism for asymmetric charge distribution in Rashba-type surface states and the origin of the energy splitting scale. Phys Rev B 88:205408

Kimura T (2007) Spiral magnets as magnetoelectrics. Annu Rev Mater Res 37:387–413

Kimura T, Goto T, Shintani H, Ishisaka K, Arima T, Tokura Y (2003) Magnetic control of ferroelectric polarization. Nature 426:55–58

King-Smith R, Vanderbilt D (1993) Theory of polarization of crystalline solids. Phys Rev B 47:1651–1654

Lebeugle D, Colson D, Forget A, ViretM, Bataille A, Gukaso A (2008) Electric-field-induced spin flop in $BiFeO_3$ single crystals at room temperature. Phys Rev Lett 100:227602

Liebmann M, Rinaldsi C, Di Sante D, Kellner J, Pauly C, Wang RN, Boschker JE, Giusssani A, Bertoli S, Cantoni M, Baldrati L, Asa M, Vobornik I, Panaccione G, Marchenko D, Sanchez-Barriga J, Rader O, Calarco R, Picozzi S, Bertacco R, Morgenstern M (2016) Giant Rashba-type spin splitting in ferroelectric GeTe(111). Adv Mater 28:560–565

Liu S, Kim Y, Tan L, Rappe A (2016) Strain-induced ferroelectric topological insulator. Nano 16:1663–1668

Lorentz B, Wang Y, Chu W (2007) Ferroelectricity in perovskite $HoMnO_3$ and $YMnO_3$. Phys Rev B 76:104405

Malashevic A, Vanderbilt D (2008) First principles study of improper ferroelectricity in $TbMnO_3$. Phys Rev Lett 101:037210

Manchon A, Koo H, Nitta J, Frolov S, Duine R (2015) New perspectives for Rashba spin–orbit coupling. Nat Mater 14:871

Meier D, Seidel J, Cano A, Delaney K, Kumagai Y, Mostovoy M, Spaldin NA, Ramesh R, Fiebig M (2012) Anisotropic conductance at improper ferroelectric domain walls. Nat Mater 11:284

Mochizuki M, Nagaosa N, Tokura Y (2011) Theory of spin-phonon coupling in multiferroic manganese perovskites RMnO$_3$ Phys Rev B 84:144409

Mundy JA, Brooks CM, Holtz ME, Moyer JA, Das H, Rebola AF, Heron JT, Clarkson JD, Disseler SM, Liu Z, Farhan A, Held H, Hovden R, Padgett E, Mao Q, Paik H, Misra R, Kourkoutis LR, Arenholz E, Scholl A, Borchers JA, Radcliff WD, Ramesh R, Fennie CJ, Schiffer P, Mueller DA, Schlom DG (2016), Atomically engineered ferroic layers yield a room-temperature magnetoelectric multiferroic. Nature 537:523–527

Nakamura M, Tokunaga Y, Taguchi Y, Tokura Y (2011) Multiferroicity in an orthorhombic $YMnO_3YMnO_3$ single-crystal film. Appl Phys Lett 98:082902

Narayan A (2015) Class of Rashba ferroelectrics in hexagonal semiconductors. Phys Rev B 92:220101

Neaton JB, Ederer C, Waghmare UV, Spaldin NA, Rabe KM (2005) First-principles study of spontaneous polarization in multiferroic $BiFeO_3$. Phys Rev B 71:014113

Ohno H (1999) Properties of ferromagnetic III–V semiconductors. J Magn Magn Mat 1:110

Orobengoa D, Capillas C, Aroyo M, Perez-Mato J (2009) AMPLIMODES: symmetry-mode analysis on the Bilbao Crystallographic Server. J Appl Cryst 42:820–833

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865

Perez-Mato J, Aroyo M, Garcia A, Blaha P, Schwarz K, Schweifer K, Parlinski K (2004) Competing structural instabilities in the ferroelectric Aurivillius compound $SrBi_2Ta_2O_9$. Phys Rev B 70:214111

Picozzi S (2014) Ferroelectric Rashba semiconductors as a novel class of multifunctional materials. Front Phys 2:10

Picozzi S, Ederer C (2009) First principles studies of multiferroic materials. J Phys Condens Matter 21:303201

Picozzi S, Yamauchi K, Sanyal B, Sergienko I, Dagotto E (2007) Dual nature of improper ferroelectricity in a magnetoelectric multiferroic. Phys Rev Lett 99:227201

Pimenov A, Mukhin AA, Ivanov VYU, Travkin VD, Balbashov AM, Loidl A (2006) Possible evidence for electromagnons in multiferroic manganites. Nat Phys 2:97–100

Plekhanov E, Barone P, Di Sante D, Picozzi S (2014) Engineering relativistic effects in ferroelectric SnTe. Phys Rev B 90:161108(R)

Rabe KT (2007) Physics of ferroelectrics: a modern perspective. Springer, Berlin

Ramesh R, Spaldin N (2007) Multiferroics: progress and prospects in thin films. Nat Mater 6:21–29

Raoux S, Wuttig M (2009) Phase change materials: science and application. Springer, New York

Rashba E (1960) Properties of semiconductors with an extremum loop. 1. Cyclotron and combinational resonance in a magnetic field perpendicular to the plane of the loop. Sov Phys Solid State 2:1109

Resta R (1994) Macroscopic polarization in crystalline dielectrics: the geometric phase approach. Rev Mod Phys 66:899–916

Roessler U, Bogdanov A, Pfleiderer C (2006) Spontaneous skyrmion ground states in magnetic metals. Nature 442:797

Schmid H (1994) Multi-ferroic magnetoelectrics. Ferroelectrics 162:317–338

Scott J (2000) Ferroelectric memories. Springer, Berlin

Seidel J, Martin LW, He Q, Zhan Q, Chu YH, Rother A, Hawkridge ME, Maksymovich P, Yu P. Gajek M, Balke N, Kalinin SV, Gemming S, Wang F, Catalan G, Scott JF, Spaldin NA, Orenstein, Ramesh R (2009) Conduction at domain walls in oxide multiferroics. Nat Mater 8:229

Seki S, Yu X, Ishiwata S, Tokura Y (2012) Observation of skyrmions in a multiferroic material. Science 336:198

Sergienko I, Dagotto E (2006) Role of the Dzyaloshinskii-Moriya interaction in multiferroic perovskites. Phys Rev B 73:094434

Sergienko I, Sen C, Dagotto E (2006) Ferroelectricity in the magnetic E-Phase of orthorhombic perovskites. Phys Rev Lett 97:227204

Seshadri R, Hill N (2001) Visualizing the role of Bi 6s "lone pairs" in the off-center distortion in ferromagnetic $BiMnO_3$. Chem Mater 13:2892

Skyrme T (1961) A nonlinear field theory. Proc R Soc A 260:127

Slawinska J, Narayan A, Picozzi S (2016) Hidden spin polarization in nonmagnetic centrosymmetric $BaNiS_2$ crystal: signatures from first principles. Phys Rev B 94:241114(R)

Stokes HA (n.d.). ISOTROPY software suite. Retrieved from http://iso.byu.edu

Stroppa A, Barone P, Jain P, Perez-Mato J, Picozzi S (2013) Hybrid improper ferroelectricity in a multiferroic and magnetoelectric metal-organic framework. Adv Mater 25:2284

Stroppa A, Di Sante D, Barone P, Bokdam M, Kresse G, Franchini C, Whangbo M-H, Picozzi S (2014) Tunable ferroelectric polarization and its interplay with spin–orbit coupling in tin iodide perovskites. Nat Commun 5:5900

Takahashi Y, Yamasaki Y, Tokura Y (2013) Terahertz magnetoelectric resonance enhanced by mutual coupling of electromagnons. Phys Rev Lett 111:037204

Tao L, Paudel T, Kovalev A, Tsymbal E (2017) Reversible spin texture in ferroelectric HfO_2. Phys Rev B 95:245141

Tokunaga Y, Iguchi S, Arima T, Tokura Y (2008) Magnetic-field-induced ferroelectric state in $DyFeO_3$. Phys Rev Lett 101:097205

Tokura Y, Seki Y, Nagaosa N (2014) Multiferroics of spin origin. Rep Prog Phys 77:076501

Wang J, Neaton J, Zheng H, Nagarajan V, Ogale S, Liu B, Viehland D, Vaithyanathan V, Schlom DG, Waghmare UV, Spaldin NA, Rabe KM, Wuttig M, Ramesh R (2003) Epitaxial $BiFeO_3$ multiferroic thin film heterostructures. Science 299:1719–1722

Weingart C, Spaldin N, Bousquet E (2012) Noncollinear magnetism and single-ion anisotropy in multiferroic perovskites. Phys Rev B 86:094413

Wu H, Burnus T, Hu Z, Martin C, Maignan A, Cezar JC, Tanaka A, Brookes NB, Khomskii DI, Tjeng LH (2009) Ising magnetism and ferroelectricity in Ca_3CoMnO_6. Phys Rev Lett 102:026404

Xiang H, Wei S, Whangbo M, Silva J (2008) Spin-orbit coupling and ion displacements in multiferroic $TbMnO_3$. Phys Rev Lett 101:037209

Xiang H, Lee C, Koo H, Gong X, Whangbo M (2013a) Magnetic properties and energy-mapping analysis. Dalton Trans 42:823–853

Xiang H, Wang P, Whangbo MH, G X (2013b) Unified model of ferroelectricity induced by spin order. Phys Rev B 88:054404

Xu X, Yao W, Xiao D, Heinz T (2014) Spin and pseudospins in layered transition metal dichalcogenides. Nat Phys 10:343

Yamauchi K, Fukushima T, Picozzi S (2009) Ferroelectricity in multiferroic magnetite Fe_3O_4 driven by noncentro-symmetric Fe^{2+}/Fe^{3+} charge-ordering: first-principles study. Phys Rev B 79:212404

Yamauchi K, Barone P, Shishidou T, Oguchi T, Picozzi S (2015) Coupling ferroelectricity with spin-valley physics in oxide-based heterostructures. Phys Rev Lett 115:037602

Young S, Rappe A (2012) First Principles Calculation of the Shift Current Photovoltaic Effect in Ferroelectrics, Phys Rev Lett 109:116601

Yu PC (2005) Fundamentals of semiconductors: physics and materials properties, 3rd edn. Springer, Berlin

Zeches RJ, Rossell MD, Zhang JX, Hatt AJ, He Q, Yang C-H, Kumar A, Wang CH, Melville A, Adamo C, Sheng G, Chu Y-H, Ihlefeld JF, Erni R, Ederer C, Gopalan V, Chen LQ, Schlom DG, Spaldin NA, Martin LW, Ramesh R (2009) A strain-driven morphotropic phase boundary in $BiFeO_3$. Science 326:977–980

Zhang X, Liu Q, Luo J, Freeman A, Zunger A (2014) Hidden spin polarization in inversion-symmetric bulk crystals. Nat Phys 10:387

Zheng F, Tan L, Liu S, Rappe A (2015) Rashba spin–orbit coupling enhanced carrier lifetime in $CH_3NH_3PbI_3$. Nano Lett 15:7794–7800

Applications of Multi-scale Modeling to Spin Dynamics in Spintronics Devices

19

Riccardo Hertel

Contents

Abstract

The design of modern spintronic applications relies on the accurate simulation of nanomagnetic structures in realistic situations. Simulations can be particularly challenging when they need to consider effects that unfold on vastly different length scales. Multiscale modeling techniques are a powerful way to treat such cases, especially when the different length scales are coupled, such

R. Hertel (✉)
Université de Strasbourg, CNRS, Institut de Physique et Chimie des Matériaux de Strasbourg, Strasbourg, France
e-mail: riccardo.hertel@ipcms.unistra.fr

that the coarse length scale cannot be modeled without considering the fine scale or vice versa. This chapter discusses various approaches and features of multiscale simulations of the magnetization in micro- and nanomagnetism. It gives an overview of numerical techniques used in this scientific domain to bridge different length scales and discusses recent multiscale studies in three-dimensional nanomagnetism. The transition to three dimensions that is currently occurring in nanomagnetic research entails new challenges for the modeling of the magnetization. One of the reasons is the occurrence of Bloch points, as three-dimensional magnetic switching processes often involve the nucleation and propagation of these singularities of the magnetization field. Bloch points are three-dimensional magnetization structures that extend over several tens of nanometers and which require an atomistic treatment of their core region. Fully dynamic micromagnetic multiscale simulations with atomistic-continuum coupling have recently been used to simulate the field-driven motion of Bloch points. The transfer of multiscale modeling known from other domains of material science to the simulation of magnetic structures has made it possible to perform numerical studies providing insight into the structure and dynamics of Bloch points and their interaction with the atomic lattice.

1   Introduction

Simulation studies play a central role in the research on magnetic nanostructures and their application in spintronic devices. Modern computer simulations can provide insight into magnetic processes and structures that are often inaccessible by experimental observation, and they can thereby support the design of spintronic applications. To ensure high reliability and ultimately predictive power for such purposes, simulations must be applicable beyond idealized model cases. In order to reproduce realistic situations, simulations often have to take into account processes and effects that occur on disparate length scales. In spite of the ongoing increase of available computational power and the advent of GPU-based supercomputing, which has made it possible to perform micromagnetic simulations of ever larger systems over the past years (Kakay et al. 2010; Li et al. 2010; Lopez-Diaz et al. 2012; Van de Wiele et al. 2012), this progress is, in itself, not sufficient to achieve multiscale functionality that is capable to bridge length scales differing by orders of magnitude. Unlike general-purpose simulation algorithms, which can be applied out-of-the-box to a variety of situations by simply changing the input parameters, multiscale modeling techniques generally require a thorough preparation of the model that is used for the specific situation, such as a spatial subdivision into a fine-scale and a coarse-scale domain. Multiscale modeling tasks are particularly complicated when the fine scale is treated with a different physical model than the coarse scale, as is the case in simulations with atomistic-continuum coupling. This chapter discusses the application of multiscale modeling techniques to micro- and nanomagnetic systems. It is argued that multiscale simulations of the magnetic structure could be particularly helpful to address the new challenges which

arise from the recent trend toward three-dimensional nanomagnetism (Fernández-Pacheco et al. 2017). In particular, a combination of atomistic and continuum models is often necessary to reliably simulate magnetic switching processes which involve the propagation of Bloch points (Arrott et al. 1979). These topological structures, discussed in detail in Sect. 4, are point defects of the magnetization, and they represent a singularity in the continuum model. They require atomistic models to address the features of their core region.

2   Overview of Multiscale Modeling in Magnetism

The literature on multiscale modeling in magnetism comprises a large spectrum of studies, treating a wide variety of situations and often using vastly different numerical methods. Before entering deeper into the subject, it is useful to briefly introduce the main approaches of multiscale simulations in solid-state magnetism. A detailed classification of the numerous multiscale models applied in the broader scope of material science can be found in Miller and Tadmor (2009).

Most multiscale simulations are effectively dual-scale, i.e., they do not include a multitude of different length scales but only two: a coarse scale and a fine scale. In the implementation, the models representing these length scales can be connected to each other in different ways. Multiscale techniques are particularly important in situations where the different length scales are strongly connected, i.e., when an accurate simulation of the coarse or macroscopic scale is not possible without considering microscopic effects unfolding on the fine length scale and vice versa. The usual approach consists in subdividing the problem into distinct regions which are treated with a different degree of accuracy, according to the local requirements. A locally higher accuracy can be achieved in the fine-scale region by refining the discretization cells or by replacing the physical model with a more microscopic one or both. When different physical models are employed, it is important to define suitable coupling mechanisms between the fine-scale and the coarse region. In material modeling, the first multiscale simulations involving an atomistic model embedded into a macroscopic continuum emerged in the 1990s (Kohlhoff et al. 1991). These models were used to investigate the propagation of cracks in brittle materials. More recently, many similar types of multiscale methods have been applied to simulate the spin structure and dynamics in magnetic materials (Jourdan et al. 2008; Andreas et al. 2014b; De Lucia et al. 2016; Poluektov et al. 2018).

Two main branches of multiscale simulations can be distinguished depending on whether or not different physical models are employed to treat the coarse scale and the fine scale, respectively. To emphasize this qualitative difference, the term *multimodel* simulations is sometimes used to describe algorithms that combine distinct modeling paradigms, such as atomistic and continuum.

In simulations involving only one physical model, the main challenge consists in solving the underlying set of equations in a large domain in which strong spatial variations of the accuracy requirements need to be considered while keeping the numerical costs within acceptable limits (Pechstein 2013). This is typically achieved

by employing discretization cells of strongly varying sizes, where a refined mesh is used in the fine-scale region. Sometimes the grid in the fine-scale region is said to be "fully refined," which means that the cell size is refined down to the atomic scale. From a technical perspective, it is usually relatively easy to bridge several length scales by means of local mesh refinement, as long as the underlying equations and the physical model remain the same. Multimodel simulations, on the other hand, are more complex as they require a seamless coupling between the fine-scale and the coarse-scale domains that are governed by different physical models. In the case of simulations with atomistic-continuum coupling, special strategies must be applied at the interface region. The coupling strategies typically involve a hybrid transition region where both models are used and fully refined meshes of the continuum model near the interface.

Another distinguishing feature of multiscale codes is whether the region of interest is spatially defined a priori or whether its location has to be determined automatically by the algorithm. Often the geometrical setup or the physical situation is such that the location of the critical region which requires a fine-scale treatment is known beforehand. In these cases, the model can be prepared accordingly, and the simulation can rely on preprocessed meshes and predefined model parameters. The situation is more complicated if the location of the region of interest needs to be determined dynamically. Such cases require a continuous adaptation of the model during the simulation, controlled by suitable criteria and indicators to identify the regions necessitating a fine-scale treatment. Because of this difference between the simulations, the term "dynamic" can be ambiguous in multiscale modeling, since it may not be immediately clear whether it only refers to the simulation of the magnetization dynamics or also to the spatial dynamics of the multiscale model.

In some cases, multiscale simulations not only combine effects unfolding on vastly different length scales but also on very different time scales. A typical example are the highly complex processes governing the ultrafast laser-induced magnetization dynamics (cf. e.g., Koopmans et al. 2010), which include electronic, phononic, and magnetic effects spanning from the femtosecond to the nanosecond range. Such problems can often be treated in the framework of sequential models (Sect. 2.4). Usually, however, multiscale simulations refer to the simulation of systems that are evolving on a common time scale but on different length scales.

2.1 Multiscale Simulation with Adaptive Meshes

In the modeling of magnetic structures, single-model multiscale simulations with adaptive mesh refinement generally address situations where the domain in which the solution is sought is much larger than the micromagnetic exchange length, discussed in Sect. 3.2. This material-dependent length scale is always in the range of at least several nanometers and is comparable to the extension of fundamental magnetic structures, like the width of domain walls or the size of vortex cores. Micromagnetic simulations generally require that the discretization size does not exceed the exchange length, but in macroscopic samples, it can be prohibitively

expensive to use cell sizes that are everywhere small enough to resolve this fine structure of the magnetization. In these cases adaptive mesh refinement strategies can be used to detect the regions which require a local refinement while using comparatively coarse cells where the magnetic structure is sufficiently smooth (Forster et al. 2002; Hertel and Kronmüller 1998; Hertel et al. 2005). In micromagnetism, the gain in computational resources obtained by adaptive mesh refinement methods has become much less important since the advent of GPU-based high-performance computing (Kakay et al. 2010; Li et al. 2010) which allows for rapid large-scale computations with desktop computers (Van de Wiele et al. 2012; Gliga et al. 2013). Moreover, data compression techniques by means of hierarchical matrix methods (Hackbusch 2015; Börm and Grasedyck 2004) have contributed significantly to the reduction of numerical costs in micromagnetic finite element modeling, thereby enabling their application to large-scale simulations (Knittel et al. 2009; Popović and Praetorius 2005).

In continuum simulations, the size of the discretization cells is merely a numerical parameter without physical meaning. While a decrease of the cell size generally reduces the discretization error due to the discrete numerical representation of the continuum equations, it does not affect the nature of the physical problem. Accordingly, there is no lower limit concerning the admissible cell size, which can be chosen even smaller than the atomistic scale. In fact, early micromagnetic simulation studies attempted to eliminate discretization errors by extrapolating the results to zero cell size (infinite discretization density) (Rave and Hubert 2000; Hertel and Kronmüller 2002). Choosing too small discretization cells is normally a waste of computational resources since the results tend to converge rapidly below a certain cell size and remain unaffected by further refinement, but there is no fundamental criterion that would prevent an infinitely fine discretization.

A transition toward an atomistic model is required when the continuum model ceases to be sufficiently accurate. In such situations, coarse-grained models may fail to capture essential physical features that can only be described on a more fine-grained level. The need for a transition from the continuum to an atomistic model results from physical considerations, which will be discussed in Sect. 3.4.

2.2 Atomistic-Continuum Coupling: Finite Differences

While finite element models are prominent in the continuum modeling of materials in general, finite difference formulations are widespread in micromagnetic continuum simulations because they provide a convenient framework for the computation of the long-range magnetostatic interaction. One way to achieve a combination of atomistic and continuum model in magnetic multiscale simulations based on finite differences consists in refining the discretization grid down to the atomistic level and connecting the surrounding continuum region by means of a coarse-graining procedure. Such models have been applied successfully, e.g., in the simulations of skyrmions (De Lucia et al. 2017) and grain boundaries (Garcia-Sanchez et al. 2005). In these models, the discretization cells in the atomically refined region are
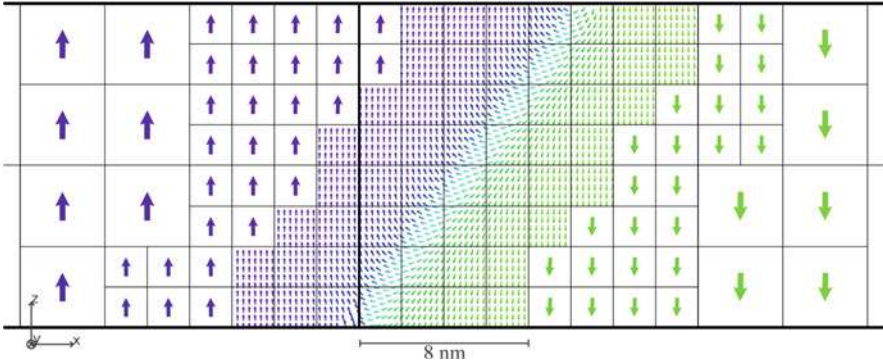
Fig. 1 Example of a finite difference multiscale simulation with atomistic resolution in a domain wall. The cell size is gradually reduced to the atomic scale as the domain wall is approached. Reprinted figure with permission from (Jourdan et al. 2008) Copyright 2008 by the American Physical Society

interpreted as atomistic magnetic moments. The interface between the continuum and the atomistic region can either be sharp (De Lucia et al. 2016) or gradual (Jourdan et al. 2008) (Fig. 1).

Unlike finite element methods, where a continuous variation of the cell size is unproblematic, a local change of the discretization size in finite difference formulations involves averaging procedures and assumptions that connect the smaller cells to the larger ones. In micromagnetic modeling, such coarse-graining procedures can play the role of an atomistic-continuum interface, where the discretization cells belonging to the continuum adjacent to the atomistic region effectively represent coarse-grained atomistic moments. The atomistic cells, on the other hand, can be coupled to the larger cells by means of "ghost" nodes, i.e., magnetic moments located inside the continuum cell, thereby creating an overlap between atomistic and continuum region. This technique can be applied to any level of refinement and coarsening and is not limited to the atomistic-continuum transition.

In micromagnetic finite difference formulations, the equations applied to the atomistic region are often very similar to those used in the continuum region, with the only qualitative difference being the number of interacting neighbors. This similarity arises from the fact that the discretized version of the continuum term of the exchange energy takes the same form as the atomistic Heisenberg exchange model (Donahue and McMichael 1997). Also, the dipolar field can be computed with the same summation scheme for micromagnetic and atomistic cells (Jourdan et al. 2008). These types of multiscale formulations therefore share similarities with mesh refinement techniques (Garcia-Cervera and Roma 2006) applied within a single modeling paradigm. They can nevertheless be regarded as models with genuine atomistic-continuum coupling, since the refinement of the cells down to the atomic scale introduces numerical entities that share sufficient similarities with a Heisenberg model so that the refined part can be interpreted as an atomistic region. While a local cell refinement down to the atomic scale provides an acceptable

model for the atomic part, it should be noted that the atomic cells in this type of multiscale simulations are located on a regular cubic grid defined by the discretization scheme. From a crystallographic perspective, such an arrangement is unrealistic for magnetic materials, and thus one drawback of this method is that the cells representing the atomic moments are not placed at realistic atomic positions of a crystal lattice.

2.3   Atomistic-Continuum Coupling: Finite Elements

In other domains of material modeling, multimodel simulations with atomistic-continuum coupling often provide a more realistic representation of the atomistic structure in the fine-scale region (Tadmor and Miller 2012; Zamora et al. 2012). In these simulations, atomically small cells of the continuum model are used as an ingredient of the interface region but are not directly identified as parts of an atomistic model. By using the finite element method for the continuum and finite difference formulation for the atomistic part, they combine qualitatively different models, which are adapted to the modeling of the respective scales (Fig. 2, right).

In magnetism, a simulation concept with such atomistic-continuum coupling was suggested a long time ago by Kronmüller et al. (1997) as a possibility to treat grain boundaries in composite permanent magnets (Fig. 2, left). The grain boundary problem has later been treated within a sequential atomistic-continuum multiscale approach (Hrkac et al. 2010) (cf. Sect. 2.4) and by using finite difference multiscale simulations with atomistic discretization (Garcia-Sanchez et al. 2005). Finite element implementations of multimodel/multiscale simulations with atomistic-continuum coupling have been reported only recently in micro- and
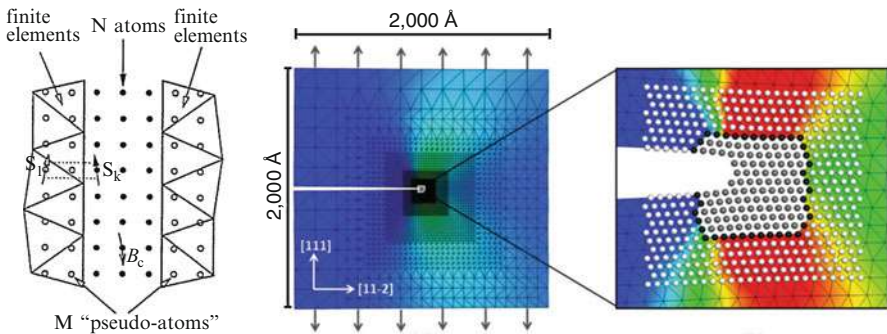
Fig. 2 Left: Schematic representation of a combined finite element/atomistic model to simulate magnetic grain boundaries. Reprinted from (Kronmüller et al. 1997), Copyright 1997, with permission from Elsevier. Right: Example of a finite element/multiscale study with atomistic-continuum coupling, simulating the propagation of a crack in brittle material. The finite element mesh is gradually refined to the atomistic scale, where it couples to the atomistic region. Reprinted figure with permission from (Zamora et al. 2012) Copyright 2012 by the American Physical Society

nanomagnetism (Poluektov et al. 2018; Andreas et al. 2014b). In such models, the coupling mechanism between the atomistic fine-scale domain and an adjacent continuum domain is usually the most complicated aspect. This connection is typically achieved by means of some kind of overlap of the atomistic region extending into the continuum domain, which can contain a "handshake" region where both models are evaluated. The handshake region includes contributions from the atomistic and the continuum model, which are weighted depending on the distance from the interface, thereby achieving a smooth transition from purely continuum to fully atomistic. In addition to these strategies, the continuum domain is often fully refined at the interface region, such that the cell size is compatible to the atomic lattice constant. Due to the complexity of the underlying meshes and their connection within an overlapping region, it is challenging to apply these models to situations in which the location of the atomistic-continuum interfaces changes dynamically. Fully dynamic three-dimensional micromagnetic multiscale code with atomistic-continuum coupling has recently been reported by Andreas et al. (2014b). In their algorithm, atomistic fine-scale regions are inserted and displaced automatically according to the peculiarities of the magnetic structure that develops during the simulation.

2.4   Sequential Models

In micro- and nanomagnetic simulations, the term multiscale modeling is also used occasionally to describe models that connect different length scales in a sequential way. Unlike the previously discussed cases, where coarse and fine scale are solved concurrently, sequential models take the results obtained from the modeling on one length scale and transfer them as input for subsequent simulations on a different scale. The classical case of sequential models consists in calculating material parameters on a microscopic scale and to use them for simulations on a coarser scale, which is employed to address larger problems. There are several examples of this type in magnetism, where electronic calculations provide input parameters for micromagnetic simulations (Schweflinghaus et al. 2016) or where information is passed from the atomistic scale to the continuum (Kazantseva et al. 2008a; Hrkac et al. 2010). The opposite situation is also possible, where continuum models yield information on the magnetic structure, which is then used as an input for electronic calculations (Seemann et al. 2012). Such sequential multiscale simulations are mentioned here only for completeness. The rest of this chapter will deal with concurrent multiscale simulations.

3   Micromagnetic Continuum Theory

Micromagnetic theory (Brown 1963; Aharoni 2000; Hubert and Schäfer 2012; Kronmüller and Fähnle 2003) plays a central role in the modeling of ferromagnetic nanostructures since it provides a theoretical framework to describe

ferromagnetic solids at mesoscopic length scales. This continuum theory bridges the gap between the macroscopic Maxwell theory of electromagnetic fields and the atomistic treatment of condensed matter based on quantum theory. Instead of using macroscopic effective quantities such as the magnetic permeability or susceptibility, micromagnetic theory describes the properties of a ferromagnet (like, e.g., its response to an external field) by analyzing in detail the internal structure of the magnetization. Micromagnetism is the method of choice to study the properties of magnetic domains, domain walls, vortices, and other fundamental structures of the magnetization which are essential in spintronic devices (Zhu and Zheng 2002).

3.1   Basic Assumptions and Fundamental Equations

The range of validity of micromagnetic theory spans over a broad range of length scales, ranging between a few nanometers and several microns. On the large scale, the validity is limited mostly by practical considerations and by the resources that are available for computer simulations. If the problem size is too large, a micromagnetic treatment is either prohibitive or unnecessary. On the lower scale, the limits of validity can be reached in cases when the spin structure displays important features on the atomic scale that cannot be captured within a continuum model. These situations are rather exceptional in magnetic nanostructures, but they deserve particular attention. In micromagnetic modeling, situations that require a multiscale approach are typically located at the boundaries of the validity of micromagnetic theory. As a continuum theory, micromagnetism traditionally takes the role of the coarse-scale model in multiscale simulations, with atomistic models being used for the fine scale. There are however also opposite situations, where the micromagnetic model represents the fine-scale part of the simulation, while in the macroscopic part, the ferromagnet is described by effective quantities, such as the magnetic susceptibility (Bruckner et al. 2013, 2014; Schrefl et al. 2005).

In micromagnetic theory, the magnetic structure is represented as a smooth vector field of the magnetization $M(r, t)$, which is generally a function of space and time. The magnetization, defined as the density of magnetic moments, is a directional field with constant magnitude. Its magnitude is the spontaneous magnetization $M_s = |M|$, which is a material-specific constant. Besides the constraint of constant magnitude $|M|$, micromagnetic theory relies on the assumption that the magnetic structure of a ferromagnet remains mostly homogeneous on the atomic scale. This assumption is justified qualitatively by the dominant influence of the exchange interaction on short length scales, which tends to avoid strong inhomogeneities, and it is confirmed quantitatively by the value of the exchange length discussed in Sect. 3.2, which is much larger than the interatomic distance. The near homogeneity of the spin structure at the atomic scale allows to introduce the magnetization as a meaningful spatial average of the magnetic moments and thus to use a continuum model to describe the magnetic structure.

The fundamental equation of motion of the magnetization is the Landau-Lifshitz-Gilbert equation (Landau and Lifshitz 1935; Gilbert 2004)

$$\frac{d\boldsymbol{M}}{dt} = -\gamma \boldsymbol{M} \times \boldsymbol{H}_{\text{eff}} + \frac{\alpha}{M_s} \left(\boldsymbol{M} \times \frac{d\boldsymbol{M}}{dt} \right) \qquad (1)$$

where γ is the gyromagnetic ratio, α is a phenomenological damping constant, and $\boldsymbol{H}_{\text{eff}}$ is the effective field acting on the local magnetization. The effective field

$$\boldsymbol{H}_{\text{eff}} = -\frac{1}{\mu_0} \frac{\delta e}{\delta \boldsymbol{M}} \qquad (2)$$

(μ_0 is the vacuum permeability) comprises all micromagnetic energy contributions to the local energy density e, the most important of which are usually the ferromagnetic exchange, the magnetocrystalline anisotropy, the Zeeman term, and the nonlocal magnetostatic interaction energy. A detailed description of the micromagnetic energy terms and the corresponding effective fields entering the Landau-Lifshitz-Gilbert equation can be found in, e.g., Refs. (Brown 1963; Hubert and Schäfer 2012; Aharoni 2000; Kronmüller and Parkin 2012).

Given an initial magnetic structure, the standard task of micromagnetic simulations consists in numerically integrating this equation of motion of the vector field $\boldsymbol{M}$ in time until a converged state is obtained.

3.2   Exchange Lengths: Intrinsic Length Scale

The problem of different interacting length scales is ubiquitous in micromagnetism, even in situations which do not require multiscale modeling techniques. At the nanoscale, the magnetic structure in nanoelements consists of basic topological entities like domain walls or vortices that are often much smaller than the magnet itself. In spite of their small size, these fundamental structures generally have a great impact on the magnetic properties of elements used in spintronic devices, and it is necessary to use sufficiently fine discretization grids in order to ensure that these structures are resolved numerically. Vortices and domain walls have a characteristic extension that results from a balance between the long-range magnetostatic interaction and the short-range exchange interaction. This interplay of energy terms acting on different scales leads to a characteristic length scale, the exchange length,

$$l_s = \sqrt{2A/(\mu_0 M_s^2)} \qquad (3)$$

where A is the exchange constant, a material-specific parameter. The exchange length provides an intrinsic length scale that serves as an estimate of the extension of micromagnetic structures, like domain walls or vortices (Hubert and Schäfer 2012). In all known ferromagnetic materials, the exchange length is larger than at least several nanometers (Kronmüller and Fähnle 2003). This is also true for another type of micromagnetic exchange length that can be defined for magnetic

materials with strong magnetocrystalline anisotropy (Kronmüller 1962). Magnetic inhomogeneities in a ferromagnet therefore extend over a range that is much larger than the atomic lattice constant. Accordingly, atomistic models are unnecessary to simulate ordinary topological structures like vortices or domain walls. There may nevertheless be exceptional situations where the magnetic structure in a ferromagnet develops strong inhomogeneities on atomistic distances. In these cases the continuum model becomes inaccurate and should be replaced by an atomistic model that could be combined with a continuum model in a multiscale simulation. Multiscale micromagnetic simulations with atomistic-continuum coupling are typically employed in cases where pronounced inhomogeneities of the magnetization develop only in certain parts of the domain, while most regions of the nanomagnet can be safely described by means of traditional micromagnetic theory.

Another situation that requires multiscale modeling occurs on the opposite spectrum of length scales, when the magnetic structure needs to be modeled in an environment that is of a macroscopic size, i.e., a domain which is by orders of magnitude larger than the exchange length. This is often the case in the modeling of recording processes, where the micromagnetic model of the magnetic layer on a hard disk has to be embedded into macroscopic simulations of the magnetic field generated by the write head and its macroscopic coil (Schrefl et al. 2005). Sometimes it is also necessary to include thermal effects in the model, a task which may involve a description on different length scales if the heat flow is modeled on a macroscopic scale, while temperature also has an impact on the microscopic spin dynamics. The latter can be described by the stochastic version of the Landau-Lifshitz-Gilbert equation (Lyberatos et al. 1993).

3.3   Connection Between Atomistic and Continuum Theory

In the hierarchy of physical theories of ferromagnetic materials, atomistic models with classical magnetic moments provide the framework with next finer granularity following micromagnetic theory. This more fine-grained approach is not constrained by the assumptions that are used in micromagnetic theory: the direction of magnetic moments at neighboring atomic lattice sites can enclose arbitrarily large angles, and it is straightforward to extend the model to account for situations where the magnetic moment is not constant. The latter feature is implemented in the Landau-Lifshitz-Bloch equation (Chubykalo-Fesenko et al. 2006), which generalizes the Landau-Lifshitz-Gilbert equation by allowing for temporal and spatial variations of the magnitude of the magnetic moment. Accounting for a temporal collapse of the magnetic moment is particularly important in the case of ultrafast, laser-induced magnetization dynamics (Kazantseva et al. 2008b). In spite of qualitative differences, atomistic models share many similarities with micromagnetic theory. In particular, the description of the spin dynamics is governed by an atomistic version of the Landau-Lifshitz-Gilbert equation which has the same form, with effective field terms accounting for the various energy terms. The atomistic and the continuum model differ mainly in the form of two energy representations, the exchange

term and the dipolar (magnetostatic) term, and their respective effective field terms. Local energy terms like the magnetocrystalline anisotropy or the Zeeman term remain unaffected by the transition from a discrete to a continuum model.

The continuum form of the energy terms in micromagnetic theory is formally derived from microscopic principles and is obtained by means of a transition from a discrete representation to a continuum form. This is achieved by a spatial averaging procedure through which an ensemble of magnetic moments and magnetic dipoles is transformed into a continuous magnetization and into a distribution of magnetostatic volume and surface charges (Brown 1962). Atomistic models, on the other hand, are located at an intermediate level between the fundamental electronic theory that provides the basis of the material's magnetic properties and the continuum model which removes all granularity by introducing averaged quantities. They are fine-scaled with respect to the continuum model but large-scaled compared to the electronic level. Due to this intermediate position, some heuristics is used for the calculation of the dipolar field in atomistic models. In atomistic models the magnetostatic field is usually represented by a sum of point dipoles attributed to the lattice sites, while in the continuum model, the field is typically calculated by solving the Poisson equation $\Delta U = \nabla \cdot \boldsymbol{M}$ for the magnetostatic potential U, from which the demagnetizing field is obtained as a gradient field $\boldsymbol{H} = -\nabla U$. By choosing atomistic dipole moments, the atomic model introduces a coarse-grained representation of the sources of the magnetostatic fields which are of electronic origin and could in principle be represented by a continuous spin density on the atomistic scale. However, if the goal is to simulate nontrivial three-dimensional spin structures that extend beyond a size of several nanometers, a self-consistent calculation of the electronic spin density is prohibitive, and dipole sums are a natural and practical solution in view of the granularity of the model. Moreover, since the impact of the magnetostatic term is almost negligible on short distances, a possible loss of accuracy introduced by the point dipole model on the atomic scale can be considered to be insignificant.

The most important difference between the micromagnetic and the atomistic models concerns the exchange term. In the atomistic case, a Heisenberg model replaces the continuum expression. The latter considers gradients in the vector field of the magnetization, while the former contains a sum describing the interaction between neighboring magnetic moments.

$$e_{\text{exc}}^{\text{atomistic}} = -J_{ij} \sum_{i \neq j} \hat{\boldsymbol{\mu}}_i \cdot \hat{\boldsymbol{\mu}}_j \quad . \tag{4}$$

Here $\hat{\mu}_i$ is the normalized magnetic moment at the lattice site i, the sum extends over nearest neighbor sites, and J_{ij} is the Heisenberg exchange constant, while the continuum representation is

$$e_{\text{exc}}^{\text{continuum}} = \frac{A}{M_s^2} \left[\left(\nabla M_x\right)^2 + \left(\nabla M_y\right)^2 + \left(\nabla M_z\right)^2 \right] \quad . \tag{5}$$

These two energy forms are connected since the continuum form (5) can be derived from the atomistic Heisenberg term (4) by means of a Taylor expansion by assuming small angles between neighboring magnetic moments (Kittel 1949), i.e.,

$$\left| \hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j \right| \ll 1 \quad . \tag{6}$$

The validity of the continuum model is challenged whenever this assumption does not hold. When particularly strong magnetic inhomogeneities develop, the continuum term should be replaced by the atomistic term.

3.4   Strongly Inhomogeneous Spin Structures

The micromagnetic exchange lengths cease to be suitable a measure when the assumption of near homogeneity of the spin structure on the atomic scale is not fulfilled. This can occur, e.g., in heterogeneous, nanocrystalline magnets, where the spin structure can be strongly inhomogeneous at the grain boundaries. Moreover, the exchange lengths are derived assuming an energetic balance resulting from a static energy minimization. Therefore, the scale described by the exchange length is not meaningful in systems that are far out of equilibrium, like in the case of laser-induced magnetic switching (Kazantseva et al. 2008b). A further case in which the magnetization becomes inhomogeneous on length scales that are much smaller than the exchange length are Bloch points, a special type of three-dimensional topological structures (Döring 1968; Feldtkeller 1965). Bloch points can form in homogeneous magnetic material and without thermal excitation. Whenever the criterion (6) is not fulfilled and the exchange lengths lose their meaning, the continuum term of the exchange energy (5) cannot be applied reliably. In such situations the fundamental assumptions of micromagnetic theory are violated, and it becomes necessary to treat the magnetic structure with a more fine-scaled, atomistic theory.

Having established that micromagnetic continuum theory is not an accurate framework to describe microscopic spin structures with inhomogeneities that are too strong to be negligible on an atomistic scale, a quantitative measure is needed to determine the degree of inhomogeneity up to which it is acceptable to apply the continuum theory. Such a measure is an important metric for dynamic multiscale simulations as it allows to identify regions which contain spin structures that require an atomistic treatment. To address this question, Andreas et al. have quantified (Andreas et al. 2014a) the degree of magnetic inhomogeneity beyond which the continuum form of the exchange energy should be replaced by a Heisenberg term. Using a one-dimensional spin spiral as a test system, the local error due to the continuum approximation was calculated as a function of the strength of the magnetic inhomogeneity. This was achieved by comparing the exchange energy computed with the continuum expression with the value obtained from the Heisenberg term. It was found that the local error remains below 0.1% if the spins rotate by 360° within a distance of at least 15 unit cells (Fig. 3).
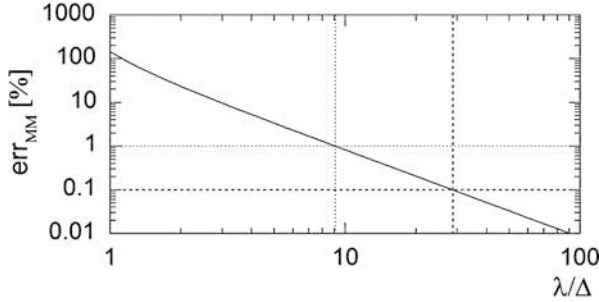
Fig. 3 Relative error of the micromagnetic exchange energy with respect to the Heisenberg term in the case of an ideal spin spiral. The length λ of the spiral is expressed in units of the distance Δ between neighboring atomic spins along the direction of the spiral. Reprinted with permission from (Andreas et al. 2014b) Copyright 2014 by the American Physical Society

4   Bloch Point Structures

Over the past years, research in spintronics and nanomagnetism has mostly focused on the properties of flat magnetic thin film elements. Owing to advances in synthesis methods, recent research in nanomagnetism has expanded into three dimensions (Fernández-Pacheco et al. 2017). This development has been further encouraged by theoretical works predicting new magnetic effects in three-dimensional nano-magnets (Yan et al. 2012, 2011) which could result in new functionalities (Staňo et al. 2018; Sanz-Hernández et al. 2017). A transition toward three-dimensional nanomagnets is accompanied by qualitative changes in the magnetic structure. One type of fundamental spin structures developing specifically in three-dimensional nanomagnetism are Bloch points. These complex spin structures display a number of unusual properties which make it impossible to simulate them accurately with usual micromagnetic algorithms. They rather represent archetypal examples of spin structures that can only be modeled reliably by means of multiscale modeling techniques.

4.1   Characteristic Properties and Importance of Bloch Points

Contrary to what their name suggests, Bloch points are not point-like entities but genuinely three-dimensional spin structures (Feldtkeller 1965; Döring 1968). Bloch point structures have a unique combination of topological, structural, and magnetostatic properties which distinguishes them from usual fundamental micro-magnetic structures. The terminology originates from the fact that they contain a point singularity in the continuum picture. Although this point can be regarded as a zero-dimensional entity, its position cannot be defined without the characteristic three-dimensional magnetic environment that surrounds it. There is a large family
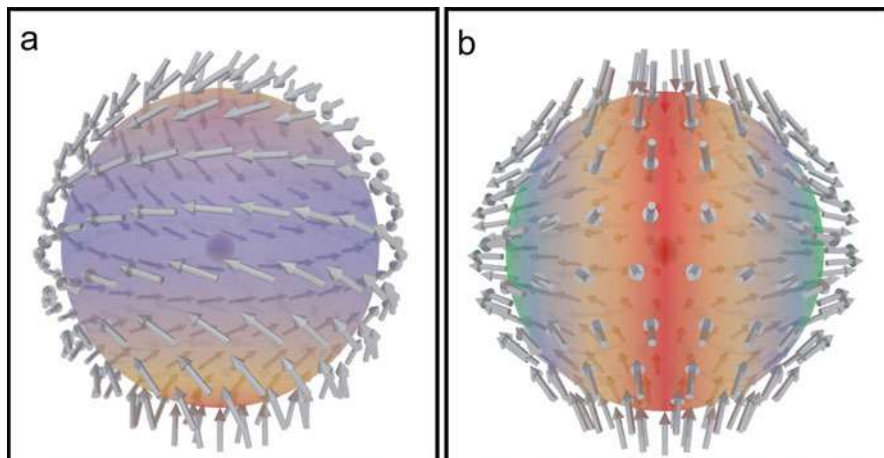
Fig. 4 Two different versions of Bloch point structures: (**a**) vortex and (**b**) antivortex type. There exist several further categories (Malozemoff and Slonczewski 1979). Reprinted from (Hertel 2016), Copyright 2016, with permission from IOP publishing

of possible Bloch point structures, which are discussed in detail in the book by Malozemoff and Slonczewski (1979). Bloch point structures are characterized by a circulation of the magnetization around a central point, such that the entire spin space is covered (Fig. 4). Insofar, they can be considered as three-dimensional variants of magnetic vortices or antivortices. One of their defining micromagnetic properties is that any magnetization direction can be found at least once on the surface of any sphere that contains a Bloch point (Hubert and Schäfer 2012). The structurally simplest version of a Bloch point is a hedgehog structure, where the magnetization points radially away from a central point. In the continuum model, the exchange energy density of such a structure diverges at the central point, although the total exchange energy remains finite. The core of the Bloch point represents a topological defect at which the magnetization direction is undefined. In the continuum model, the singularity describes a region at which the magnetization is maximally inhomogeneous, i.e., where the magnetization direction changes by 180° on an arbitrarily small distance. The occurrence of such divergences and singularities is obviously not a physical reality, but rather indicates that the approximations of the continuum model cease to be valid. In fact, there are no singularities when Bloch point structures are treated within the framework of discrete, atomistic models.

Concerning their magnetostatic properties, Bloch points are magnetic monopoles of the magnetic field $\boldsymbol{H}$. Note that, since Bloch points are not isolated entities, their monopolar magnetostatic charge is a real physical feature which does not represent a violation of fundamental laws of physics ($\nabla \cdot \boldsymbol{B} = 0$ remains valid).

In addition to their magnetostatic charge, Bloch points carry a topological charge Q, which is similar to that of skyrmions. In a two-dimensional spin skyrmion

structure, the topological charge is determined by means of a projection of the planar magnetic structure onto a unit sphere, followed by an integration over the surface of the sphere (Pfleiderer 2011). The magnetic structure of a skyrmion projected onto a unit sphere has the same topological property as the magnetization on a unit sphere that encloses a Bloch point (without performing any projection). Therefore, like in the case of skyrmions, a nonzero, integer topological charge can be assigned to Bloch points.

In spite of their topological similarity with skyrmions, Bloch points may also be regarded as three-dimensional counterparts of magnetic vortex and antivortex structures in thin films. This is because of the role that they can play in the dynamics of three-dimensional magnetic structures, which is in many ways analogous to that of vortices in two-dimensional magnetic thin films. For example, magnetic switching processes in thin film elements are often mediated by the nucleation, propagation, and expulsion of magnetic vortices. Similarly, three-dimensional reversal processes can be accomplished by Bloch points entering, traversing, and exiting the sample. A classical example thereof is the magnetization reversal of a soft magnetic cylindrical wire (Arrott et al. 1979). Another similarity between vortices and Bloch points is that they do not occur below a critical sample size, which in the case of a vortex is known as the single-domain limit; otherwise these spin structures would dissolve in order to reduce the ferromagnetic exchange energy. In ferromagnets, these structures can only exist when they are embedded into a sufficiently large volume, so that the cumulative impact of the long-range magnetostatic interaction is sufficiently strong to stabilize them.

4.2   Micromagnetic Modeling of Bloch Points

Accurately simulating Bloch point structures and their dynamics represents a major challenge for micromagnetic simulations, since two characteristic features of Bloch points violate the basic assumptions of micromagnetic theory. Firstly, the topological defect at the center of the Bloch point is in contradiction to the assumption of a smooth and continuous vector field of the magnetization, and it is also in disagreement with the constraint $|\boldsymbol{M}| = $ const. More importantly, the assumption of small variations of the magnetization on the atomic scale is not valid in the vicinity of a Bloch point. In numerical simulations, this results in important local errors of the exchange energy density and of the corresponding effective field, as well as in an artificial pinning of the Bloch point at the discretization grid.

First attempts to address these problems within the framework of micromagnetic theory consisted in using adaptively refined meshes with a high discretization density near the Bloch point (Thiaville et al. 2003; Hertel and Kronmüller 2002). A more sophisticated approach was later presented which consists in extending the continuum model by adding a degree of freedom that, similar to the LLB equation, relaxes the constraint of constant magnitude of the magnetization (Elías and Verga 2011; Galkina et al. 1993; Lebecki et al. 2012). In these models, the singularity in the core is removed by gradually reducing $|\boldsymbol{M}|$ to zero as the center of the Bloch

point is approached. Although the validity of this phenomenological approach to reduce the magnitude of the magnetization near the Bloch point center is not firmly established, these models succeed in removing the difficulties connected with the micromagnetic simulation of Bloch points and thereby provide a description that is more reliable than ordinary micromagnetic simulations with local mesh refinement.

Simulating Bloch points in ferromagnets is generally a task of intrinsically multiscale nature. While standard micromagnetic theory fails to describe the features in the vicinity of the core region, purely atomistic models, on the other hand, cannot describe the formation and motion of Bloch points. Unless specific boundary conditions of the magnetic structure are enforced (Reinhardt 1973), a Bloch point that is not embedded into a larger micromagnetic continuum dissolves in atomistic models due to its high energy density. Bloch points are usually stabilized by the cumulative effect of the long-range magnetostatic interaction within a volume that is too large to be treated atomistically. Exceptions to this are more exotic materials with a strong antisymmetric exchange, where the short-range Dzyaloshinskii-Moriya interaction can stabilize Bloch points (Milde et al. 2013; Rybakov et al. 2015). In these situations purely atomistic models can be sufficient.

Recently, fully dynamical multiscale micromagnetic simulations with atomistic-continuum coupling have been employed to study Bloch points in an ordinary ferromagnet. This has provided quite detailed insight into the dynamics of Bloch points and their interaction with the atomic lattice structure.

4.3 Dynamic Multiscale Model with Atomistic-Continuum Coupling

Since the core region of a Bloch point structure cannot be described reliably with continuum micromagnetism, computer simulations of these magnetic structures should aim at applying an atomistic model in this part of the structure. The magnetic structure surrounding the central region can be simulated safely with continuum theory, and it would thus be unnecessarily expensive (and in practice not feasible) to simulate the entire three-dimensional Bloch point structure atomistically. Purely atomistic simulations could be used to study artificial situations, where the Bloch point structure is stabilized by imposing particular boundary conditions of the magnetization (Reinhardt 1973), but it would currently not be possible to perform fully atomistic simulations of realistic cases of Bloch point structures propagating in mesoscopic ferromagnets. Arguably the best strategy to address these situations is to apply the atomistic model only selectively, i.e., in the regions of the sample where an atomistic resolution is needed, and to enclose this atomistic part into a continuum model that describes the overall, mesoscopic magnetic structure. With such a multiscale approach, it is possible to achieve the required atomistic resolution in Bloch point cores while keeping the total number of degrees of freedom in the simulation within acceptable limits.

If the goal of micromagnetic simulation studies is to investigate the dynamics of Bloch points, the position of the region that is treated atomistically cannot be
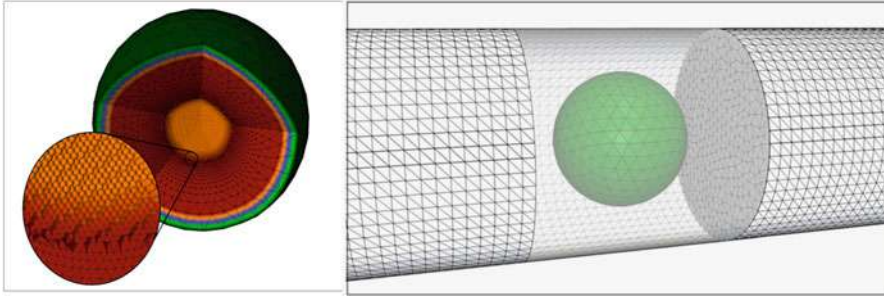
Fig. 5 Exploded view of the multiscale sphere used in Andreas et al. (2014b) The core contains an atomistic region, which is connected to a fully refined finite element mesh with cell sizes that gradually increase toward the surface. Right: The multiscale sphere can move within the volume of the ferromagnetic sample, in this case a thick magnetic nanowire. Reprinted from (Hertel 2016), Copyright 2016, with permission from IOP Publishing

defined a priori. The problem requires a fully dynamic multiscale simulation with adaptively moving atomistic-continuum interface. The first report of a code with these characteristics is due to Andreas et al. (2014b). The authors used GPU-accelerated multiscale modeling techniques to perform dynamic simulations of the magnetic structure and succeeded in bridging several length scales. By adhering to an object-oriented programming approach, the authors encapsulate the multiscale part of the simulation within a numerical object that they called the multiscale sphere (Fig. 5). The multiscale sphere has a diameter of about 50 nm and consists of a core-shell type discretization. It contains a core region that is fully atomistic, enclosed into a spherical shell which is entirely governed by continuum micromagnetics on its outer surface. The spherical shape and the discretization within this multiscale object are optimized for the simulation of Bloch points. To design the multiscale sphere, the size of the region that requires atomistic accuracy in a static Bloch point structure was determined by using the function displayed in Fig. 3 as an error indicator, and the volume of the atomistic core region of the multiscale sphere was chosen accordingly. The transition between this atomistic central region and the surrounding continuum model occurs entirely within the multiscale sphere. By enclosing the atomistic-continuum coupling in a numerical object that can be dynamically generated, displaced, and removed, it becomes possible to adaptively apply multiscale modeling techniques where it is needed. In the simulations, this feature is used to track the position of a Bloch point during its field-driven motion. This ensures that the core of the structure is always simulated with the atomistic model. The code first identifies the positions at which the inhomogeneity is so strong that an atomistic treatment is needed. It then generates or displaces the multiscale sphere as necessary and removes it when it is no longer required. The multiscale sphere can thus be thought of as an "atomistic lens" that follows the region of interest during the simulation. Inserting and displacing the multiscale sphere inside the volume of the sample is achieved without any time-consuming remeshing of

the continuum model in the background. Moreover, several multiscale spheres can operate simultaneously (we used up to 20), with their number being limited by the available numerical resources. When different multimodel spheres intersect, they can generate a larger contiguous atomistic region.

Within the multiscale sphere, the coupling between the atomistic region in the core sphere and the continuum in the outer shell is obtained by means of a hybrid handshake region. In this interface region, both atomistic and continuum model are used. The contributions of the two models are weighted according to their position within the shell, such that the inner part is fully atomistic and the outer part fully micromagnetic. This coupling, combined with a gradual refinement of the mesh down to the atomistic scale near the core region, ensures a seamless transition between the two models. In order to couple the multiscale sphere to the surrounding "background" continuum that provides the larger magnetic structure, the sphere receives the magnetic structure from the background at the surface and transfers the magnetic structure back to the continuum model via the nodes of the unrefined background mesh that are located within the volume of the sphere. More details on the model are described in Andreas et al. (2014b).

4.4   Simulated Bloch Point Dynamics, Interaction with Atomic Lattice

The dynamic multiscale modeling technique described in the previous section has made it possible to simulate in detail the field-driven motion of a Bloch point in a cylindrical nanowire, thereby providing insight into a fundamental magnetization process that was predicted about 40 years ago by Arrott et al. (1979). In cylindrical geometries, Bloch points are part of a special type of head-to-head domain walls. These domain walls, sometimes called Bloch point walls (Da Col et al. 2014), are characterized by an azimuthal vortex structure with the magnetization circulating around the central axis of the cylinder, and they separate two domains with opposite magnetization direction aligned along the cylinder axis. The Bloch point is located in the center of the vortex domain wall and is dragged along the wire as the domain wall is displaced. The simulations revealed that Bloch points driven by moderate magnetic fields (less than 100 mT) can propagate at very high velocities (above 1000 ms^{-1}) without hindering the motion of the domain wall that contains them (Hertel 2016). In spite of its vanishing net magnetization in the core region, the Bloch point structure does not appear to have a significantly inferior mobility than the domain wall. Moreover, the simulations predict that, under specific circumstances, the motion of a field-driven Bloch point in a magnetic nanocylinder can change from a linear displacement along the cylinder axis to a complex spiraling orbit. As a Fourier analysis shows, this spiraling motion can be decomposed into two circulations with opposite sign and different amplitudes. This particular gyrating motion is probably due to the impact of the topological charge of the dynamics, as described by the Thiele equation (Thiele 1973). This is corroborated by the fact that a similarly complex motion was also reported for skyrmions, which have the same

topological charge as Bloch points. In the case of skyrmions, numerical studies have shown that the topological charge leads to a motion along a hypocycloid trajectory (Moutafis et al. 2009; Moon et al. 2014; Shiino et al. 2017), which is very similar to the complex, spiraling motion that occurs for the Bloch point.

In addition to this complex Bloch point dynamics, the studies have brought into evidence an interaction between the atomic lattice and the Bloch point. In the low-field dynamics, the relative orientation of the atomic crystal with respect to the propagation direction can significantly affect the magnetic structure and the propagation velocity of Bloch points. Note that the simulation was performed assuming vanishing magnetocrystalline anisotropy. Therefore, any interaction between the lattice and the Bloch point obtained in the simulation is not of relativistic nature, but a direct consequence of the spatial arrangement of the atomic magnetic moments in the lattice. The interaction between the crystal lattice and the Bloch point is also responsible for the occurrence of a depinning field, i.e., a minimal field strength which is necessary to initiate the motion of a Bloch point. Such a pinning of Bloch points at preferential sites within the atomic lattice was predicted already in 1973 (Reinhardt 1973) and is largely confirmed by the multiscale simulations, which moreover show that the precise value of the depinning field depends on the orientation of the crystal lattice.

Owing to the atomistic resolution provided by the multiscale simulations, it is possible to identify the position of the Bloch point within the atomic lattice, in the static case, as well as in dynamic processes. At any time step of the simulation, the position can be determined by spatially interpolating the magnetic moments of the atomic lattice, since the location of the Bloch point can be uniquely defined as the point at which the interpolated magnetic moment is equal to zero. Using this definition, one can reconstruct precisely the trajectory of the Bloch point during its motion through the crystal lattice. If the instantaneous position of the Bloch point, taken at many different time steps, is projected into a unit cell, a pattern develops which displays a discrete version of the spatial probability distribution of the Bloch point in a unit cell (Fig. 6). The set of points that is thereby obtained assembles in the form of the Wigner-Seitz cell of the crystal lattice of the magnet, thereby showing that the Bloch point is located on a midplane between two atomic lattice sites. Further inspection shows that the probability density distribution is not uniform on the Wigner-Seitz cell. Preferential facets and edges can be clearly recognized, as well as high-energy facets that the Bloch point tends to avoid during its motion.

The interaction between the Bloch point and the crystal lattice is also evidenced by adding information on the local velocity to the high-resolution data of the Bloch point position. A useful representation of the data describing the trajectory of a Bloch point moving through the crystal lattice is obtained by following the previously described principle of projecting the current position into a unit cell. This allows to correlate the local Bloch point velocity and its position within the unit cell. The result emphasizes the strong tendency of the Bloch point to avoid high-energy facets: during its motion, the Bloch point may occasionally enter a high-energy facet of the Wigner-Seitz cell. When this occurs, the Bloch point tends to leave this unfavorable region very quickly, as it immediately experiences a drastic acceleration that increases the local velocity temporarily by almost a factor of two.
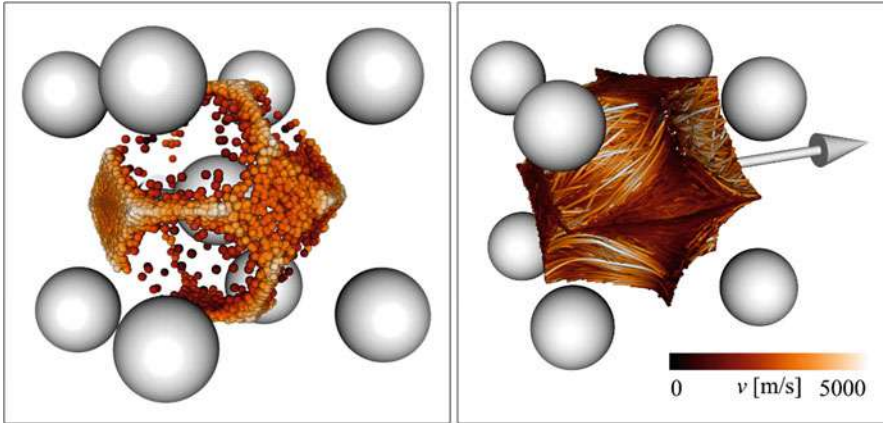
Fig. 6 Left: Ensemble of the Bloch positions (displayed in orange) in an atomistic unit cell taken at different time steps during the field-driven propagation of a Bloch point. Preferred locations of the Bloch point are the edges and the square facets of the Wigner-Seitz cell. Right: Trajectory of a Bloch point moving through the crystal lattice, projected into a single unit cell. The color code displays the local velocity. Strong accelerations occur when the Bloch point enters an energetically unfavorable facet of the Wigner-Seitz cell

It should be kept in mind that, although it is possible to ascribe a position and a local velocity to a Bloch point, the analysis of its dynamics is not comparable to the motion of a particle through the crystal lattice. The position of the Bloch point is merely a point in space that is geometrically constructed according to the surrounding spin structure. It is not an isolated entity, but represents the center of a large, three-dimensional magnetic structure. Remarkably, the relatively large micromagnetic Bloch point structure is sensitive to the fine structure of the atomic lattice not only as far as the pinning in the static case is concerned but also during its high-velocity motion through the crystal.

5   Conclusions

Multiscale modeling techniques are tailored to bridge length scales that differ by several orders of magnitude. This is achieved by simulating different parts of the computational region with different degrees of accuracy. In most cases only two length scales are involved, a fine scale and a coarse scale, which may be represented by different physical models. Carefully connecting the coarse-scale and the fine-scale model makes it possible to perform large-scale simulations with high accuracy. In micro- and nanomagnetism, dynamic multiscale simulations have recently been applied to combine atomistic and continuum models. Algorithms with such capabilities could be employed to simulate the spin dynamics in future spintronic devices. These features are particularly useful in view of the ongoing technological transition toward three dimensions in nanomagnetism, where Bloch points often play a central role in magnetic switching processes. Due to their

inherent multiscale character, Bloch point structures cannot be treated reliably with traditional micromagnetic simulations. While multiscale simulations with dynamic atomistic-continuum modeling have long been applied in other domains of material modeling, they have been developed only recently in magnetic modeling. The success of these multiscale simulations in providing insight into the dynamics of Bloch points and their interaction with the crystal lattice makes them a promising starting point for future investigations of complex, three-dimensional magnetic structures.

References

Aharoni A (2000) Introduction to the theory of ferromagnetism, 2nd edn. Oxford University Press, Oxford

Andreas C, Gliga S, Hertel R (2014a) Numerical micromagnetism of strong inhomogeneities. J Magn Magn Mater 362:7–13. https://doi.org/10.1016/j.jmmm.2014.02.097, http://www.sciencedirect.com/science/article/pii/S0304885314002212

Andreas C, Kákay A, Hertel R (2014b) Multiscale and multimodel simulation of Bloch-point dynamics. Phys Rev B 89(13):134403. http://link.aps.org/doi/10.1103/PhysRevB.89.134403

Arrott A, Heinrich B, Aharoni A (1979) Point singularities and magnetization reversal in ideally soft ferromagnetic cylinders. IEEE Trans Magn 15(5):1228–1235. https://doi.org/10.1109/TMAG.1979.1060342

Börm S, Grasedyck L (2004) Low-rank approximation of integral operators by interpolation. Computing 72(3-4):325–332. https://link.springer.com/article/10.1007/s00607-003-0036-0

Brown WF (1962) Magnetostatic principles in ferromagnetism. North-Holland, Amsterdam

Brown WF (1963) Micromagnetics. Interscience Publishers, Huntington

Bruckner F, Vogler C, Bergmair B, Huber T, Fuger M, Suess D, Feischl M, Fuehrer T, Page M, Praetorius D (2013) Combining micromagnetism and magnetostatic Maxwell equations for multiscale magnetic simulations. J Magn Magn Mater 343(Supplement C):163–168. https://doi.org/10.1016/j.jmmm.2013.04.085, http://www.sciencedirect.com/science/article/pii/S0304885313003156

Bruckner F, Suess D, Feischl M, Führer T, Goldenits P, Page M, Praetorius D, Ruggeri M (2014) Multiscale modeling in micromagnetics: existence of solutions and numerical integration. Math Models Methods Appl Sci 24(13):2627–2662. http://www.worldscientific.com/doi/abs/10.1142/S0218202514500328

Chubykalo-Fesenko O, Nowak U, Chantrell RW, Garanin D (2006) Dynamic approach for micromagnetics close to the Curie temperature. Phys Rev B 74(9):094436. https://link.aps.org/doi/10.1103/PhysRevB.74.094436

Da Col S, Jamet S, Rougemaille N, Locatelli A, Mentes TO, Burgos BS, Afid R, Darques M, Cagnon L, Toussaint JC, Fruchart O (2014) Observation of Bloch-point domain walls in cylindrical magnetic nanowires. Phys Rev B 89(18):180405. http://link.aps.org/doi/10.1103/PhysRevB.89.180405

De Lucia A, Krüger B, Tretiakov O, Kläui M (2016) Multiscale model approach for magnetization dynamics simulations. Phys Rev B 94(18):184415. https://journals.aps.org/prb/abstract/10.1103/PhysRevB.94.184415

De Lucia A, Litzius K, Krüger B, Tretiakov OA, Kläui M (2017) Multiscale simulations of topological transformations in magnetic-skyrmion spin structures. Phys Rev B 96(2):020405. https://link.aps.org/doi/10.1103/PhysRevB.96.020405

Donahue M, McMichael R (1997) Exchange energy representations in computational micromagnetics. Phys B Condens Matter 233(4):272–278. http://www.sciencedirect.com/science/article/pii/S0921452697003104

Döring W (1968) Point singularities in micromagnetism. J Appl Phys 39(2):1006–1007. https://doi.org/10.1063/1.1656144, http://jap.aip.org/resource/1/japiau/v39/i2/p1006_s1

Elías RG, Verga A (2011) Magnetization structure of a Bloch point singularity. Eur Phys J B 82(2):159–166. https://doi.org/10.1140/epjb/e2011-20146-6, http://www.springerlink.com/content/3730q17475492813/

Feldtkeller (1965) Mikromagnetisch stetige und unstetige Magnetisierungsstrukturen. Zeitschrift für angewandte Physik 19(6):530–536

Fernández-Pacheco A, Streubel R, Fruchart O, Hertel R, Fischer P, Cowburn RP (2017) Three-dimensional nanomagnetism. Nat Commun 8:15756. https://www.nature.com/articles/ncomms15756

Forster H, Schrefl T, Suess D, Scholz W, Tsiantos V, Dittrich R, Fidler J (2002) Domain wall motion in nanowires using moving grids (invited). J Appl Phys 91:6914. https://doi.org/10.1063/1.1452189, http://link.aip.org/link/JAPIAU/v91/i10/p6914/s1&Agg=doi

Galkina E, Ivanov B, Stephanovich V (1993) Phenomenological theory of Bloch point relaxation. J Magn Magn Mater 118(3):373–378. https://doi.org/10.1016/0304-8853(93)90441-4, http://www.sciencedirect.com/science/article/pii/0304885393904414

Garcia-Cervera C, Roma A (2006) Adaptive mesh refinement for micromagnetics simulations. IEEE Trans Magn 42(6):1648–1654. https://doi.org/10.1109/TMAG.2006.872199

Garcia-Sanchez F, Chubykalo-Fesenko O, Mryasov O, Chantrell RW, Guslienko KY (2005) Exchange spring structures and coercivity reduction in FePt/FeRh bilayers: a comparison of multiscale and micromagnetic calculations. Appl Phys Lett 87(12):122501. http://aip.scitation.org/doi/full/10.1063/1.2051789

Gilbert TL (2004) A phenomenological theory of damping in ferromagnetic materials. IEEE Trans Magn 40(6):3443–3449. https://doi.org/10.1109/TMAG.2004.836740

Gliga S, Kákay A, Hertel R, Heinonen OG (2013) Spectral analysis of topological defects in an artificial spin-ice lattice. Phys Rev Lett 110(11):117205. http://link.aps.org/doi/10.1103/PhysRevLett.110.117205

Hackbusch W (2015) Hierarchical matrices: algorithms and analysis, 1st edn. Springer, New York

Hertel R (2016) Ultrafast domain wall dynamics in magnetic nanotubes and nanowires. J Phys Condens Matter 28(48):483002. http://stacks.iop.org/0953-8984/28/i=48/a=483002

Hertel R, Kronmüller H (1998) Adaptive finite element mesh refinement techniques in three-dimensional micromagnetic modeling. IEEE Trans Magn 34(6):3922–3930

Hertel R, Kronmüller H (2002) Finite element calculations on the single-domain limit of a ferromagnetic cubea solution to μMAG standard problem No. 3. J Magn Magn Mater 238(2–3):185–199. https://doi.org/10.1016/S0304-8853(01)00876-9, http://www.sciencedirect.com/science/article/pii/S0304885301008769

Hertel R, Fruchart O, Cherifi S, Jubert PO, Heun S, Locatelli A, Kirschner J (2005) Three-dimensional magnetic-flux-closure patterns in mesoscopic Fe islands. Phys Rev B 72:214409. http://link.aps.org/doi/10.1103/PhysRevB.72.214409

Hrkac G, Woodcock TG, Freeman C, Goncharov A, Dean J, Schrefl T, Gutfleisch O (2010) The role of local anisotropy profiles at grain boundaries on the coercivity of $Nd_2Fe_{14}B$ magnets. Appl Phys Lett 97(23):232511. http://scitation.aip.org/content/aip/journal/apl/97/23/10.1063/1.3519906

Hubert A, Schäfer R (2012) Magnetic domains – the analysis of magnetic microstructures, 1st edn. Springer. http://www.springer.com/materials/book/978-3-540-64108-7

Jourdan T, Marty A, Lançon F (2008) Multiscale method for Heisenberg spin simulations. Phys Rev B 77(22):224428. http://link.aps.org/doi/10.1103/PhysRevB.77.224428

Kakay A, Westphal E, Hertel R (2010) Speedup of FEM micromagnetic simulations with graphical processing units. IEEE Trans Magn 46:2303–2306. https://doi.org/10.1109/TMAG.2010.2048016, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5467640

Kazantseva N, Hinzke D, Nowak U, Chantrell RW, Atxitia U, Chubykalo-Fesenko O (2008a) Towards multiscale modeling of magnetic materials: simulations of FePt. Phys Rev B 77(18):184428. http://link.aps.org/doi/10.1103/PhysRevB.77.184428

Kazantseva N, Nowak U, Chantrell RW, Hohlfeld J, Rebei A (2008b) Slow recovery of the magnetisation after a sub-picosecond heat pulse. EPL (Europhys Lett) 81(2):27004. https://doi.org/10.1209/0295-5075/81/27004, http://stacks.iop.org/0295-5075/81/i=2/a=27004

Kittel C (1949) Physical theory of ferromagnetic domains. Rev Mod Phys 21(4):541–583. http://link.aps.org/doi/10.1103/RevModPhys.21.541

Knittel A, Franchin M, Bordignon G, Fischbacher T, Bending S, Fangohr H (2009) Compression of boundary element matrix in micromagnetic simulations. J Appl Phys 105(7):07D542. https://doi.org/10.1063/1.3072032, http://link.aip.org/link/JAPIAU/v105/i7/p07D542/s1&Agg=doi

Kohlhoff S, Gumbsch P, Fischmeister HF (1991) Crack propagation in b.c.c. crystals studied with a combined finite-element and atomistic model. Philos Mag A 64(4):851–878. https://doi.org/10.1080/01418619108213953

Koopmans B, Malinowski G, Longa FD, Steiauf D, Fhnle M, Roth T, Cinchetti M, Aeschlimann M (2010) Explaining the paradoxical diversity of ultrafast laser-induced demagnetization. Nat Mater 9(3):259–265. https://doi.org/10.1038/nmat2593, https://www.nature.com/articles/nmat2593

Kronmüller H (1962) Mikromagnetische Berechnung der Magnetisierung in der Umgebung unmagnetischer Einschlüsse in Ferromagnetika. Zeitschrift für Physik 168(5):478–494. https://doi.org/10.1007/BF01378144, http://www.springerlink.com/content/w44w23jh359w8563/

Kronmüller H, Fähnle M (2003) Micromagnetism and the microstructure of ferromagnetic solids. Cambridge University Press, Cambridge

Kronmüller H, Fischer R, Hertel R, Leineweber T (1997) Micromagnetism and the microstructure in nanocrystalline materials. J Magn Magn Mater 175(1–2):177–192. https://doi.org/10.1016/S0304-8853(97)00225-4, http://www.sciencedirect.com/science/article/pii/S0304885397002254

Kronmüller H, Parkin SSP (2012) Handbook of magnetism and advanced magnetic materials. Wiley, Chichester

Landau LD, Lifshitz E (1935) On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. Phys Z Sowjet 8:153–169. http://cds.cern.ch/record/437299

Lebecki KM, Hinzke D, Nowak U, Chubykalo-Fesenko O (2012) Key role of temperature in ferromagnetic Bloch point simulations. Phys Rev B 86(9):094409. http://link.aps.org/doi/10.1103/PhysRevB.86.094409

Li S, Livshitz B, Lomakin V (2010) Graphics processing unit accelerated micromagnetic solver. IEEE Trans Magn 46(6):2373–2375. https://doi.org/10.1109/TMAG.2010.2043504

Lopez-Diaz L, Aurelio D, Torres L, Martinez E, Hernandez-Lopez MA, Gomez J, Alejos O, Carpentieri M, G Finocchio, Consolo G (2012) Micromagnetic simulations using graphics processing units. J Phys D Appl Phys 45(32):323001. https://doi.org/10.1088/0022-3727/45/32/323001, http://stacks.iop.org/0022-3727/45/i=32/a=323001

Lyberatos A, Berkov DV, Chantrell RW (1993) A method for the numerical simulation of the thermal magnetization fluctuations in micromagnetics. J Phys Condens Matter 5(47):8911–8920. https://doi.org/10.1088/0953-8984/5/47/016, http://iopscience.iop.org/0953-8984/5/47/016

Malozemoff AP, Slonczewski JC (1979) Magnetic domain walls in bubble materials. Academic, New York

Milde P, Köhler D, Seidel J, Eng LM, Bauer A, Chacon A, Kindervater J, Mühlbauer S, Pfleiderer C, Buhrandt S, Schütte C, Rosch A (2013) Unwinding of a skyrmion lattice by magnetic monopoles. Science 340(6136):1076–1080. https://doi.org/10.1126/science.1234657, http://science.sciencemag.org/content/340/6136/1076

Miller RE, Tadmor EB (2009) A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. Model Simul Mater Sci Eng 17(5):053001. https://doi.org/10.1088/0965-0393/17/5/053001, http://m.iopscience.iop.org/0965-0393/17/5/053001

Moon KW, Chun BS, Kim W, Qiu ZQ, Hwang C (2014) Control of skyrmion magnetic bubble gyration. Phys Rev B 89(6):064413. https://link.aps.org/doi/10.1103/PhysRevB.89.064413

Moutafis C, Komineas S, Bland JAC (2009) Dynamics and switching processes for magnetic bubbles in nanoelements. Phys Rev B 79(22):224429. https://link.aps.org/doi/10.1103/PhysRevB.79.224429

Pechstein C (2013) Finite and boundary element tearing and interconnecting solvers for multiscale problems. Lecture notes in computational science and engineering. Springer, Berlin/Heidelberg. https://www.springer.com/us/book/9783642235870

Pfleiderer C (2011) Magnetic order: surfaces get hairy. Nat Phys 7(9):673. https://doi.org/10.1038/nphys2081, https://www.nature.com/articles/nphys2081

Poluektov M, Eriksson O, Kreiss G (2018) Coupling atomistic and continuum modelling of magnetism. Comput Methods Appl Mech Eng 329:219–253. https://doi.org/10.1016/j.cma.2017.10.010, https://www.sciencedirect.com/science/article/pii/S0045782517302463

Popović N, Praetorius D (2005) Applications of H-matrix techniques in micromagnetics. Computing 74(3):177–204. https://link.springer.com/article/10.1007/s00607-004-0098-7

Rave W, Hubert A (2000) Magnetic ground state of a thin-film element. IEEE Trans Magn 36(6):3886–3899. https://doi.org/10.1109/20.914337

Reinhardt J (1973) Gittertheoretische Behandlung von mikromagnetischen Singularitäten. Int J Magnetism 5:263–268

Rybakov FN, Borisov AB, Blügel S, Kiselev NS (2015) New type of stable particlelike states in chiral magnets. Phys Rev Lett 115(11):117201. https://link.aps.org/doi/10.1103/PhysRevLett.115.117201

Sanz-Hernández D, Hamans RF, Liao JW, Welbourne A, Lavrijsen R, Fernández-Pacheco A (2017) Fabrication, detection, and operation of a three-dimensional nanomagnetic conduit. ACS Nano 11(11):11066–11073. https://doi.org/10.1021/acsnano.7b05105

Schrefl T, Schabes M, Suess D, Ertl O, Kirschner M, Dorfbauer F, Hrkac G, Fidler J (2005) Partitioning of the perpendicular write field into head and SUL contributions. IEEE Trans Magn 41(10):3064–3066. https://doi.org/10.1109/TMAG.2005.855227

Schweflinghaus B, Zimmermann B, Heide M, Bihlmayer G, Blgel S (2016) Role of Dzyaloshinskii-Moriya interaction for magnetism in transition-metal chains at Pt step edges. Phys Rev B 94(2):024403. https://link.aps.org/doi/10.1103/PhysRevB.94.024403

Seemann KM, Garcia-Sanchez F, Kronast F, Miguel J, Kákay A, Schneider CM, Hertel R, Freimuth F, Mokrousov Y, Blügel S (2012) Disentangling the physical contributions to the electrical resistance in magnetic domain walls: a multiscale study. Phys Rev Lett 108(7):077201. http://link.aps.org/doi/10.1103/PhysRevLett.108.077201

Shiino T, Kab-Jin K, Lee KS, Park BG (2017) Inertia-driven resonant excitation of a magnetic skyrmion. Sci Rep 7(1):13993. https://doi.org/10.1038/s41598-017-13241-2

Staňo M, Schaefer S, Wartelle A, Rioult M, Belkhou R, Sala A, Menteş TO, Locatelli A, Cagnon L, Trapp B, Bochmann S, Martin S, Gautier E, Toussaint JC, Ensinger W, Fruchart O (2018) Flux-closure domains in high aspect ratio electroless-deposited CoNiB nanotubes. SciPost Phys 5:038. https://scipost.org/10.21468/SciPostPhys.5.4.038

Tadmor EB, Miller RE (2012) Modeling materials: continuum, atomistic and multiscale techniques, 1st edn. Cambridge University Press, Cambridge/New York

Thiaville A, García JM, Dittrich R, Miltat J, Schrefl T (2003) Micromagnetic study of Bloch-point-mediated vortex core reversal. Phys Rev B 67(9):094410. http://link.aps.org/doi/10.1103/PhysRevB.67.094410

Thiele AA (1973) Steady-state motion of magnetic domains. Phys Rev Lett 30(6):230–233. http://link.aps.org/doi/10.1103/PhysRevLett.30.230

Van de Wiele B, Manzin A, Vansteenkiste A, Bottauscio O, Dupré L, De Zutter D (2012) A micromagnetic study of the reversal mechanism in permalloy antidot arrays. J Appl Phys 111(5):053915. http://aip.scitation.org/doi/full/10.1063/1.3689846

Yan M, Andreas C, Kákay A, García-Sánchez F, Hertel R (2011) Fast domain wall dynamics in magnetic nanotubes: suppression of Walker breakdown and Cherenkov-like spin wave emission. Appl Phys Lett 99(12):122505. https://doi.org/10.1063/1.3643037, http://link.aip.org/link/APPLAB/v99/i12/p122505/s1&Agg=doi

Yan M, Andreas C, Kákay A, García-Sánchez F, Hertel R (2012) Chiral symmetry break-
ing and pair-creation mediated Walker breakdown in magnetic nanotubes. Appl Phys
Lett 100(25):252401–252401–4. https://doi.org/10.1063/1.4727909, http://apl.aip.org/resource/
1/applab/v100/i25/p252401_s1

Zamora RJ, Nair AK, Hennig RG, Warner DH (2012) Ab initio prediction of environmental
embrittlement at a crack tip in aluminum. Phys Rev B 86(6):060101. http://link.aps.org/doi/
10.1103/PhysRevB.86.060101

Zhu JG, Zheng Y (2002) The micromagnetics of magnetoresistive random access memory.
In: Spin dynamics in confined magnetic structures I, Topics in applied physics. Springer,
Berlin/Heidelberg, pp 289–325. https://link.springer.com/chapter/10.1007/3-540-40907-6_9

Atomistic Spin Dynamics

20

Richard F. L. Evans

Contents

Abstract

Atomistic spin models describe a class of models which approach a discrete limit of magnetic materials, where each atom is ascribed a localized atomistic magnetic moment or *spin*. These atomic spins interact principally by the Heisenberg exchange interaction, leading to long-range magnetic order and resulting in macroscopic magnetic properties. Here we review the theoretical foundations and recent developments of atomistic spin dynamics and their application to advanced magnetic materials, thermal and ultrafast spin dynamics.

R. F. L. Evans (✉)
The Department of Physics, The University of York, Heslington, York, UK
e-mail: richard.evans@york.ac.uk

1   Introduction

The theoretical foundations of spin models were established early in the twentieth century through the discovery of the quantum mechanical property of spin (Gerlach and Stern 1924) and the formulation of the Heisenberg exchange interaction (Heisenberg 1928) giving a microscopic origin of the Weiss field (Weiss 1907). Atomistic spin models were first realized in the model of Ernst Ising (1925) which was the first attempt to formulate a description of macroscopic magnetic order from a microscopic level, which was solvable analytically. Spin models only became practically useful tools with the advent of computational physics, where the finite size properties of magnets could be studied numerically using Metropolis Monte Carlo (Metropolis et al. 1953; Binder 1969). A limitation of the Ising model is the imposition of an infinite anisotropy due to quantization of the spin direction, making it inapplicable to real magnetic materials. A generalization of the Ising model is to remove the quantization of spin and treat the atomic spins as purely classical objects, allowing them to vary freely in 3D space. This allows for more realistic simulation of magnetic materials with finite anisotropies and more complex interactions than described by a simple nearest neighbor exchange interaction, enabling Monte Carlo simulations of equilibrium properties for more realistic systems than the Ising model could describe.

With the development of general purpose computers and programming languages such as FORTRAN and C, it became possible to develop more complex algorithms and software capable of performing numerical micromagnetics. Within the micromagnetic formalism, the atomic magnetic properties are averaged over a small volume of space typically $(2–10\,\text{nm})^3$ by assuming that the magnetization direction within the cell varies continuously. This approximation is known as a continuum approximation and allows the translation of the electronic properties to equivalent large scale micromagnetic equations (Fidler and Schrefl 2000). This first allowed the solution of the ground state magnetic configuration of complex magnetic structures through energy minimization and then the development of simulated magnetization dynamics by numerical solution of the Landau-Lifshitz-Gilbert equation (Fidler and Schrefl 2000). Today micromagnetics is a standard research tool due to the availability of user friendly software packages such as OOMMF and MUMAX3 and the ability to simulate relatively large structures accessible experimentally.

Despite the ubiquity of micromagnetic simulations, they are fundamentally constrained by the continuum formulation which requires that the magnetization direction varies *slowly* in space and time. This limitation means that the dynamics of ferrimagnets and antiferromagnets cannot be properly simulated since the intrinsic magnetization direction varies strongly on the atomic scale. At high temperatures the same problem exists since the thermal fluctuations lead to atomic scale spin disorder. Many technologically important materials and devices rely on ultrathin films and structures where the material properties themselves vary on the atomic length scale and cannot be properly described by the continuum approximation. Thus, the limited applicability of the micromagnetic formalism presents a barrier to understanding the dynamics of many magnetic materials and devices.

A major innovation came with the implementation of atomistic spin dynamics, where the time-dependent behavior of atomic spins is simulated using the Landau-Lifshitz-Gilbert (LLG) equation (Landau and Lifshitz 1935; Gilbert 1955). Although pioneered in the late twentieth century (Antropov et al. 1995; Krech et al. 1998), atomistic spin dynamics simulations were too computationally expensive to be practically useful until the turn of the twenty first century. The theoretical framework with an assumed classical equation of motion allowed the natural consideration of antiferromagnets and ferrimagnets (Nowak 2007), as well as atomic level defects found in nanoparticles, surfaces, and interfaces where the magnetization direction can change significantly at the atomic scale. Atomistic spin dynamics also enables the natural study of high temperature magnetization dynamics near and above the Curie temperature where the spins are paramagnetic and also spin excitations and the spin wave density of states (DOS) (Skubic et al. 2008). All of these features avoid the fundamental continuum limitation of micromagnetics and has led to a renaissance of magnetic materials modeling where the macroscopic properties of magnetic materials can be simulated using a microscopic model with a direct link to the underlying electronic properties responsible for the emergence of magnetic interactions.

2   Atomistic Spin Models

The basis of the atomistic spin model is a discrete description of magnetism, where each atom possess a localized magnetic moment or *spin*. The validity of localized spin models is often a topic of historical debate since it seems to conflict with the Stoner model of magnetism, which explains the existence of non-integer magnetic moments in metals due to the band structure. The development of first principles calculations of the spin polarized electron density revealed that in Fe and Co the electrons responsible for the magnetic moment are well localized to the nucleus and the bonding electrons tend to have only a small net spin. This supports the localized moment picture for elemental Fe and Co, while the small moment in Ni suggests that longitudinal Stoner spin fluctuations are important (Pan et al. 2017). The local moment approximation is also not well justified for induced magnetic moments such as those found in Pt and Rh in close proximity to a $3d$ magnetic metal where the spin density is distributed over the unit cell volume.

The interactions between atomistic spins are encapsulated in the spin Hamiltonian: the full Hamiltonian of the electron system with all spin-independent terms removed. A typical spin Hamiltonian $\mathscr{H}$ consisting of Heisenberg exchange, uniaxial anisotropy, and coupling to an external applied magnetic field is given by

$$\mathscr{H} = -\sum_{i<j} J_{ij}\mathbf{S}_i \cdot \mathbf{S}_j - \sum_i k_{\mathrm{u}}\left(\mathbf{S}_i \cdot \mathbf{e}_i\right)^2 - \sum_i \boldsymbol{\mu}_i\mathbf{S}_i \cdot \mathbf{B}_{\mathrm{app}}, \qquad (1)$$

where i is the local atomic site, J_{ij} is the exchange interaction between spins, $\mathbf{S}_i$ is a unit vector giving the spin direction on atomic site i, $\mathbf{S}_j$ is a unit vector giving

the spin direction on a neighboring atomic site j, k_u is the uniaxial anisotropy constant on site i, $\mathbf{e}_i$ is a unit vector giving the magnetic easy axis for site i, $\boldsymbol{\mu}_i$ is the magnitude of the local atomic spin moment on site i, and $\mathbf{B}_\mathrm{app}$ is a vector describing the external applied magnetic (Zeeman) field.

In atomistic spin models, the exchange interaction is typically the strongest interaction by 2–4 orders of magnitude and is responsible for the alignment of atomic spins. In ferromagnets where $J_{ij} > 0$ this is apparent from the formation of magnetic domains which give rise to classic macroscopic magnetic behavior. In antiferromagnets where $J_{ij} < 0$, the atomic spins are aligned antiparallel, and so in weak external fields, the net magnetization is zero.

For applications the most important factor is often the magnetic stability of the material and its resistance to losing or changing its magnetic state. The stability of the material is determined by its preference for aligning along particular crystallographic orientations, a property usually known as magnetocrystalline anisotropy or magnetic anisotropy energy (MAE). The simplest form of anisotropy is uniaxial, where one magnetic axis is preferred over all others. The common magnets Fe and Ni have cubic anisotropy where the spins prefer to lie along six or eight symmetric crystal directions.

For small systems with a uniform shape such as nanoparticles, the exchange and anisotropy represent the most important interactions in the system. However, for larger systems, thin films, or elongated nanostructures, the dipole field becomes important and can dominate the magnetic structure. For a magnetic sample consisting of point dipoles, the magnetic dipole field $\mathbf{B}_\mathrm{dipole}^i$ at site i can be expressed as

$$\mathbf{B}_\mathrm{dipole}^i = \frac{\mu_0}{4\pi} \left(\sum_{i \neq j} \frac{3(\mathbf{S}_j \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{S}_j}{r^3} \right), \tag{2}$$

where r is the separation distance between atomic dipoles i and j and $\hat{\mathbf{r}}$ is a unit vector in the direction $i \rightarrow j$. This long-ranged interaction leads to a demagnetizing field which minimizes the surface magnetic charges. The main influence of the dipole field is shape anisotropy, where different shapes have preferred magnetic orientations. For example, thin films prefer an in-plane orientation of the magnetization or elongated nanoparticles prefer to align their magnetization along the long axis of the particle.

Numerically dipole fields represent a challenging problem since each and every dipole interacts with every other, making direct solution of Eq. 2 impossible for all but the smallest systems with a size of around 10,000 atoms or less. Therefore approximate solutions are often used treating the dipole fields with a micromagnetic approximation where the atomic magnetizations are averaged within a small volume of space (Evans et al. 2014). This allows the separation of time and length scales from the atomistic spin dynamics while incorporating the important physical effects of the demagnetizing field.

A realistic spin Hamiltonian for a specific material can invoke a number of more complex interactions, such as two-ion anisotropy (Callen and Callen 1966;

Mryasov et al. 2005), surface anisotropy (Yanes et al. 2007), Dzyaloshinskii-Moriya interactions (Dzyaloshinsky 1958; Moriya 1960), higher-order magnetocrystalline anisotropies (Skomski 2012), and higher-order four-spin terms (Barker and Chantrell 2015). These complex interactions are surprisingly common in most technologically relevant materials, but their application in spin dynamics simulations is a very recent development.

3   Spin Dynamics

While the spin Hamiltonian can describe a plethora of magnetic interactions and properties, by itself it is not especially useful. The many body problem of solving the time evolution of several interacting spins is not solvable analytically and so it is necessary to adopt classical simulations where the spin relaxation is treated adiabatically and without quantum mechanics. It is important to note however that all magnetic interactions are fundamentally quantum mechanical, and these are simply approximated as a classical potential with a classical evolution of the spin states. This allows for classical Monte Carlo simulations to calculate equilibrium thermodynamic properties, but these are unable to capture the time-dependent dynamics of the spin system. At the turn of the twenty first century, atomistic spin dynamics simulations became computationally feasible for the first time and over the past 20 years have evolved to become a standard modeling technique for magnetic materials.

The theoretical basis of spin dynamics is the time evolution of a single atomic spin given by the phenomenological Landau-Lifshitz-Gilbert (LLG) equation, first derived by Landau and Lifshitz (1935) and then extended by Gilbert to include a critical damping (Gilbert 1955). The equation incorporates a precession term which arises from the intrinsic quantum mechanical precession of spins and a relaxation term which aligns the net spin direction with an externally applied magnetic field. The LLG atomistic equation (Ellis et al. 2015) is given by

$$\frac{\partial \mathbf{S}_i}{\partial t} = -\frac{\gamma_i}{(1 + \lambda_i^2)}[\mathbf{S}_i \times \mathbf{B}_i + \lambda_i \mathbf{S}_i \times (\mathbf{S}_i \times \mathbf{B}_i)], \tag{3}$$

where $\mathbf{S}_i$ is the classical spin vector of length $|\mathbf{S}_i| = 1$, $\gamma_i = 1.76 \times 10^{11}$ $\mathrm{T}^{-1}\mathrm{s}^{-1}$ is the gyromagnetic ratio, λ_i is the dimensionless Gilbert damping parameter representing the coupling to the heat bath, and $\mathbf{B}_i$ is the effective magnetic field in units of tesla. At the atomistic level the, Gilbert damping is a coupling of the spin to the electronic system and lattice and as the spin precesses and leads to a loss of energy. Materials containing atoms with strong spin-orbit coupling such as Pt and Ta will typically have a high Gilbert damping (Pan et al. 2016), while insulating materials will typically have low damping. In general the lattice contribution to the Gilbert damping is weak. The effective magnetic field on each atomic spin is given by the derivative of the total spin Hamiltonian with respect to the local spin moment

plus a Langevin thermal term representing the thermal spin fluctuations (Brown 1963) and given by

$$\mathbf{B}_i = \boldsymbol{\zeta}_i(t) - \frac{1}{\mu_i} \frac{\partial \mathcal{H}}{\partial \mathbf{S}_i}, \tag{4}$$

where μ_i is the spin magnetic moment and $\boldsymbol{\zeta}_i(t)$ is a time-dependent random thermal field. The thermal fluctuations are represented by a stochastic Langevin process where the Brownian motion of the spins is simulated by a random thermal magnetic field. The statistical properties of the thermal field in time are described by a white noise process given by

$$\boldsymbol{\zeta}_i = \langle \boldsymbol{\zeta}_i^a(t) \boldsymbol{\zeta}_j^b(t) \rangle = 2\delta_{ij}\delta_{ab}(t - t') \frac{\lambda_i k_B T}{\mu_i \gamma_i}, \tag{5}$$

where

$$\langle \boldsymbol{\zeta}_i^a(t) \rangle = 0. \tag{6}$$

This stochastic Landau-Lifshitz-Gilbert equation of motion, often abbreviated to *sLLG*, describes the fundamental time dependence of classical spins in a Langevin heat bath. This enables the simulation of time-dependent phenomena and phase transitions of complex systems with an almost endless variety of fundamental interactions, properties, and structures. In particular spin waves are modeled explicitly encapsulating the essential physics of many magnetic materials at the atomic level.

The Landau-Lifshitz-Gilbert equation is the essential equation describing magnetization dynamics from atomic to micromagnetic and macroscopic length scales. The development of numerical micromagnetics required numerical solution of the LLG equation and over the years has attracted a wide range of numerical schemes including simple predictor-corrector methods, Runge-Kutta, semi-implicit, midpoint, and symplectic integration schemes. For micromagnetics the effective magnetic fields in the LLG equation are generally small (a few Tesla) and slowly varying and so the integration is essentially limited by the ability to exactly conserve the magnetization length during integration. For atomistic spin dynamics, the requirements are somewhat different since the exchange fields are large (typically 100–1000 T), and the thermal fluctuations are of a similar magnitude at high temperatures due to the atomic values of the spin moment. This firstly requires a numerical integration scheme that leads to the Stratonovich interpretation of the stochastic LLG equation, such as the Heun scheme (García-Palacios and Lázaro 1998). Secondly small time steps around 0.1 fs are required to correctly resolve the dynamics of the spins. This is because spin excitations at elevated temperatures and in ferrimagnets and antiferromagnets are especially large, leading to a rapidly changing effective field in both magnitude and direction. The rapidly changing field means that the fundamental dynamics are not limited by the numerical integration scheme but by the physics of spin dynamics at the atomic scale. Practically this means that there is no value in integration schemes more complex than Heun due to their higher numerical cost. Alternative approaches such as pulsed noise (Garanin

2017) may alleviate some of the problems with large spin excitations and allow larger time steps, but it is not yet clear that this is beneficial for antiferromagnets where the dynamics are intrinsically on the femtosecond timescale.

3.1   Quantum Statistics

In the above classical treatment, all quantum mechanical variables have been replaced with classical variables. The presence of a Langevin thermostat means that infinitesimal changes to spin orientations are allowed. In reality of course spin, transitions are discrete and caused by spin-flip scattering processes induced by incoming photons or electrons interacting with the spins. This means that there is a natural gap in the spectrum where low energy electron and photon excitations are unable to cause a spin flip. This behavior naturally leads to quantum statistics, where there are distinct spin energy levels that are populated. This can be represented in a spin model by considering the intensity of thermal fluctuations and changing their intensity to give the expected magnon distribution (Bergqvist and Bergman 2018) or parametrically by rescaling the effective temperature in the simulation (Evans et al. 2015) to give the experimentally measured temperature-dependent magnetization, which is a direct estimate of the magnon population. However, this still avoids the fundamental nature of the underlying microscopic spin-flip processes. Therefore, the thermodynamic properties of magnetic materials are only approximate and pure classical spin models typically give an incorrect temperature-dependent magnetization curve due to the classical nature of the statistics of the thermal bath.

4   Advanced Models of Magnetic Materials

Simple spin models are based on a straightforward extension of finite difference micromagnetics to atomic resolution, where each cell is the size of an atom and the continuum micromagnetic interactions are replaced with their discrete equivalents with a simple cubic lattice. While these models are qualitatively useful in describing thermal effects and atomic scale magnetism, they suffer from a weak relation to reality where the atomic structure, electronic properties, and magnetism interplay to give a complex range of magnetic phenomena in a wide diversity of magnetic materials. To better understand the magnetic properties with a high degree of specificity, it is necessary to formulate a realistic model of a material including the correct crystal structure and a realistic model of the magnetic interactions. Examples of atomistic spin models of specific materials include FePt (Mryasov et al. 2005), GdFe alloys (Ostler et al. 2011), Fe_3O_4 (Nedelkoski et al. 2017), FeRh (Barker and Chantrell 2015), YIG (Barker and Bauer 2016), $Nd_2Fe_{14}B$ (Toga et al. 2016), and IrMn (Szunyogh et al. 2009; Yanes et al. 2013), where the structure and magnetic interactions are included explicitly in the model. For atomistic spin models, such material-specific models are a relatively recent development, starting

around 20 years ago. In the future such models will be essential to fully understand the fundamental and thermodynamic properties of magnetic materials.

4.1   Amorphous GdFe Alloys

One of the main advantages of atomistic spin models over micromagnetics is the ability to correctly simulate the dynamics of ferrimagnetic and antiferromagnetic materials. In these materials, the localized atomic spins are exchange coupled in an antiparallel state. The partial or complete compensation of the magnetization at the atomic level leads to inertial spin dynamics with a rich variety of behavior making them particularly interesting at the fundamental level as well as practically important. Technologically one of the most important ferrimagnetic materials is GdFe, an amorphous alloy of the rare earth (RE) metal Gd and transition metal (TM) Fe. Here the Gd and Fe atoms form two magnetic sublattices that oppose each other and due to the different origin of the magnetic moments, *4f* and *3d* electrons, respectively. The Fe atoms draw their magnetic moments from the *3d* electrons, and the spin polarized electron density is quite well localized near the nucleus with an effective spin magnetic moment of $1.92\,\mu_B$ per atom. The Gd atoms have a magnetic moment due to the *4f* electron shell, which is lower in energy than the *6s* and *5d* shells and prefers unpaired filling of the orbitals. This leads to a large effective spin moment of $7.63\,\mu_B$ per atom. Since the *4f* electron shell is located away from the *5d* bonding electrons, the exchange coupling is weaker than in the transition metal elements, leading to a lower Curie temperature of 298 K for pure Gd metal. However, when alloyed together as GdFe, the effective Curie temperature is around 500 K which is sufficiently above room temperature to be useful for applications such as magneto-optical recording. This is because the exchange interaction from the Fe sublattice polarizes the Gd spin moments causing an increase in the effective exchange interaction and thus the Curie temperature for the coupled system.

It is possible to model the static and dynamic properties of GdFe alloys using an extended spin model explicitly considering Fe and Gd spin moments and Heisenberg exchange coupling (Ostler et al. 2011). Experimentally GdFe alloys are assumed to be amorphous where the Fe and Gd atoms are almost homogeneously distributed in the material with no short-range or long-range crystallographic order. This would add a significant degree of complexity to the simulations, with a wide range of exchange interaction strengths, coordination number, and density fluctuations. The simplest approach is to consider a magnetically amorphous material where the Gd and Fe sites are randomly distributed within a periodic crystal structure which captures the essence of the magnetic interactions and ferrimagnetic properties. Alloys used for magneto-optical recording applications typically have a composition close to $Gd_{25}Fe_{75}$ to balance the requirement for a low Curie temperature and larger magnetic moment when heating the material to enable recording of information. In thin film form when sandwiched between SiN glass, GdFe films have a strong perpendicular magnetic anisotropy, usually assumed to arise due to interfacial or surface anisotropy. Having defined the interactions in the systems, we can write down the spin Hamiltonian

$$\mathcal{H} = \mathcal{H}_{Gd} + \mathcal{H}_{Fe} \tag{7}$$

$$\mathcal{H}_{Gd} = -\sum_{i<j} J_{Gd}\mathbf{S}_i \cdot \mathbf{S}_j - \sum_{i<\delta} J_{GdFe}\mathbf{S}_i \cdot \mathbf{S}_\delta$$

$$-\sum_i k_u \left(\mathbf{S}_i \cdot \mathbf{e}_i\right)^2 - \boldsymbol{\mu}_{Gd}\sum_i \mathbf{S}_i \cdot \mathbf{B}_{app} \tag{8}$$

$$\mathcal{H}_{Fe} = -\sum_{\nu<\delta} J_{Fe}\mathbf{S}_\nu \cdot \mathbf{S}_\delta - \sum_{\nu<j} J_{GdFe}\mathbf{S}_\nu \cdot \mathbf{S}_j$$

$$-\sum_\nu k_u \left(\mathbf{S}_\nu \cdot \mathbf{e}_\nu\right)^2 - \boldsymbol{\mu}_{Fe}\sum_\nu \mathbf{S}_\nu \cdot \mathbf{B}_{app} \tag{9}$$

which explicitly lists inter-sublattice ($J_{GdFe} = -1.09 \times 10^{-21}$ J/link) and intra-sublattice ($J_{Gd} = 1.26 \times 10^{-21}$ J/link and $J_{Fe} = 2.835 \times 10^{-21}$ J/link) exchange interactions and different couplings to the applied field (Ostler et al. 2011). The anisotropy is assumed to be uniaxial due to the surface effect of SiN capping layers used in experiments with a value of $k_u = 8.07246 \times 10^{-24}$ J/atom. The form given here is encapsulated in the simplified form in Eq. 1 but is more specific in identifying relevant sets of interactions and values in the system.

Having defined the interactions in the system, it is then possible to simulate the equilibrium temperature-dependent properties using atomistic spin dynamics. Figure 1 shows the calculated temperature dependence of the equilibrium sub-lattice and total magnetizations and reproduces the main features of the material seen experimentally: a magnetic compensation point and different temperature-dependent magnetizations of the Fe and Gd sublattices (Ostler et al. 2011). It is important to note that the two magnetic sublattices in GdFe have the same Curie temperature, while their different temperature dependencies come from the different intrinsic exchange interactions, being much weaker for Gd spins compared with Fe. At the compensation point, the net magnetization is close to zero and so macroscopically the system of spins behaves like an antiferromagnet. The existence of a compensation point is only possible due to the different temperature dependences of the Fe and Gd sublattices. This arises because the Gd spins are polarized by the exchange interaction from neighboring Fe spins, which leads to order in the Gd sublattice above its intrinsic Curie temperature of 298 K. However, the spin fluctuations on the Gd sites are larger than the Fe sites due to the lower total exchange, which significantly reduces the Curie temperature of the combined system.

4.2 Magnetite

Magnetite is the oldest known magnetic material, first documented by the ancient Greeks and known in medieval times as lodestone and used as a primitive compass. Despite its long history, the origins of its magnetism are still an active subject of investigation. Magnetite is an oxide of iron with formula unit Fe_3O_4 and has a complex inverse spinel crystal structure, as shown in Fig. 2. The structural
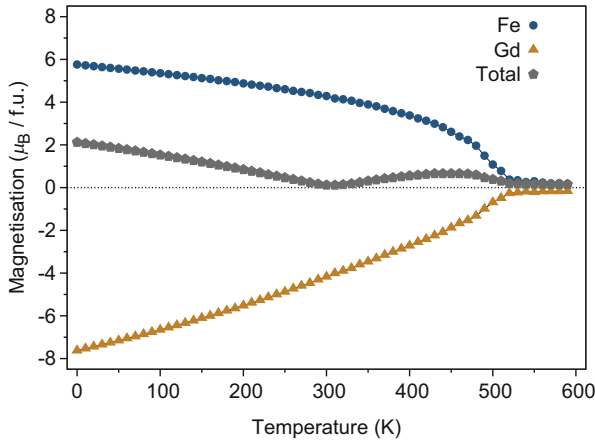
Fig. 1 Plot of the simulated sublattice temperature-dependent magnetization of $Gd_{25}Fe_{75}$ random alloy showing the effect of increased spin fluctuations on the total magnetization. The polarizing effect of the Fe spins on the Gd spins leads to a different temperature dependence of the magnetization as seen from the shape of sublattice magnetization curves. The Fe sublattice exhibits a typical classical $M_s(T)$ curve showing critical behavior near the Curie temperature T_C, while the Gd sublattice shows an almost linear behavior. The ferrimagnetic nature of the exchange coupling between the sublattices and the different temperature-dependent magnetizations naturally leads to a magnetic compensation point where the net magnetization is close to zero. The compensation point leads to a range of complex behaviors due to the inertial magnetization dynamics

Fig. 2 Visualization of the inverse spinel crystal structure of magnetite (Fe_3O_4) showing the different magnetic site symmetries. (Image courtesy of Daniel Meilak)
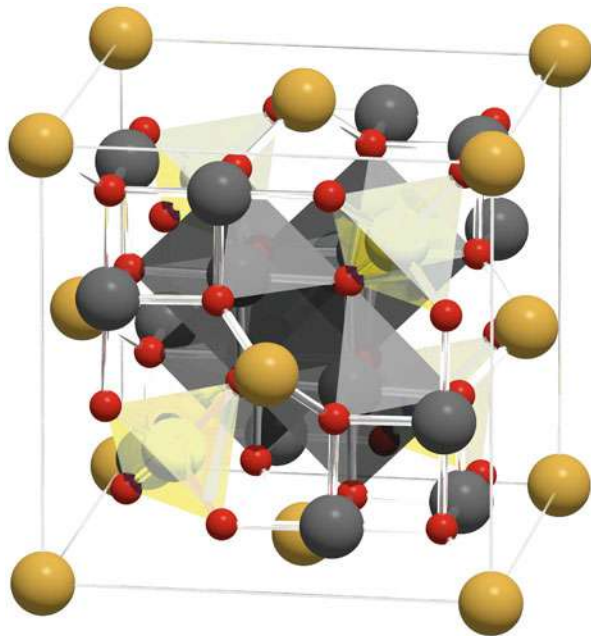
complexity is typical of the vast majority of magnetic materials and represents a recent development in materials modeling where the structural and magnetic properties are strongly related and must be properly represented in the simulations. In the case of magnetite, the exchange interactions between Fe sites arise due to the double exchange interaction mediated by the oxygen atoms where its strength is determined by the relative overlap of the atomic orbitals. The crystal structure of magnetite contains two distinct sites based on the local oxygen symmetry: tetrahedral and octahedral. The site symmetries define two distinct magnetic sites with different coordination number and localized magnetic moments on the Fe ions due to the different valence states of the Fe. Expanding the spin Hamiltonian into these components we therefore have

$$\mathscr{H} = \mathscr{H}_{\text{Fe}}^{\text{T}} + \mathscr{H}_{\text{Fe}}^{\text{O}} \tag{10}$$

$$\mathscr{H}_{\text{Fe}}^{\text{T}} = -\frac{1}{2} \sum_{i,j} J_{\text{Fe}}^{\text{TT}} \mathbf{S}_i \cdot \mathbf{S}_j - \frac{1}{2} \sum_{i,\delta} J_{\text{Fe}}^{\text{TO}} \mathbf{S}_i \cdot \mathbf{S}_\delta$$

$$-\frac{k_c}{2} \sum_i \left(S_x^4 + S_y^4 + S_z^4 \right) - \mu_{\text{Fe}}^{\text{T}} \sum_i \mathbf{H}_{\text{app}} \cdot \mathbf{S}_i \tag{11}$$

$$\mathscr{H}_{\text{Fe}}^{\text{O}} = -\frac{1}{2} \sum_{v,\delta} J_{\text{Fe}}^{\text{OO}} \mathbf{S}_v \cdot \mathbf{S}_\delta - \frac{1}{2} \sum_{v,j} J_{\text{Fe}}^{\text{TO}} \mathbf{S}_v \cdot \mathbf{S}_j$$

$$-\frac{k_c}{2} \sum_v \left(S_x^4 + S_y^4 + S_z^4 \right) - \mu_{\text{Fe}}^{\text{O}} \sum_v \mathbf{H}_{\text{app}} \cdot \mathbf{S}_v, \tag{12}$$

where T and O indicate tetrahedrally coordinated and octahedrally coordinated Fe ions, respectively. The exchange coupling constants $J_{\text{Fe}}^{\text{TT}} = -2.2 \times 10^{-22}$ J/link, $J_{\text{Fe}}^{\text{TO}} = -5.256 \times 10^{-21}$ J/link, and $J_{\text{Fe}}^{\text{OO}} = -1.0206 \times 10^{-21}$ J/link represent the super exchange interactions between the tetrahedrally and octahedrally coordinated spins. Magnetite exhibits cubic crystal symmetry and so cubic magnetocrystalline anisotropy is included at each Fe site where $k_c = 8.16 \times 10^{-25}$ J/atom is the effective cubic anisotropy at room temperature. Due to the double exchange interaction, the leading term $J_{\text{Fe}}^{\text{TO}}$ is antiferromagnetic which causes magnetite to be ferrimagnetic with opposed tetrahedral and octahedral magnetic sublattices. Unlike GdFe there is no magnetic compensation point in magnetite as the sublattices have a similar temperature dependence.

5   Applications of Spin Dynamics

The core methodology of spin dynamics has been established for over 20 years, but its usefulness at that time was previously limited by a lack of fundamental materials knowledge and the computational expense of modeling reasonably sized systems with simple serial codes. Materials knowledge has advanced considerably

in that time, in particular new capabilities of first principles methods such as the Korringa-Kohn-Rostoker (KKR) method and conventional relativistic density functional theory (rDFT) to determine magnetic properties and parameters and map them to an effective Heisenberg model. The development of parallel spin dynamics codes such as VAMPIRE (https://vampire.york.ac.uk) and UPPASD (http://physics.uu.se/forskning/materialteori/pagaende-forskning/uppasd/) has enabled the simulation of much larger systems than previously possible, using thousands of cores for large-scale simulations at sizes comparable with micromagnetics. These improvements have enabled spin dynamics simulations to address fundamental challenges in magnetic materials modeling in a wide range of applications including heat-assisted magnetic recording (HAMR), magnetic random access memory (MRAM), nanoparticle magnetism, ultrafast magnetism, permanent magnets, and spintronics. It is impossible to cover the full range of applications in the present chapter, but to give a flavor of the capabilities of spin dynamics, we illustrate a few topical examples from slow to ultrafast magnetic processes.

5.1   Calculation of the Effective Gilbert Damping

The Gilbert damping parameter describes the ability of a magnetic material to respond to a magnetic field. The intrinsic damping is mainly electronic in origin (Pan et al. 2016) arising from spin-orbit coupling. However, real materials are often far from ideal and exhibit enhanced damping due to defects and impurities. Another important origin of enhanced damping is caused by thermal spin fluctuations, more commonly known as spin wave scattering. These extrinsic contributions to damping can be modeled directly using spin dynamics by performing a simple magnetic relaxation experiment. The idea is to simulate the relaxation of the net magnetization $\mathbf{M}$ toward a magnetic field $\mathbf{B}_{app}$ after performing an arbitrary rotation of the spin direction by an angle θ. Since the dynamics of the total magnetization follows the Landau-Lifshitz-Gilbert equation which has an analytical solution (Evans et al. 2014), it is possible to extract an effective damping constant from the simulation by fitting to the time-dependent relaxation of the net magnetization.

Here we consider a simple ferromagnetic $(10\,\text{nm})^3$ cube of α-Fe with zero magnetocrystalline anisotropy. The effect of spin fluctuations on the damping is calculated by simulating the time-dependent relaxation of a small block of material. The system is first equilibrated at the simulation temperature to thermalize the spins in a strong field. The net magnetization of the system is then instantaneously rotated to an angle of 30 degrees to the applied field direction which causes the net magnetization to precess around the field and gradually relax toward the field direction. Since the microscopic and macroscopic equations of motion are the same, in the absence of spin fluctuations, the input and calculated damping are expected to be the same. The ability to extract the effective Gilbert damping from the simulation requires that applied magnetic field is much larger than the equivalent anisotropy field; otherwise, the spin-dependent fields lead to a large error in the fitting. Zero anisotropy removes any restrictions on the field which can be used,
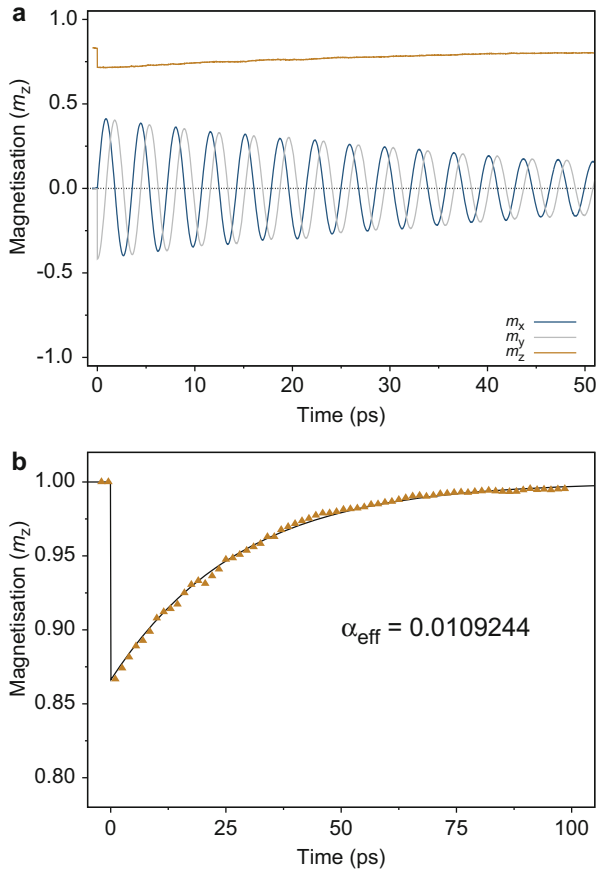
Fig. 3 (**a**) Plot of the simulated magnetization dynamics of a $(10\,nm)^3$ block of α-Fe at $T = 800\,K$ in a $10\,T$ applied field after an instantaneous rotation of the magnetization to $30°$ from the field direction. The rotation causes the magnetization to precess around the field direction and relax toward the applied field direction. The intrinsic Gilbert damping at $T = 0\,K$ is $\alpha = 0.01$. (**b**) Fitting of the normalized z-component of the magnetization from the analytic solution (Evans et al. 2014). The effective damping shows a small enhancement due to thermal spin fluctuations at elevated temperatures

while for materials with a large anisotropy, much larger fields can be applied. The result of the relaxation simulation for a 10T applied field and intrinsic damping of $\lambda = 0.01$ is shown in Fig. 3 for a temperature of 800 K, a significant fraction of the Curie temperature of 1048 K where spin fluctuations are relatively large. The lines show a fit to the analytic solution incorporating a phase shift due to the initial starting angle of the magnetization of $30°$ from the applied field direction. Here the thermal spin fluctuations lead to a small increase in the effective damping of around 10%. Modeling other contributions to the Gilbert damping is at an early stage, but spin-lattice dynamics offer an enticing route to understanding spin-phonon

interactions, particularly in magnetic oxides such as yttrium iron garnet ($Y_3Fe_5O_{12}$) and magnetite (Fe_3O_4) where the electronic damping is small.

5.2   Magnetization Dynamics in Fe_3O_4 Nanoparticles

An important property of nanoparticles is superparamagnetism, where the thermal fluctuations are sufficiently strong to make the magnetization direction of the particle unstable. The magnetocrystalline anisotropy of magnetite is relatively small, and so superparamagnetism is a common feature of nanoparticles in the 5–15 nm size range typically used for applications in magnetic hyperthermia. Using atomistic spin dynamics, we are able to directly simulate the thermodynamic properties of small magnetite nanoparticles. Using the model of magnetite described earlier, we can build an idealized single-crystal magnetite nanoparticle with an octahedral particle geometry, as shown in Fig. 4. The particle faceting is representative of experimental samples where [111] facets are energetically favorable.

The dynamics of the particle are simulated for a total of 10 ns, but even over this very short timescale, the magnetization is found to be unstable. Due to the cubic magnetocrystalline anisotropy the, minimum energy directions are along the [111] crystal directions. This can be seen from the simulations since the magnetization fluctuates in three dimensions between these eight minima. Extended calculations such as these provide detailed information about the relaxation rates of the particles which is important for other modeling techniques which assume a particular relaxation attempt frequency. Spin dynamics simulations enable this important
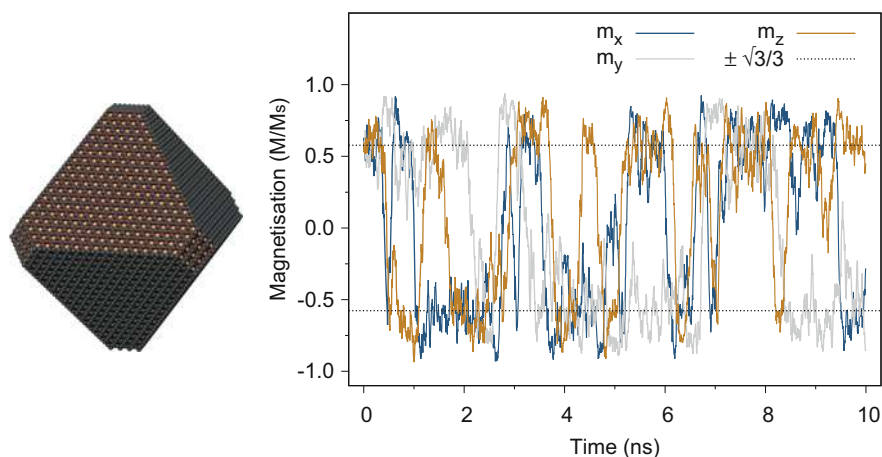
Fig. 4 Visualization of a 12 nm Octahedral magnetite nanoparticle (left) and plot of the simulated time-dependent magnetization dynamics at $T = 300$ K (right). The data show fluctuations of the different magnetization components between the 8 minima along the [111] crystal directions due to the cubic magnetocrystalline anisotropy, indicated by the dashed lines

property to be directly simulated for a range of situations such as enhanced surface anisotropy or in the presence of defects such as antiphase boundaries (Nedelkoski et al. 2017).

5.3   Ultrafast Demagnetization of Ni

The dynamics of magnetic materials were once thought to be slow, and limited by the precession frequency of a magnetic spin in a field, given by the Larmor frequency $\omega = -\gamma|\mathbf{B}|$, where $\gamma = 1.76 \times 10^{11}$ T^{-1}s^{-1} is the gyromagnetic ratio and $|\mathbf{B}|$ is the applied magnetic field strength. For everyday magnetic field strengths of around 1 T and where the Gilbert damping is included, this gives a characteristic precession time $\tau_p = -2\pi(1+\alpha^2)/(\gamma|\mathbf{B}|)$ of around 70 picoseconds $(70 \times 10^{-12}$ s). The relaxation time is a multiple of the precession time $\tau_r = \tau_p/\alpha$, where the magnetization has to reorient itself with an external magnetic field. For critical Gilbert damping $\alpha = 1$, the relaxation time is the same as the precession time. However most magnetic materials have much lower damping between 0.001 and 0.1 leading to dynamics on the nanosecond timescale.

In 1996, pioneering experiments showed that metallic Ni responded to a femtosecond laser pulse on the sub-picosecond timescale (Beaurepaire et al. 1996). At the time this was a controversial discovery since it shows magnetic relaxation three orders of magnitude faster than the intrinsic relaxation time from the known properties of magnets. The ultrafast relaxation is caused by thermal spin fluctuations induced through rapid laser heating, though the microscopic mechanism responsible for the transfer of angular momentum from the spins is still debated. At the time the transfer of energy to the spins was described with a simple phenomenological three-temperature (3TM) model (Kirilyuk et al. 2010). The model describes the coupling of electron, lattice, and spin systems as reservoirs for energy. The laser pulse interacts directly with the electrons in the material causing a large increase in their thermal energy. Over the course of a few picoseconds, the hot electrons interact with the lattice causing an increase in the lattice temperature. The heat capacity of the electron system is quite small and so the electron system reaches a much high temperature than the lattice. In a magnetic material, the spins act as another reservoir which can exchange energy with the electrons and lattice. Atomistic spin dynamics is able to simulate this thermal energy transfer directly by calculating the effect of rapid heating, where it can be assumed that the hot electrons interact directly with the 3d spins in the material. In this case one can use the two-temperature model to simulate the dynamic response of the electron and lattice temperatures to the laser pulse. This is expressed by the coupled equations

$$\left(T_e C_e \frac{\partial T_e}{\partial t}\right) = -G_{el}\left(T_e - T_l\right) + P(t) \tag{13}$$

$$C_l \frac{\partial T_l}{\partial t} = G_{el}\left(T_e - T_l\right), \tag{14}$$

where T_e is the electron temperature, C_e is the electron heat capacity, G_{el} is the electron-lattice coupling, T_l is the lattice temperature, C_l is the lattice heat capacity, and $P(t)$ is a time-dependent energy source from the laser heating. The electron heat capacity is assumed to scale with the electron temperature, known as the free electron approximation. The laser heating is assumed to have a Gaussian profile in time $P(t)$ given by

$$P(t) = \frac{P_0}{\sqrt{2\pi}} \exp\left(\frac{-(t - t_0)}{t_p^2}\right), \tag{15}$$

where P_0 is the laser power, t_p is the pulse width, t is the time, and t_0 is the arrival time of the pulse. These equations can now be solved numerically to determine the response of the electron and lattice temperatures of a Ni surface to a 50 fs laser pulse. A typical time evolution is shown in Fig. 5a. The smaller heat capacity of the electron leads to a large transient increase in the electron temperature caused by the laser energy. After a few hundred femtoseconds, the electron system begins to transfer energy to the lattice which causes a slower increase in the lattice temperature. The final equilibrium temperature of the system is much lower than the peak electron temperature due to the much larger heat capacity of the lattice.

We now simulate the effect of this transient temperature rise caused by the laser pulse on the magnetization of Ni by coupling the $3d$ spins to the electron temperature. We simulate the effect by considering a small segment of the Ni film $(10\,\text{nm})^3$ with periodic boundary conditions in the $x - y$ plane. The response of the z-component of the magnetization is shown in Fig. 5b. Initially before the pulse, the magnetization shows incomplete magnetic order due to the natural thermal spin fluctuations at an ambient temperature of 300 K. When the laser pulse arrives, there is a rapid decrease in the magnetization due to the rapid temperature rise. This causes large thermal spin fluctuations at the atomic level and observed as a transient decrease in magnetic order. A snapshot visualization of the spins during the maximum demagnetization is shown in Fig. 5c. During this process the net angular momentum is transferred to the lattice by electron-lattice interactions. After the pulse has gone, the reduction in the electron temperature leads to a recovery of the magnetization on a longer timescale than the initial demagnetization process. At longer timescales the magnetization reaches a new equilibrium value due to the deposition of heat in the sample by the laser pulse.

Through spin dynamics simulations, we are therefore able to model and understand the basic thermalization process of spins due to laser heating. However, this leaves several outstanding questions. The light matter interactions are complex due to the intensity of the pulse, leading to transient changes of the electronic structure and ultrafast spin currents that are not included within the simple approach above. Also the model completely neglects the underlying spin-flip scattering processes that are responsible for the microscopic spin fluctuations and angular momentum transfer. For the first few fs, the fundamental magnetic parameters are also strongly time and temperature dependent due to the effects of the laser on the underlying

Fig. 5 (**a**) Plot of the simulated time dependence of the electron and lattice temperatures from numerical solution of the two-temperature model. (**b**) Simulated transient magnetization dynamics of a $(10\,\text{nm})^3$ block of Ni excited by a 50 fs laser pulse. The simulations show excellent agreement with experimental measurements (Beaurepaire et al. 1996). (**c**) Visualization of the spin configuration during demagnetization showing thermalization of the spins
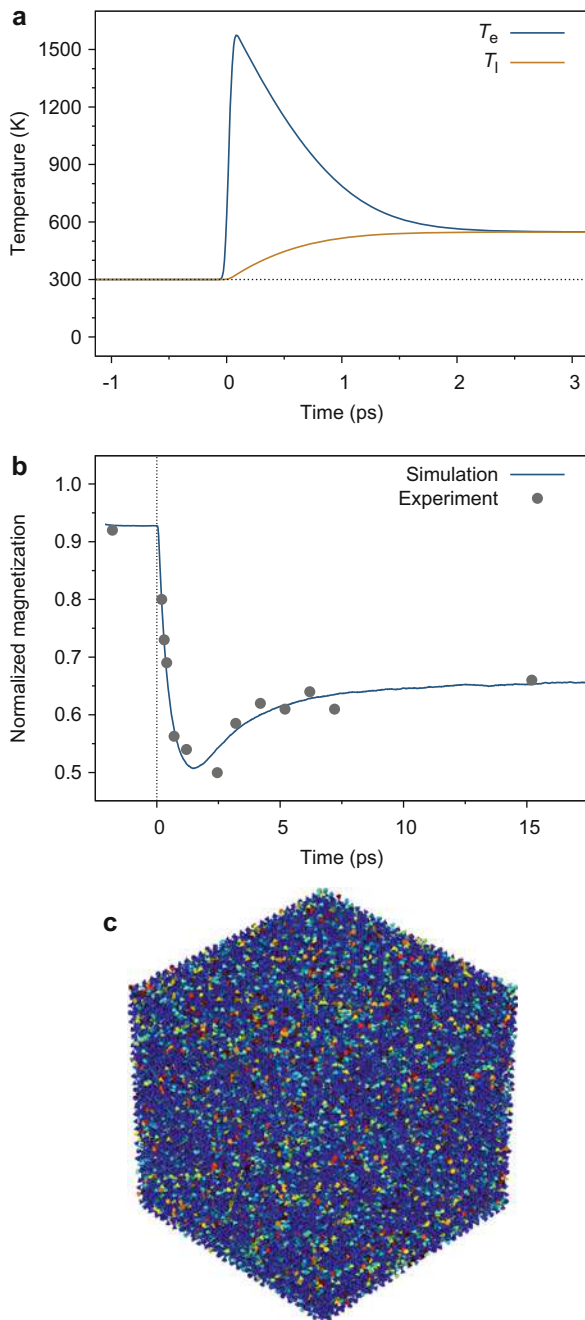
electronic structure (Simoni et al. 2017). This makes the field of ultrafast magnetism one of the most exciting areas of the field today and will almost certainly lead to huge advances in our understanding of magnetic processes at the fundamental atomic level.

5.4   Heat-Induced Switching of GdFe

In the previous section, the thermalization of spins described the demagnetization of Ni. More recently the ability to completely control magnetic order just with light was demonstrated in amorphous GdFe (Stanciu et al. 2007). Here pulses of circularly polarized light were able to change the direction of the magnetization, rather than just demagnetization. At the time this was thought to be due to the inverse Faraday effect which induces a net angular moment from the circular polarization of the light. However, further investigation of the sublattice magnetization dynamics with X-ray magnetic circular dichroism (XMCD) and complementary modeling with atomistic spin dynamics (Radu et al. 2011) showed a complex switching mechanism where the Fe and Gd sublattice magnetizations demagnetized at different rates and at some point are aligned in parallel, despite the strong antiferromagnetic coupling between them. These complex dynamics were confirmed by atomistic spin dynamics simulations (Radu et al. 2011) and showed that angular momentum transfer was an important component of the switching process. During the course of this work, the spin dynamics simulations made a startling prediction that the reversal occurred in the absence of a magnetic fields and linearly polarized laser light, essentially *heat* induced magnetic switching. This was possible due to the realistic model of the GdFe material that had been developed, explicitly including the dynamics of Fe and Gd sublattices and their response to an ultrafast heat pulse. This remarkable prediction was confirmed experimentally (Ostler et al. 2012) through laser heating and more recently through pure hot electron transport (Yang et al. 2017), confirming that ultrafast heating alone is sufficient to cause magnetic switching in GdFe.

We will now reproduce this predictive result with a spin dynamics simulation, combining the earlier GdFe model with an ultrafast heat pulse. The model parameters are unchanged apart from a more realistic value of the Gilbert damping of $\lambda = 0.1$ since we are interested in reproducing realistic dynamics rather than equilibrium properties. The response of the system to the heat pulse is shown in Fig. 6. As with the Ni simulation earlier, the initial magnetization shows reduced order before the pulse due to thermal spin fluctuations. When then laser pulse arrives, the transient increase in the electron temperature causes rapid demagnetization of the Fe and Gd sublattices due to the heating. However, due to the different local atomic magnetic moments, the demagnetization times are different for each sublattice. This is because the larger magnetic moments on the Gd sites are less strongly affected by thermal fluctuations responsible for demagnetization. This in turn affects the magnetization dynamics, making the Gd sublattice slow to respond to the laser heating pulse. At the same time, the Fe sublattice fully demagnetizes. After the
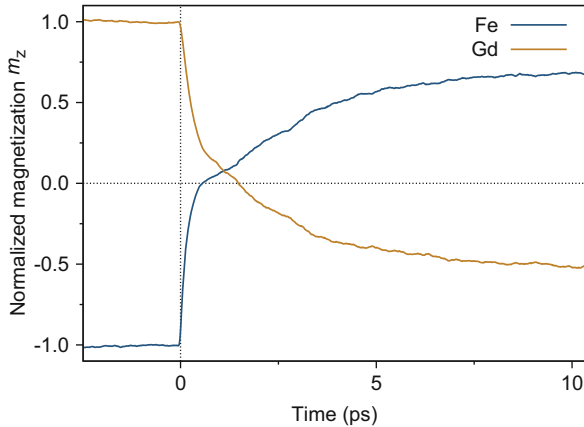
Fig. 6 Simulated heat-induced magnetic switching of GdFe be 50 fs laser pulse

pulse is gone, the system attempts to re-magnetize, but the exchange field from the partially demagnetized Gd sublattice causes the Fe sublattice to process rapidly. The Fe precession excites a mutual precession of the Gd sublattice which leads to a dynamic switching process. A fraction of a picosecond later, the magnetization recovers sufficiently to cause the sublattices to align with the magnetic easy axis, completing the switching process (Atxitia et al. 2013). After cooling back to ambient temperature and pulsing again, this process is repeatable and deterministic. This concisely demonstrates the predictive power of atomistic spin dynamics simulations which will only improve over the coming years.

6   The Future

Despite the impressive capabilities of spin dynamics simulations, significant challenges remain in being able to fully understand the dynamics of magnetic materials both in terms of atomic structure and magnetic properties. A new class of model has recently emerged which couples spin and molecular dynamics in a single simulation (Ma et al. 2008; Perera et al. 2016, 2017). This allows an investigation of spin-lattice effects in a direct manner including contributions to the Gilbert damping and magneto volume effects, where the magnetic properties are linked to the lattice parameter. Such models are more computationally expensive than spin models but offer significant improvements in our ability to understand the dynamic properties of magnetic materials.

Another more fundamental limitation of the classical Heisenberg model is the use of classical statistics for the Langevin heat bath. In reality the underlying microscopic mechanisms are discrete electron excitations, causing small jumps in the local spin direction. This process is much better represented by quantum statistics, where infinitesimal spin fluctuations are forbidden. For larger ensembles

of spins, the temperature rescaling method includes an average effect of the lack of infinitesimal spin states by including smaller thermal fluctuations. However, smaller systems such as molecular magnets or isolated spins will require a better treatment of spin fluctuations and single spin dynamics.

Another class of magnetic moment not well treated by a fixed spin approximation induced magnetic moments or moments that are more Stoner-like rather than Heisenberg-like, where their length varies significantly in time. In this case the fixed moment neglects an important degree of freedom in the material and is in most cases a poor approximation. This is most evident in heavy metals such as Pt which has a relatively strong induced moment due to the magnetic proximity effect, where a nearby magnetic atom induces a net moment due to the exchange interaction (Mryasov et al. 2005). Ni also shows some Stoner-like behavior that is not particularly well represented by a fixed spin approximation. However, models of variable spin are much more computationally expensive and so are not currently as widely applicable as standard spin dynamics. In the future models that can better deal with these induced moments will be needed to better understand their behavior.

7   Conclusion

Spin dynamics is becoming a standard research method in the field of magnetism and spintronics due to the ability to model a diverse range of magnetic materials and phenomena, with an increasing ability to be predictive. Through a combination of advanced materials modeling of the structure of materials, their complex magnetic interactions, and spin dynamics simulations, it is now possible to model a wide range of magnetic phenomena which were poorly understood only 20 years ago. This rapidly developing capability will complement ongoing experimental efforts to make devices for a range of practical applications.

References

Antropov VP, Katsnelson MI, van Schilfgaarde M, Harmon BN (1995) *Ab Initio*. Phys Rev Lett 75:729–732. https://doi.org/10.1103/PhysRevLett.75.729

Atxitia U, Ostler T, Barker J, Evans RFL, Chantrell RW, Chubykalo-Fesenko O (2013) Ultrafast dynamical path for the switching of a ferrimagnet after femtosecond heating. Phys Rev B 87(22):224417. http://prb.aps.org/abstract/PRB/v87/i22/e224417

Barker J, Bauer GEW (2016) Thermal spin dynamics of yttrium iron garnet. Phys Rev Lett 117:217201. https://doi.org/10.1103/PhysRevLett.117.217201

Barker J, Chantrell RW (2015) Higher-order exchange interactions leading to metamagnetism in FeRh. Phys Rev B 92:094402. https://doi.org/10.1103/PhysRevB.92.094402

Beaurepaire E, Merle JC, Daunois A, Bigot JY (1996) Ultrafast spin dynamics in ferromagnetic nickel. Phys Rev Lett 76:4250–4253. https://doi.org/10.1103/PhysRevLett.76.4250

Bergqvist L, Bergman A (2018) Realistic finite temperature simulations of magnetic systems using quantum statistics. Phys Rev Mater 2:013802. https://doi.org/10.1103/PhysRevMaterials.2.013802

Binder K (1969) A Monte-Carlo method for the calculation of the magnetization of the classical Heisenberg model. Phys Lett A 30(5):273–274. http://www.sciencedirect.com/science/article/pii/037596016990989X

Brown WF (1963) Thermal fluctuations of a single-domain particle. Phys Rev 130(5):1677–1686. https://doi.org/10.1103/PhysRev.130.1677

Callen HB, Callen E (1966) The present status of the temperature dependence of magnetocrystalline anisotropy, and the l(l+1)2 power law. J Phys Chem Solids 27(8):1271–1285. http://www.sciencedirect.com/science/article/pii/0022369766900126

Dzyaloshinsky I (1958) A thermodynamic theory of weak ferromagnetism of antiferromagnetics. J Phys Chem Solids 4(4):241–255. https://doi.org/10.1016/0022-3697(58)90076-3, http://www.sciencedirect.com/science/article/pii/002236975890%0763

Ellis MOA, Evans RFL, Ostler TA, Barker J, Atxitia U, Chubykalo-Fesenko O, Chantrell RW (2015) The Landau-Lifshitz equation in atomistic models. Low Temp Phys 41(9):705–712. https://doi.org/10.1063/1.4930971

Evans RFL, Fan WJ, Chureemart P, Ostler TA, Ellis MOA, Chantrell RW (2014) Atomistic spin model simulations of magnetic nanomaterials. J Phys Condens Matter 26:103202

Evans RFL, Atxitia U, Chantrell RW (2015) Quantitative simulation of temperature-dependent magnetization dynamics and equilibrium properties of elemental ferromagnets. Phys Rev B 91:144425. https://doi.org/10.1103/PhysRevB.91.144425

Fidler J, Schrefl T (2000) Micromagnetic modelling – the current state of the art. J Phys D Appl Phys 33(15):R135. http://stacks.iop.org/0022-3727/33/i=15/a=201

Garanin DA (2017) Pulse-noise approach for classical spin systems. Phys Rev E 95:013306. https://doi.org/10.1103/PhysRevE.95.013306

García-Palacios JL, Lázaro FJ (1998) Langevin-dynamics study of the dynamical properties of small magnetic particles. Phys Rev B 58:14937–14958. https://doi.org/10.1103/PhysRevB.58.14937

Gerlach W, Stern O (1924) Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. Zeitschrift für Physik 9(1):349–352

Gilbert TL (1955) A Lagrangian formulation of the gyromagnetic equation of the magnetic field. Phys Rev 100:1243. https://doi.org/10.1103/PhysRevB.100.1235

Heisenberg W (1928) Zur theorie des ferromagnetismus. Z Phys 49:619

Ising E (1925) Contribution to the theory of ferromagnetism. Z Phys 31:253–258

Kirilyuk A, Kimel AV, Rasing T (2010) Ultrafast optical manipulation of magnetic order. Rev Mod Phys 82:2731–2784. https://doi.org/10.1103/RevModPhys.82.2731

Krech M, Bunker A, Landau D (1998) Fast spin dynamics algorithms for classical spin systems. Comput Phys Commun 111(1):1–13. http://www.sciencedirect.com/science/article/pii/S0010465598000095

Landau LD, Lifshitz EM (1935) Theory of the dispersion of magnetic permeability in ferromagnetic bodies. Phys Z Sowietunion 8:153

Ma PW, Woo CH, Dudarev SL (2008) Large-scale simulation of the spin-lattice dynamics in ferromagnetic iron. Phys Rev B 78:024434. https://doi.org/10.1103/PhysRevB.78.024434

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21(6):1087. https://doi.org/10.1063/1.1699114

Moriya T (1960) Anisotropic superexchange interaction and weak ferromagnetism. Phys Rev 120:91–98. https://doi.org/10.1103/PhysRev.120.91

Mryasov ON, Nowak U, Guslienko KY, Chantrell RW (2005) Temperature-dependent magnetic properties of fept: effective spin hamiltonian model. Euro Phys Lett 69(5):805. http://stacks.iop.org/0295-5075/69/i=5/a=805

Nedelkoski Z, Kepaptsoglou D, Lari L, Wen T, Booth RA, Oberdick SD, Galindo PL, Ramasse QM, Evans RFL, Majetich S et al (2017) Origin of reduced magnetization and domain formation

in small magnetite nanoparticles. Sci Rep 7. https://doi.org/10.1038/srep45997, http://www.osti.gov/pages/servlets/purl/1366451

Nowak U (2007) Classical spin models. Wiley. https://doi.org/10.1002/9780470022184.hmm205

Ostler TA, Evans RFL, Chantrell RW, Atxitia U, Chubykalo-Fesenko O, Radu I, Abrudan R, Radu F, Tsukamoto A, Itoh A, Kirilyuk A, Rasing T, Kimel A (2011) Crystallographically amorphous ferrimagnetic alloys: comparing a localized atomistic spin model with experiments. Phys Rev B 84:024407. https://doi.org/10.1103/PhysRevB.84.024407

Ostler TA, Barker J, Evans RFL, Chantrell RW, Atxitia U, Chubykalo-Fesenko O, El Moussaoui S, Le Guyader L, Mengotti E, Heyderman LJ, Nolting F, Tsukamoto A, Itoh A, Afanasiev D, Ivanov BA, Kalashnikova AM, Vahaplar K, Mentink J, Kirilyuk A, Rasing T, Kimel AV (2012) Ultrafast heating as a sufficient stimulus for magnetization reversal in a ferrimagnet. Nat Commun 3:666. http://www.nature.com/doifinder/10.1038/ncomms1666

Pan F, Chico J, Hellsvik J, Delin A, Bergman A, Bergqvist L (2016) Systematic study of magnetodynamic properties at finite temperatures in doped permalloy from first-principles calculations. Phys Rev B 94:214410. https://doi.org/10.1103/PhysRevB.94.214410

Pan F, Chico J, Delin A, Bergman A, Bergqvist L (2017) Extended spin model in atomistic simulations of alloys. Phys Rev B 95:184432. https://doi.org/10.1103/PhysRevB.95.184432

Perera D, Eisenbach M, Nicholson DM, Stocks GM, Landau DP (2016) Reinventing atomistic magnetic simulations with spin-orbit coupling. Phys Rev B 93:060402. https://doi.org/10.1103/PhysRevB.93.060402

Perera D, Nicholson DM, Eisenbach M, Stocks GM, Landau DP (2017) Collective dynamics in atomistic models with coupled translational and spin degrees of freedom. Phys Rev B 95:014431. https://doi.org/10.1103/PhysRevB.95.014431

Radu I, Vahaplar K, Stamm C, Kachel T, Pontius N, Dürr HA, Ostler TA, Barker J, Evans RFL, Chantrell RW, Tsukamoto A, Itoh A, Kirilyuk A, Rasing T, Kimel AV (2011) Transient ferromagnetic-like state mediating ultrafast reversal of antiferromagnetically coupled spins. Nature 472(7342):205–208. http://www.nature.com/doifinder/10.1038/nature09901

Simoni J, Stamenova M, Sanvito S (2017) Ultrafast demagnetizing fields from first principles. Phys Rev B 95:024412. https://doi.org/10.1103/PhysRevB.95.024412

Skomski R (2012) Simple models of magnetism. OUP, Oxford

Skubic B, Hellsvik J, Nordström L, Eriksson O (2008) A method for atomistic spin dynamics simulations: implementation and examples. J Phys Condens Matter 20(31):315203. https://doi.org/10.1088/0953-8984/20/31/315203, http://stacks.iop.org/0953-8984/20/i=31/a=315203?key=crossref.74bd3abd0f58484825bcbd4af5661ec8

Stanciu CD, Hansteen F, Kimel AV, Kirilyuk A, Tsukamoto A, Itoh A, Rasing T (2007) All-optical magnetic recording with circularly polarized light. Phys Rev Lett 99:047601. https://doi.org/10.1103/PhysRevLett.99.047601

Szunyogh L, Lazarovits B, Udvardi L, Jackson J, Nowak U (2009) Giant magnetic anisotropy of the bulk antiferromagnets IrMn and $IrMn_3$ from first principles. Phys Rev B 79:020403. https://doi.org/10.1103/PhysRevB.79.020403

Toga Y, Matsumoto M, Miyashita S, Akai H, Doi S, Miyake T, Sakuma A (2016) Monte Carlo analysis for finite-temperature magnetism of $Nd_2Fe_{14}B$ permanent magnet. Phys Rev B 94:174433. https://doi.org/10.1103/PhysRevB.94.174433

Weiss P (1907) L'hypothèse du champ moléculaire et la propriété ferromagnétique. J Phys Theor Appl 1:661–690. https://doi.org/10.1051/jphystap:019070060066100

Yanes R, Chubykalo-Fesenko O, Kachkachi H, Garanin DA, Evans R, Chantrell RW (2007) Effective anisotropies and energy barriers of magnetic nanoparticles with néel surface anisotropy. Phys Rev B 76:064416. https://doi.org/10.1103/PhysRevB.76.064416

Yanes R, Jackson J, Udvardi L, Szunyogh L, Nowak U (2013) Exchange bias driven by Dzyaloshinskii-Moriya interactions. Phys Rev Lett 111:217202. https://doi.org/10.1103/PhysRevLett.111.217202

Yang Y, Wilson RB, Gorchon J, Lambert CH, Salahuddin S, Bokor J (2017) Ultrafast magnetization reversal by picosecond electrical pulses. Sci Adv 3(11):e1603117. http://advances.sciencemag.org/content/3/11/e1603117

Check for updates

Ultra-fast Dynamics for Heat-Assisted Magnetic Recording

21

Simon J. Greaves

Contents

Abstract

This chapter describes models and techniques for the simulation of magnetization dynamics at elevated temperatures. These can be applied to heat-assisted magnetic recording systems, or other systems in which magnetic materials are heated close to or above their Curie temperature. In many cases demagnetization takes place on picosecond time scales as energy is absorbed by the electrons responsible for the magnetic moment. As such, an understanding of ultra-fast magnetization dynamics is important when modelling these systems. Models that can be used to simulate ultra-fast magnetization dynamics are presented, together with an overview of heat sources and other mechanisms that can give rise to ultra-fast dynamics.

S. J. Greaves (✉)
Research Institute of Electrical Communication, Tohoku University, Sendai, Japan
e-mail: simon@riec.tohoku.ac.jp

1   Introduction

The growth of the information society has resulted in the creation of vast amounts of data that need to be stored and processed. This has led to an insatiable demand for high density, high speed and low cost data storage systems and devices. One candidate technology that may be used in the future is heat assisted magnetic recording, which involves localized heating of a recording medium to reduce the coercivity, allowing information to be stored. If a magnetic material is heated near to its Curie temperature the magnetization dynamics can be quite different from a conventional recording system in which recording takes place at room temperature. Depending on the source and nature of the heating mechanism, the recording medium may become demagnetized on picosecond time scales and the spins, electrons and lattice atoms may need to be described by different temperatures. Hence, models that can account for ultra-fast magnetization dynamics are necessary. This chapter describes micromagnetic and atomistic models that can be used for this task.

2   Heat-Assisted Magnetic Recording

To increase the density of data stored on a magnetic disk medium, the size of the polycrystalline grains comprising the medium should be reduced. For a single-phase grain with volume V and uniaxial anisotropy K_u, the average time, τ, between random, thermally induced changes of the magnetization direction is given by

$$\frac{1}{\tau} = f_0 e^{-K_u V / kT} \tag{1}$$

where T is the temperature, f_0 is the attempt frequency, and k is the Boltzmann's constant. The attempt frequency for high energy barriers ($K_u V \gg kT$) is given by Brown (1963)

$$f_0 = \frac{\gamma \alpha}{1 + \alpha^2} \sqrt{\frac{H_k^3 M_s V}{2\pi kT}} \tag{2}$$

where γ is the gyromagnetic ratio, α the damping constant, M_s the saturation magnetization, and $H_k = 2K_u/M_s$ the anisotropy field.

Suppose the attempt frequency is 10^{10} Hz: if we wish to store data for 10 years at a temperature of 300 K, the ratio $K_u V/kT$ should be about 43 (energy barrier $\Delta E = K_u V \approx 1.78 \times 10^{-12}$ erg) to ensure that at least half of the grains in a medium retain their original magnetization. In reality, higher values of $K_u V/kT$ may be necessary to maintain a sufficiently high signal-to-noise ratio (SNR) as grains may be subjected to magnetostatic fields arising from other medium grains and external magnetic fields. If the grain volume is reduced, K_u, and therefore H_k, must be increased to compensate.

   To record on a magnetic medium, the head field should exceed the anisotropy
field of the medium. However, the recording field that can be generated by a write
head is limited by the saturation magnetization of head materials. The field near the
surface of a magnetic material with magnetization perpendicular to the surface is
$2\pi M_s$. Therefore, the maximum field in-between two parallel planes of the same
material is $4\pi M_s$. In a perpendicular recording system, the air bearing surface
(ABS) of the main pole of the write head generates a field $H_{head} = 2\pi M_s$ and a soft
magnetic underlayer generates an additional field $H_{SUL} < 2\pi M_s$. Some additional
magnetic flux may also be present from the coil driving the write head, but is ignored
here. The maximum M_s of write head materials is around $1910\,\mathrm{emu/cm^3}$ (Yun et al.
1996), giving a maximum write field of 24 kOe.
   The maximum head field places an upper limit on K_u and a lower limit on V.
More complicated media structures, such as composite or exchange spring media,
can be used to reduce the switching field of a grain, but ultimately a limit is reached
and the grain cannot be made any smaller. To overcome this problem, the concept
of heat-assisted magnetic recording (HAMR) was introduced.
   HAMR is a technology that enables recording on media with extremely high
room temperature uniaxial anisotropy. A schematic of a HAMR system is shown in
Fig. 1a. A laser locally heats a portion of the medium during recording, lowering the
anisotropy field below the head field. As the medium cools, the magnetization aligns
with the head field; once the medium returns to room temperature, the anisotropy
field is sufficiently high that the medium is thermally stable. This process is shown

(a)   (b)
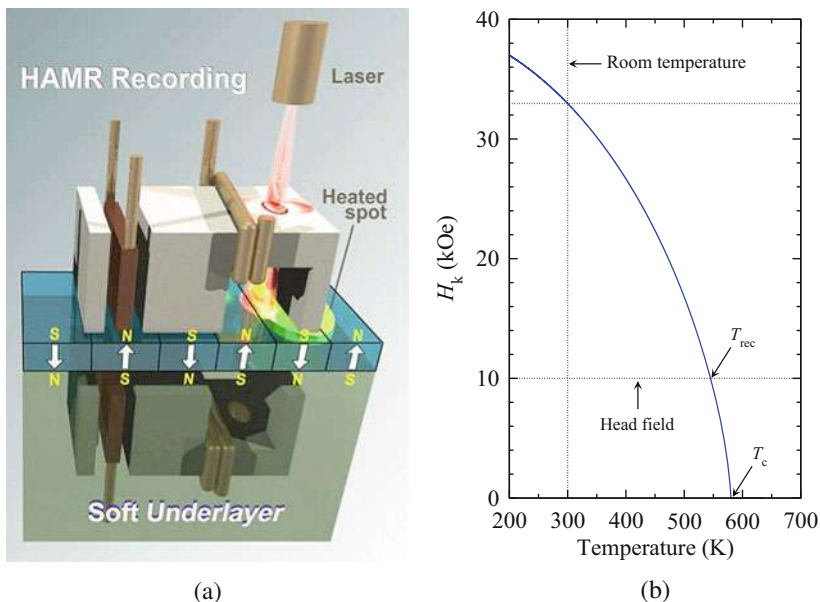
Fig. 1 Schematic of a HAMR system (left) (courtesy of Seagate) and recording process (right) (**a**)
HAMR schematic. (**b**) H_k vs. T

in Fig. 1b, in which the temperature dependence of H_k is shown for a medium with an assumed Curie temperature of 580 K. At room temperature, H_k is about 33 kOe, far higher than the assumed head field of 10 kOe. Heating the medium near to, or above, T_c reduces H_k below the head field. As the medium cools, data can be written at temperatures around T_{rec} (545 K in this example), at which point $H_k \approx H_{head}$.

Factors controlling the quality of the recorded bits and the sharpness of transitions in a HAMR system are the thermal gradient in the medium, dT/dy; the rate of change of the anisotropy field in the vicinity of the recording temperature, dH_k/dT; and the head field gradient, dH/dy, where y is the down-track axis. These parameters combine to give an effective head field gradient of

$$\frac{dH_{eff}}{dy} = \frac{dH_k}{dT}\frac{dT}{dy} + \frac{dH}{dy} \tag{3}$$

Head field gradients are limited by geometrical considerations to around 500 Oe/nm, whereas thermal gradients of 5 K/nm and dH_k/dT of 200 Oe/K are easily obtained. As a result the head field gradient becomes less important in a HAMR system and the recording performance is dominated by the thermal gradient and dH_k/dT.

Typical recording media used in HAMR systems are FePt alloys (Weller et al. 2013, 2016). These are favored for their extremely high uniaxial anisotropy, theoretically allowing grains with diameters of 3–4 nm to be thermally stable at room temperature for many years.

The above discussion relates to conventional recording systems, but in recent years, alternative methods of switching the magnetization direction have been proposed which do not rely on an external magnetic field. For example, in all-optical recording, circularly polarized light is used to determine the direction of magnetization switching based on the inverse Faraday effect (Stanciu et al. 2007). In other experiments multiple light pulses were found to have a cumulative effect on the magnetization switching in a process ascribed to magnetic circular dichroism (El Hadri et al. 2016). Electrical pulses may also be used to demagnetize materials on ultrashort time scales (Wilson et al. 2017).

Whatever the nature of the heat source, the approach to modelling the magnetization dynamics is the same: a micromagnetic or atomistic model is used to model the magnetization, with the stimulus provided by the heat included as an additional magnetic field term. The difficulty lies in correctly accounting for the temperature dependence of parameters, such as M_s, K_u, and most other inputs to the model, as well as accurately calculating the temperature itself. The remainder of this section deals with the models used to describe magnetization dynamics, while Sect. 3 considers how to treat the heat absorption.

2.1 The LLG Equation

Standard micromagnetic simulations use the Landau-Lifshitz-Gilbert (LLG) equation to model the change in direction of a magnetization vector **M** when subject to a magnetic field **H**:

$$\frac{d\mathbf{M}}{dt} = -\gamma \left(\mathbf{M} \times \left(\mathbf{H} - \frac{\alpha}{\gamma M_s} \frac{d\mathbf{M}}{dt} \right) \right) \tag{4}$$

$\mathbf{M}$ can represent the magnetization of a single spin or the average magnetization of a discrete volume (cell) containing many spins (a macrospin). When the LLG equation is applied to cells containing many spins, a temperature-dependent function, $M_s(T)$, must be introduced to reflect the reduction of M_s at elevated temperatures and $M_s(T)$ should reach zero at the Curie temperature, T_c.

However, if $M_s(T)$ becomes zero, there would be a singularity in Eq. 4. Strictly speaking, M_s represents the average magnetization of a number of atoms whose magnetic moment is independent of temperature. As such, it could be argued that although the time-averaged magnetization is zero at, or above, the Curie temperature, the instantaneous magnetization is small, but nonzero. Therefore it is sufficient for $M_s(T)$ to be a small, nonzero value at $T \geq T_c$.

The field $\mathbf{H}$ in the LLG equation is not just the external magnetic field applied by the write head; it also includes the magnetostatic, anisotropy, exchange coupling, and thermal fields. A detailed discussion of each of these terms can be found in Greaves (2007). The latter four terms are functions of M_s and are therefore temperature dependent, while the anisotropy field includes a K_u term which also varies with temperature.

The thermal field is used to represent the effect of thermal energy on the magnetization. It is a random value chosen from a Gaussian distribution with an average value of zero and a standard deviation of

$$\sigma = \sqrt{\frac{2k\alpha T}{M_s V \gamma dt}} \tag{5}$$

The thermal field is uncorrelated in space and time, and new values are generated at each point every time the LLG equation is evaluated. The effect of the thermal field is to impart a randomness to the motion of $\mathbf{M}$. If $\mathbf{M}$ represents the magnetization of a single domain particle with uniaxial anisotropy, then at a temperature of zero Kelvin, $\mathbf{M}$ will lie along the easy axis. As the temperature is increased from zero Kelvin, the thermal field will cause $\mathbf{M}$ to fluctuate about the easy axis such that multiple samples of the angle between $\mathbf{M}$ and the easy axis will form a Maxwell-Boltzmann distribution. Interestingly, this also means that at nonzero temperatures, $\mathbf{M}$ can never be perfectly aligned with the easy axis.

2.2 Temperature-Dependent Parameters

The temperature dependence of various physical parameters must be considered in heat-assisted recording models. The form of $M_s(T)$ can be calculated using the Weiss theory and an appropriate value for J (Weiss 1907). The parameter J is used in the Brillouin function $B(\alpha, J)$ to find the value of $x \equiv M_s(T)/M_s(0)$ that satisfies the relation
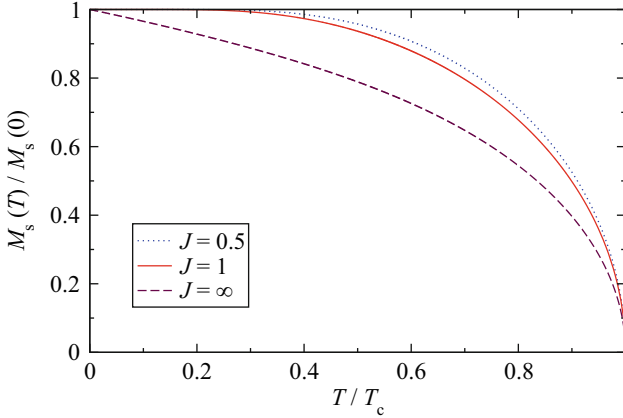
Fig. 2 Reduced saturation magnetization vs. T/T_c. J is the Brillouin function parameter used in the Weiss theory of ferromagnetism

$$x = B\left(\frac{3JT_c x}{T(J+1)}, J\right).$$ (6)

Figure 2 shows the temperature dependence of M_s calculated using the Weiss theory of ferromagnetism. Materials such as Fe and Co have a temperature dependence of M_s which lies close to the curves with $J = 0.5$ and $J = 1$.

Alternatively, an analytical function for $M_s(T)$ may be used, with the advantage that a better fit to the experimental data may be obtained. For example, Kuzmin (Kuzmin 2005) proposed a function of the form

$$\frac{M_s(T)}{M_s(0)} = \left(1 - s\left(\frac{T}{T_c}\right)^{3/2} - (1-s)\left(\frac{T}{T_c}\right)^p\right)^{1/3}$$ (7)

where s and p are fitting parameters. For many materials, $p = 2.5$ and various values of s were found to give good agreement with experimental data. In the case of Fe, $p = 4$ was better.

Having established the temperature dependence of M_s, the uniaxial anisotropy can often be assumed to follow the relationship

$$\frac{K_u(T)}{K_u(0)} = \left(\frac{M_s(T)}{M_s(0)}\right)^n$$ (8)

where the power n has been measured in experiments to be $n = 2$ for FePt and FeNiPt alloys (Thiele et al. 2002) and $n = 3$ for CoCr alloy thin films (Inaba et al. 2000). For pure ferromagnetic elements, much larger values of n are appropriate (Zener 1954; Carr 1958), but these do not make good recording media.

Exchange coupling is also a temperature-dependent parameter. In a model in which cells are used to represent the magnetization of individual grains in a recording medium, the temperature dependence of the exchange coupling energy density can be described by a linearly decreasing function (Huang and Kryder 2014).

Finally, the damping constant can also vary with temperature. The microscopic origins of damping are not fully understood, and many different mechanisms may contribute in larger or smaller amounts to the overall damping. In general, scattering of conduction electrons by the magnetization and defects are believed to make the largest contributions (Kambersky 2007). The temperature dependence of these effects is difficult to calculate, even for pure metals, and can vary greatly in alloys with different compositions. Examples of measurements on pure ferromagnetic metals such as nickel (Bhagat and Chicklis 1969) and iron (Cochran et al. 1991) show the damping constant decreasing with temperature in the former and increasing with temperature in the latter. Note that the damping constant in these papers, λ, can be related to the damping constant in the LLG equation, α, by $\lambda = \alpha \gamma M_s$. Measurements of the temperature dependence of α can also vary significantly depending on the shape of the sample and the measurement technique used. Given these difficulties, determining the effect of α at high temperatures is perhaps best approached by assuming a simple, linear, temperature-dependent function $\alpha(T)$ which will at least allow the modeller to decide if $\alpha(T)$ has a significant effect on the system being simulated.

2.3   Atomistic Models

If a micromagnetic model is one in which $\mathbf{M}$ represents the average magnetization of a number of atoms, an atomistic model is one in which every (magnetic) atom is represented by an individual magnetic moment or spin $\mathbf{S}$. Because the number of atoms is large in even a small object, e.g., a 4.4 nm cube with a 0.2 nm atomic spacing contains 10,648 atoms, useful calculations are currently limited to single grains, collections of grains, or smaller objects. As computing power increases (and the size of magnetic devices decreases), solving larger problems and modelling of entire devices become more tractable. If the atoms can be assumed to lie on a regular lattice, graphics processing units (GPUs) can be used to accelerate computations, in particular the magnetostatic field calculation.

In an atomistic model, the LLG equation is applied to each spin individually, with $\mathbf{M}$ in Eq. 4 replaced by $\mathbf{S}$ and $M_s V$ replaced by the magnetic moment of the atom. The total magnetization is then given by $\mathbf{M} = \sum_i \mathbf{S_i}$. The exchange coupling calculation may be extended to second and third (or more) nearest neighbors, depending on the nature of the material being modelled.

In atomistic models the magnetic moments are generally assumed to be independent of temperature, with $M_s(T)$ and the Curie temperature being determined by the strength of the exchange coupling interaction and the atomic spacing. At

low temperatures, the exchange coupling field causes the magnetic moments of neighboring atoms in a ferromagnet to align in the same direction. As the temperature is increased, the thermal energy, represented in the model by the thermal field, causes the magnetic moments to become disordered, reducing the spontaneous magnetization. At the Curie temperature, the time-averaged magnetization along any axis reaches zero, although the magnetization at any particular instant is likely to be nonzero as $\sum_i \mathbf{S_i}$ along any particular direction will rarely be exactly zero.

At low temperatures magnetization reversal proceeds via precessional rotation of the magnetization, as in the micromagnetic model. Nearer to the Curie temperature, the reversal mode undergoes a transition to a linear mode. The magnetic moments of individual atoms continue to precess, governed as they are by the LLG equation, but the loss of coherency at high temperatures means that the total magnetization, $\mathbf{M}$, can switch from one direction to the opposite direction without developing a large orthogonal magnetization component (Barker et al. 2010). In this respect the atomistic model behaves quite differently to the micromagnetic model. Linear magnetization reversal can be much faster than precessional reversal, and therefore, where ultra-fast dynamics are concerned, it is important that the model be able to represent the linear reversal mode.

A common problem with atomistic models is that the calculated variation of M_s with temperature follows the $J = \infty$ curve in Fig. 2, which is not representative of materials such as Co or Fe. This discrepancy arises due to the way in which the thermal field is used. As noted in Sect. 2.1, the magnon energies represented by the thermal fields are chosen from a Gaussian distribution and have a constant power spectral density, i.e., the thermal energies form a white noise spectrum. As a result high energy magnon excitations are over represented at low temperatures, leading to lower $M_s(T)$ than in real materials. One solution to this problem is to use colored noise (Woo et al. 2015), although this is nontrivial to generate.

2.4   The LLB Equation

The use of atomistic models is limited by the sheer number of atoms that are present in all but the smallest objects. An alternative way to model high temperature magnetization dynamics is to use the Landau-Lifshitz-Bloch (LLB) equation. First introduced by Garanin (1997), the LLB equation can be used as a replacement for the LLG equation at high temperatures, allowing larger volumes, e.g., polycrystalline grains, to be treated as one discrete object described by an average magnetization vector.

Various forms of the LLB equation exist: Evans et al. (2012), Tzoufras and Grobis (2015), and Xu and Zhang (2012). All of them have in common the fact that at low temperatures, they reduce to the LLG equation, and at high temperatures, an additional term provides for longitudinal magnetization fluctuations. To take an example from the first of these three papers, the LLB equation is given as
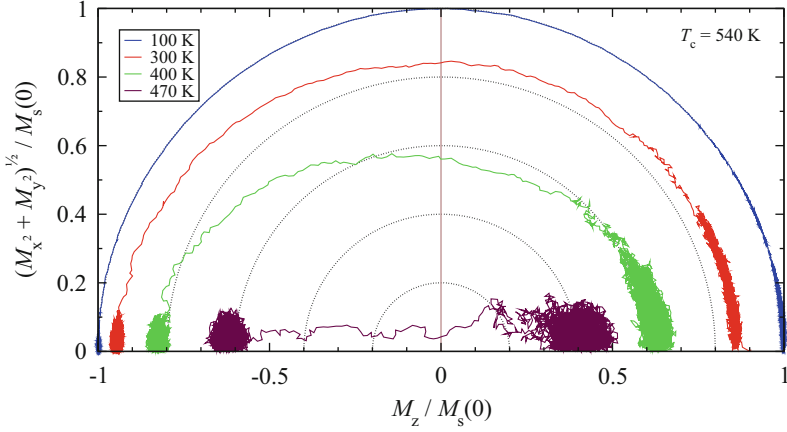
Fig. 3 Magnetization reversal of an 8 nm cube at various temperatures. (Data from Greaves et al. 2015)

$$\frac{d\mathbf{m}}{dt} = -\gamma\,(\mathbf{m} \times \mathbf{H}) + \frac{\gamma\alpha_\parallel}{m^2}\,(\mathbf{m} \cdot \mathbf{H})\,\mathbf{m} - \frac{\gamma\alpha_\perp}{m^2}\,(\mathbf{m} \times (\mathbf{m} \times (\mathbf{H} + \boldsymbol{\eta}_\perp))) + \boldsymbol{\eta}_\parallel \qquad (9)$$

where the η terms represent random, thermal fields and the α terms are temperature-dependent damping parameters. At $T = 0\,\mathrm{K}$ $\alpha_\parallel$ is equal to zero and Eq. 9 is basically the same as Eq. 4, albeit expressed in a slightly different form.

Another difference between the LLB and LLG equations is that in the LLB equation, the magnetization vector $\mathbf{m}$ does not have a constant length. The average length of $\mathbf{m}$ is equal to $M_s(T)/M_s(0)$, but the actual length fluctuates from one time step to the next due to the thermal fields. An additional "entropic" field term is added to $\mathbf{H}$ to govern the length of $\mathbf{m}$, such that the entropic field is negative when $|\mathbf{m}| > M_s(T)/M_s(0)$, positive when $|\mathbf{m}| < M_s(T)/M_s(0)$, and zero when $\mathbf{m}$ is the correct length.

The second term on the right hand side of Eq. 9 gives rise to longitudinal magnetization fluctuations at high temperatures, with $\alpha_\parallel = \alpha\frac{2T}{3T_c}$. The effect of these fluctuations can be seen in Fig. 3, which shows how the magnetization of an 8 nm cube reverses from the $+z$ direction to the $-z$ direction during the calculation of a hysteresis loop. At a temperature of $100\,\mathrm{K}$ the magnetization reverses via precessional motion about the z axis, with the magnetization vector lying on the surface of a sphere. Thus, when M_z is small, the in-plane component of M is large. As the temperature is increased, a gradual transformation from precessional to linear reversal takes place (Barker et al. 2010). At 470 K the magnetization changes from $+z$ to $-z$ while expressing only a very small in-plane magnetization component, i.e., the reversal is linear.

The original form of the LLB equation proposed by Garanin supposed that the temperature dependence of M_s followed the $J = \infty$ curve in Fig. 2, and this is

reflected in the formulation of the entropic field equations. To model materials with different values of J, the expressions in Greaves et al. (2015) can be used to resolve the discontinuity in the entropic field at T_c inherent in Garanin's approach when $J \neq \infty$. Alternatively, a quantum LLB equation has been proposed (Nieves et al. 2014) to achieve a similar result. The form of the quantum LLB equation is the same as Eq. 9, with changes to the expressions for the damping constants and entropic field.

One potential problem with the LLB equation is that it cannot be used in an atomistic model as the length of **m** is not constant. Therefore, there must be a certain cell volume below which the LLB equation becomes invalid. However, even a 1 nm^3 cell can contain in excess of 100 atoms, and the use of the LLB equation should not be a problem at these scales, the exception being materials such as dilute magnetic alloys in which the number of magnetic spins in a given volume is relatively small.

2.5 Other Models

Alternative models for high temperature magnetization dynamics should also be mentioned. One approach is to renormalize the temperature-dependent parameters, such as M_s, K_u, α and the exchange coupling constant A, of a block of atoms to match the values obtained from an atomistic model of the same-sized block (Victora and Huang 2013). The block size is typically 1–2 nm, i.e., smaller than the grain size in a recording medium, but allowing much faster calculations than an atomistic model.

3 Heat Transport

The heat flow equation

$$\frac{dT}{dt} = \frac{K}{C}\nabla^2 T + \frac{P(t)}{CV} \tag{10}$$

can be used to calculate the change of temperature of an object over time. Like the LLG and LLB equations, the heat flow equation shares the differential operator d/dt, and temperature distributions can be calculated in parallel with the LLG or LLB equations, although it may be necessary to use a different time step. In Eq. 10 K is the thermal conductivity and C is the specific heat capacity. The last term in the equation is used when heat is added to the medium by an external source, e.g., a laser or near field transducer. In that case $P(t)$ is the energy delivered by the heat source and V the cell volume into which the energy is absorbed. In general $P(t)$ has a spatial distribution, e.g., the laser spot might have a Gaussian intensity profile on the surface of the medium and the laser intensity would decay with depth as the light is absorbed.

   The intensity of light at a depth z within a medium can be calculated using the extinction coefficient of the medium, k_e, as

$$I(z, t) = (1 - R)I(0, t)e^{-4\pi k_e z/\lambda} \tag{11}$$

where λ is the wavelength of the light, R is the bulk reflectivity, and $I(0)$ is the intensity of incident light on the surface of the medium. Both R and k_e are functions of λ and k_e is the complex part of the refractive index. Equation 11 can be integrated over the thicknesses of the layers in the model to determine the total energy absorbed by each layer, giving $P(t)$ in Eq. 10. For a commonly used magnetic material such as Co, the extinction coefficient for $\lambda = 600$ nm is about 4, meaning that 55% of incident energy is absorbed within the top 10 nm of a medium and 90% within the top 27 nm. As a result a large vertical temperature gradient can exist within a medium.

   Both the thermal conductivity, K, and the specific heat, C, are functions of temperature. In the case of pure cobalt and iron, the thermal conductivity decreases as the temperature is increased, although in the case of alloys of Co and Fe, the rate of reduction can be much reduced, or even change sign (Terada et al. 2003; Shelton 1934). As data are difficult to obtain, particularly for more complicated alloys, the temperature dependence of bulk thermal conductivity is often omitted from models and the same is true for specific heat capacities. There is some evidence that the specific heat capacity can also depend on the direction of the magnetization and that the temperature dependence of the specific heat capacity is itself related to the uniaxial anisotropy, albeit very weakly (Lanchester et al. 1975).

   Modern, granular recording media consist of magnetic grains segregated by nonmagnetic oxides such as SiO_2 which have a relatively low thermal conductivity. As a result thermal conduction through the depth of the film is likely to be higher than in the film plane, and the model should take account of this anisotropic thermal conductivity. Experiments using FePt, a common material for heat-assisted recording media, have shown that the ratio of the vertical to in-plane thermal conductivities could be as high as 10:1, depending on the FePt grain size (Ho et al. 2013). The use of heat sink layers underneath the recording layer may also increase the ratio of vertical to in-plane thermal conductivity.

3.1 Two- and Three-Temperature Models

For the study of ultra-fast dynamics, it is not sufficient to calculate only the temperature of the bulk, or lattice, as effects such as laser-induced demagnetization take place on a much shorter time scale than changes in the lattice temperature. A two-temperature model can be used to represent the different temperatures of the electrons and lattice (Anisimov et al. 1974). The electrons absorb the heat entering the system, and because they have a much smaller specific heat than the lattice, their temperature increases much more rapidly. Subsequently, heat is transferred from the electrons to the lattice and the two temperatures converge over time. Experimental

Fig. 4 Heat flow between electrons, spins, and lattice atoms in the three-temperature model. $P(t)$ represents the heat from a laser or near field light source
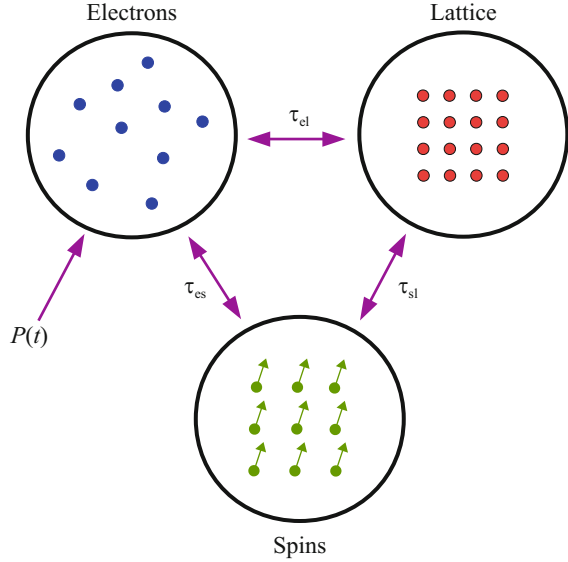
measurements on metals find the two-temperature model gives a good description of the behavior of many metals (Hohlfeld et al. 2000), but in magnetic materials, the temperature of the spins must also be considered.

In the three-temperature model, separate temperatures are ascribed to the electrons, spins, and the lattice atoms (Beaurepaire et al. 1996). Each component has a different (temperature dependent) specific heat, and coupled differential equations describe the heat flow between the three components. A schematic of the model is shown in Fig. 4 with the constants τ_{el}, τ_{es}, and τ_{sl} determining the rate of energy transfer between the three components and $P(t)$ representing the heat entering the system.

Equation 12 shows the equations used to calculate the temperatures of the electrons, spins, and lattice. If necessary, further terms can be added to account for thermal conduction out of the heated region, although such conduction may take place on much longer time scales than the picosecond dynamics of the spins and electrons and could be neglected in many cases:

$$\frac{dT_e}{dt} = \frac{-\tau_{el}(T_e - T_l) - \tau_{es}(T_e - T_s) + P(t)}{C_e(T_e)}$$

$$\frac{dT_l}{dt} = \frac{-\tau_{el}(T_l - T_e) - \tau_{sl}(T_l - T_s)}{C_l(T_l)}$$

$$\frac{dT_s}{dt} = \frac{-\tau_{es}(T_s - T_e) - \tau_{sl}(T_s - T_l)}{C_s(T_s)} \tag{12}$$

Many experiments have shown that energy transfer between electrons and spins can be extremely rapid, such that the spin temperature lags the electron temperature by $\approx$10–100 fs. In such materials Eq. 12 can be simplified to van Kampen (2003)

$$\frac{dT_e}{dt} = \frac{-\tau_{el}(T_e - T_l) + P(t)}{C_e(T_e)}$$

$$\frac{dT_l}{dt} = \frac{-\tau_{el}(T_l - T_e)}{C_l(T_l)}$$

$$\frac{dT_s}{dt} = \frac{(T_e - T_s)}{t_d} \tag{13}$$

where t_d is the delay between the electron and spin temperatures and Eq. 13 is basically the same as the two-temperature model. This approach has been used in an atomistic model of spin dynamics to calculate the evolution of the magnetic moment over femtosecond to nanosecond time scales (Chimata et al. 2012).

Experimental measurements of the specific heat of, e.g., nickel, show a roughly linear variation of specific heat over the temperature range from 300 to 600 K (Grew 1934), with a sharp peak around the Curie temperature at 631 K. To model this behavior, as a first approximation, the specific heat of the lattice can be assumed to be constant, with the electronic specific heat assumed to vary as

$$C_e(T_e) = \gamma T_e \tag{14}$$

and the total specific heat being the sum of the two. The γ parameter can be used as a fitting parameter, with values in the literature for nickel ranging from 1.1×10^3 to 6×10^3 J/m^{3}K^2. Another fitting parameter in the simplified three-temperature model is τ_{el}, which can be of the order 1×10^{18} W/m^{3}K. The parameter controls the rate of decay of the electron temperature. In the full three-temperature model, the other coupling constants were estimated by Beaurepaire et al. (1996) to be $\tau_{es} = 6 \times 10^{17}$ W/m^{3}K and $\tau_{sl} = 0.3 \times 10^{17}$ W/m^{3}K.

3.2 Heat Sources

The nature of the incident energy source used to heat the medium may also need to be considered in the model. In many pump-probe experiments, the medium is heated by an ultra-short laser pulse lasting for less than 100 fs, with the size of the illuminated area measured in tens of microns. Given that micromagnetic models of these experiments are generally limited to simulating regions of $\leq 1\mu$m^2, the model can be simplified by assuming a uniform distribution of the incident laser power $I(0, t)$ over the surface of the medium. The time dependence of the laser pulse intensity on the surface of the medium can be represented by a Gaussian of the form

$$I(0, t) = I_{\max} e^{-(\frac{t - t_{\max}}{t_p})^2} \tag{15}$$

where I_{max} is the maximum intensity, t the time, t_{max} the time at which the pulse reaches its maximum intensity, and t_p the pulse duration. As the laser intensity is uniform over the modelled region, there may be no thermal gradient in the plane of the medium and hence no thermal conduction. A thermal gradient in the vertical direction will still exist and thermal conduction into the air and substrate should be considered.

3.2.1 Near Field Transducers

In modern recording systems, the size of written bits, measured in the tens of nanometers, is much smaller than the minimum (diffraction limited) size of an optical spot that can be obtained using lenses. Instead, near field transducers (NFT) are used to generate the extremely small heat spots used in heat-assisted magnetic recording. Such NFTs usually consist of a thin metallic film, often gold, in which surface plasmons are induced by incident light from a laser or waveguide. Depending on the geometry of the NFT, the plasmon energy can be concentrated into a small area and coupled into an adjacent recording medium.

NFT designs have been studied in detail, with common shapes being "bow tie," "E," "nanobeak," and "lollipop" (Zhou et al. 2014). Of these the most common design reported in the literature as being used in actual HAMR systems is the "lollipop" (Challener et al. 2009). The energy distribution in and around a NFT can be calculated using a finite-difference time-domain (FDTD) model, such as the free software MEEP (Oskooi et al. 2010). As a first approximation, the recording medium can be assumed to consist of homogeneous strata, but a more detailed analysis would also take account of the granular structure of most recording media, with different absorption and thermal conductivity characteristics assigned to the magnetic grains and nonmagnetic grain boundaries (Nakagawa et al. 2007). Such analyses show energy absorption being concentrated in the grains, raising their temperature by more than would be expected from a uniform layer model. There is also recent evidence that the shape of the grains themselves can lead to localized hot spots, for example, in the vicinity of protrusions or locations around the edge of a grain with a smaller radius of curvature relative to the rest of the grain.

3.2.2 Alternative Heating Mechanisms

The three-temperature model provides a good representation of how energy from an incident laser pulse is absorbed by a magnetic material. But in recent years experiments have revealed alternative heating mechanisms that may induce ultra-fast magnetization dynamics, and these are briefly discussed here.

Magnetic circular dichroism is a phenomenon in which the absorption (and reflectivity) of left and right circularly polarized light by a magnetic material differs depending on the orientation of the material's magnetization. This means that if a multi-domain medium with perpendicular anisotropy is subjected to a circularly polarized light pulse, the "up" domains will absorb more energy than

the "down" domains, or vice versa, depending on the polarization of the light. As a result one set of domains will be heated more than the other, creating a temperature difference between the two. The temperature difference creates a higher probability of thermally induced magnetization reversal for one of the magnetization orientations, such that after multiple light pulses, the area or number of the hotter domains will shrink and that of the cooler domains grow. A simple model for this process is described in Gorchon et al. (2016). To include the effect in a dynamic model, a magnetization-dependent absorption term of the form

$$P(t) = A_0 + \Delta A \times m_z/|\mathbf{m}| \tag{16}$$

could be used. Here, A_0 is the light energy absorbed by the medium, obtained by integrating Eq. 11 over the medium thickness; $m_z/|\mathbf{m}|$ is the normalized, vertical component of magnetization; and ΔA represents the magnetic circular dichroism. ΔA is material dependent, but the results in Gorchon et al. (2016) suggest that ΔA of 0.5% is enough to result in all optical switching after a sufficiently large number of light pulses.

The Faraday effect (FE) is the action of a magnetic material in rotating the plane of polarization of light passing through the material. Similarly, the inverse Faraday effect (IFE) is the induction of magnetization in a material by circularly polarized light. Expressions have been derived for the magnitude of the IFE in Battiato et al. (2014) using a quantum mechanical approach. The method has been applied to a model of magnetization switching in FePt media in John et al. (2017). In that model the IFE was implemented as an additional contribution to the magnetization with a characteristic decay time slightly longer than the duration of the light pulse and an initial magnitude of 3–7% of M_s. The initial magnitude of the IFE depends on the light intensity, wavelength, helicity, and the magnetic material.

Finally, electrical pulses can also be used to induce ultra-fast magnetization dynamics. An electric current passing through a resistive material will cause Joule heating, and this may result in demagnetization of a sample if the Curie temperature is reached. But it is also possible to demagnetize a sample using ultra-short duration electric current pulses, where Joule heating of the bulk sample is negligible. The experiments described by Wilson et al. (2017) compare magnetization dynamics in samples subjected to ultra-short (ps) light and electric current pulses of similar energy. In both cases the magnetization dynamics could be described by the three-temperature model, but the rate of energy transfer between the electrons and the lattice, τ_{el}, had to be about twice as large when electric current pulses were used, compared with light pulses, in order to fit the experimental results. The reason for the difference was suggested to be due to how energy is transferred into the medium: an electric current pulse excites many electrons by a small amount, whereas a light pulse excites relatively few electrons by a large amount. The resulting difference in the distribution of electron energies governs the rate of energy transfer τ_{el}.

4   Conclusions

At first glance, creating a model of heat-assisted magnetic recording may appear daunting due to the large number of parameters that must be considered and the fact that nearly all of them are temperature dependent. Indeed, it is often difficult to find appropriate values for many commonly used materials in the literature, and then there is the question as to whether bulk values are representative of thin films. However, on ultra-fast time scales, it is often the case that most of the parameters only have a small influence on the result, meaning that knowledge of their precise values is unnecessary. Magnetization dynamics for macrospin cells can be calculated using the LLB model, while the LLG model is suitable for atomistic calculations. Depending on the size of the illuminated region, thermal conduction calculations may not be needed, and it may be sufficient to assume a uniform temperature distribution. Despite these approximations, many experimental results have been successfully described using micromagnetic and atomistic models, and once a basic model has been implemented, it is not too difficult to add support for more exotic phenomena, such as the inverse Faraday effect, or magnetic circular dichroism, as needed.

References

Anisimov SI, Kapeliovich BL, Perelman TL (1974) Electron emission from metal surfaces exposed to ultrashort laser pulses. Sov Phys JETP 39:375–377

Barker J, Evans RFL, Chantrell RW, Hinzke D, Nowak U (2010) Atomistic spin model simulation of magnetic reversal modes near the Curie point. Appl Phys Lett 97:192504–1–3

Battiato M, Barbalinardo G, Oppeneer PM (2014) Quantum theory of the inverse Faraday effect. Phys Rev B 89:014413–1–9

Beaurepaire E, Merle JC, Daunois A, Bigot JY (1996) Ultrafast spin dynamics in ferromagnetic nickel. Phys Rev Lett 76:4250–4253

Bhagat SM, Chicklis EP (1969) Ferromagnetic resonance in nickel at high temperatures. Phys Rev 178:828–832

Brown WF (1963) Thermal fluctuations of a single domain particle. Phys Rev 130:1677–1686

Carr WJ (1958) Temperature dependence of ferromagnetic anisotropy. Phys Rev 109:1971–1976

Challener WA, Peng C, Itagi AV, Karns D, Peng W, Peng Y, Yang X, Zhu X, Gokemeijer NJ, Hsia YT, Ju G, Rottmayer RE, Seigler MA, Gage EC (2009) Heat-assisted magnetic recording by a near-field transducer with efficient optical energy transfer. Nat Photonics 3:220–224

Chimata R, Bergman A, Bergqvist L, Sanyal B, Eriksson O (2012) Microscopic model for ultrafast remagnetization dynamics. Phys Rev Lett 109:157201–1–5

Cochran JF, Rudd JM, Muir WB, Trayling G, Heinrich B (1991) Temperature dependence of the Landau-Lifshitz damping parameter for iron. J Appl Phys 70:6545–6547

El Hadri MS, Pirro P, Lambert CH, Bergeard N, Petit-Watelot S, Hehn M, Malinowski G, Montaigne F, Quessab Y, Medapalli R, Fullerton EE, Mangin S (2016) Electrical characterization of all-optical helicity-dependent switching in ferromagnetic Hall crosses. Appl Phys Lett 108:092405–1–5

Evans RFL, Hinzke D, Atxitia U, Nowak U, Chantrell RW, Chubykalo-Fesenko O (2012) Stochastic form of the Landau-Lifshitz-Bloch equation. Phys Rev B 85:014433–1–9

Garanin DA (1997) Fokker-Planck and Landau-Lifshitz-Bloch equations for classical ferromagnets. Phys Rev B 55:3050–3057

Gorchon J, Yang Y, Bokor J (2016) Model for multishot all-thermal all-optical switching in ferromagnets. Phys Rev B 94:020409–1–5

Greaves SJ (2007) Micromagnetic simulations of magnetic recording media. In: Roller S, Lammers P, Furui T, Galle M, Bez W. High performance computing on vector systems. Springer, Berlin, pp 229–244

Greaves SJ, Muraoka H, Kanai Y (2015) Modelling of heat assisted magnetic recording with the Landau-Lifshitz-Bloch equation and Brillouin functions. J Appl Phys 117:17C505–1–4

Grew KE (1934) The specific heat of nickel and of some nickel-copper alloys. Proc R Soc A 145:509–522

Ho H, Sharma AA, Ong WL, Malen JA, Bain JA, Zhu JG (2013) Experimental estimates of in-plane thermal conductivity in FePt-C granular thin film heat assisted magnetic recording media using a model layered system. Appl Phys Lett 103:131907–1–4

Hohlfeld J, Wellershoff SS, Güdde J, Conrad U, Jähnke V, Matthias E (2000) Electron and lattice dynamics following optical excitation of metals. Chem Phys 251:237–258

Huang EY, Kryder MH (2014) Effects of temperature on intergranular exchange coupling in L1$_0$ FePt thin films. J Appl Phys 115:213906–1–5

Inaba N, Uesaka Y, Futamoto M (2000) Compositional and temperature dependence of basic magnetic properties of CoCr alloy thin films. IEEE Trans Magn 36:54–60

John R, Berritta M, Hinzke D, Müller C, Santos T, Ulrichs H, Nieves P, Walowski J, Mondal R, Chubykalo-Fesenko O, McCord J, Oppeneer PM, Nowak U, Münzenberg M (2017) Magnetisation switching of FePt nanoparticle recording medium by femtosecond laser pulses. Sci Rep 7:4114–1–8

Kambersky V (2007) Shape-orbital Gilbert damping in common magnetic metals. Phys Rev B 76:134416–1–10

Kuzmin MD (2005) Shape of temperature dependence of spontaneous magnetization of ferromagnets: quantitative analysis. Phys Rev Lett 94:107204–1–4

Lanchester PC, Whitehead NF, Wells P, Scurlock RG (1975) The low temperature specific heat of hexagonal cobalt in a magnetic field. J Phys F 5:247–254

Nakagawa K, Kim J, Itoh A (2007) Finite difference time domain simulation on near-field optics for granular recording media in hybrid recording. J Appl Phys 101:09H504–1–3

Nieves P, Serantes D, Atxitia U, Chubykalo-Fesenko O (2014) Quantum Landau-Lifshiz-Bloch equation and its comparison with the classical case. Phys Rev B 90:104428–1–11

Oskooi AF, Roundy D, Ibanescu M, Bermel P, Joannopoulos JD, Johnson SG (2010) MEEP: a flexible free-software package for electromagnetic simulations by the FDTD method. Comput Phys Commun 181:687–702

Shelton SM (1934) Thermal conductivity of some irons and steels over the temperature range 100 to 500 C. Bur Stan J Res 12:441–450

Stanciu CD, Hansteen F, Kimel AV, Kirilyuk A, Tsukamoto A, Itoh A, Rasing T (2007) All-optical magnetic recording with circularly polarized light. Phys Rev Lett 99:047601–1–4

Terada Y, Ohkubo K, Mohri T, Suzuki T (2003) Thermal conductivity of cobalt-base alloys. Metall Mater Trans A 34:2026–2028

Thiele JU, Coffey KR, Toney MF, Hedstrom JA, Kellock AJ (2002) Temperature dependent magnetic properties of highly chemically ordered Fe$_{55-x}$Ni$_x$Pt$_{45}$ L1$_0$ films. J Appl Phys 91:6595–6600

Tzoufras M, Grobis MK (2015) Dynamics of single-domain magnetic particles at elevated temperatures. New J Phys 17:103014–1–6

van Kampen M (2003) Ultrafast spin dynamics in ferromagnetic metals. Ph.D. Thesis, TU Eindhoven, p 51

Victora RH, Huang PW (2013) Simulation of heat-assisted magnetic recording using renormalized media cells. IEEE Trans Magn 49:751–757

Weiss P (1907) L'hypothèse du champ moléculaire et la propriété ferromagnétique. J Phys Theor Appl 6:661–690

Weller D, Mosendz O, Parker G, Pisana S, Santos TS (2013) L1$_0$ FePtX-Y media for heat-assisted magnetic recording. Phys Stat Solidi A 210:1245–1260

Weller D, Parker G, Mosendz O, Lyberatos A, Mitin D, Safonova NY, Albrecht M (2016) Review article: FePt heat assisted magnetic recording media. J Vac Sci Technol B 34:060801–1–10

Wilson RB, Yang Y, Gorchon J, Lambert CH, Salahuddin S, Bokor J (2017) Electric current induced ultrafast demagnetization. Phys Rev B 96:045105–1–9

Woo CH, Wen H, Semenov AA, Dudarev SL, Ma PW (2015) Quantum heat bath for spin-lattice dynamics. Phys Rev B 91:104306–1–12

Xu L, Zhang S (2012) Magnetization dynamics at elevated temperatures. Physica E 45:72–76

Yun EJ, Win W, Walser RM (1996) Magnetic properties of RF diode sputtered Co_xFe_{100-x} alloy thin films. IEEE Trans Magn 32:4535–4537

Zener C (1954) Classical theory of the temperature dependence of magnetic anisotropy energy. Phys Rev 96:1335–1337

Zhou N, Xu X, Hammack AT, Stipe BC, Gao K, Scholz W, Gage EC (2014) Plasmonic near-field transducer for heat-assisted magnetic recording. Nanophotonics 3:141–155

Magnetic Impurities on Surfaces: Kondo and Inelastic Scattering

22

Deung-Jang Choi and Nicolás Lorente

Contents

Abstract

The scanning tunneling microscope is a unique tool for injecting electrons with great energy and spatial resolution. These electrons can interact with local magnetic moments and reveal the rich phenomena associated with spin-flip dynamics. This chapter is a review of these interesting effects with emphasis on physical concepts and recent results. The Kondo effect is of paramount

D.-J. Choi (✉) · N. Lorente
Centro de Física de Materiales CFM/MPC (CSIC-UPV/EHU), Donostia-San Sebastian, Spain

Donostia International Physics Center (DIPC), Donostia-San Sebastian, Spain
e-mail: djchoi@dipc.org

importance in condensed matter physics. When Kondo correlations and systems of spins of increasing complexity are used, new phenomena appear such as heavy-fermion phases as recently revealed by quasiparticle-interference experiments. New developments showing the relationship between atomic spin entanglement and appearance of Kondo correlations are reviewed. Non-equilibrium Kondo brings in its own features although in the tunneling regime, the Kondo effect is largely accounted for its equilibrium properties. Kondo peaks, inelastic electron tunneling spectroscopy, electron paramagnetic resonance of single atoms, molecules, and atomic objects are explored.

1   Introduction

Adsorbates on solid surfaces can present localized magnetic moments. The properties of these local magnetic moments give rise to very rich and interesting physics. Recently, it has been shown that the scanning tunneling microscope (STM) is able to measure the spectra of magnetic excitations of these adsorbates (Heinrich et al. 2004; Ternes et al. 2009; Gauyacq et al. 2012; Ternes 2015, 2017), having in this way a probe into the quantum world of atomic adsorbates. Among the rich properties of this quantum world, the Kondo effect has a predominant role revealing the complexity of many-body correlations.

The combination of the just-mentioned magnetic spectroscopy, with the atomic manipulation abilities of the STM, is giving rise to new exciting possibilities. Atoms can be located on a surface in ways that were never thought of before. This permits us to show new properties that only existed in theoretical works. Tantalizing examples are the creation of tailored spin chains with antiferromagnetic ordering (Hirjibehedin et al. 2006; Loth et al. 2012), with ferromagnetic spin waves (Spinelli et al. 2014), with exotic XXZ phases (Toskovic et al. 2016), or with Kondo correlations that did not exist in the isolated components before atom manipulation (Choi et al. 2017b).

Another example of the fast-track evolution of local magnetic moment physics and the STM is the advent of electron paramagnetic resonance (EPR) techniques (Manassen et al. 1989; Komeda and Manassen 2008; Balatsky et al. 2012; Müllegger et al. 2014; Baumann et al. 2015). The measurement of time-resolved signals with the STM opens a new door to hyperfine interactions (Heinrich, 2017, personal communication) and to studying spin lifetimes (T_1) and decoherence times (T_2) (Delgado and Fernández-Rossier 2017). Pure dephasing times (T_2^*) understood as the component of T_2 that deregulate the spin phase via elastic interactions (and not because the population of a given spin state has been quenched after a time T_1) need to be known and understood before any serious coherent manipulation of spins can be successfully undertaken in techniques such as quantum computing (Pla et al. 2012).

In the present chapter, we will make a brief review of the Kondo effect of magnetic adsorbates with special emphasis on developing an understanding of the underlying physics and on exploring recent results. We will also briefly summarize

magnetic inelastic effects at the core of single-atom magnetic spectroscopy. We will show how the study of magnetic impurities can become of increasing complexity by making them mutually interact leading to correlations, entanglement, and new states of matter. In particular, the use of quasiparticle-interference (QPI) analysis in the STM has become a mature technique and is widely used in the study of heavy-fermion systems caused by Kondo correlations (Morr 2017). Due to the exciting new developments of single-molecule EPR, we briefly summarize the developing activities in this area at the end of the chapter.

2   Spin Interaction Between Substrate and Adsorbate

As early as 1930, reports on the behavior of electrical resistivity of noble metals with temperature showed a perplexing upturn of the resistivity as temperature decreased (Meissner and Voigt 1930). This finding was clearly at odds with all the body of knowledge on the electronic and vibrational properties of solids at low temperatures. As temperature is reduced, the resistivity should diminish as $\sim T^5$ due to the quenching of electron-phonon scattering and eventually level off at very low temperatures due to lattice imperfections (van Sciver 2012). It was soon realized that a small concentration of magnetic impurities seemed to be responsible for the anomalous behavior (Kondo 2005). But reports were unclear. In 1964, experimental research was still active in determining when a minimum in the resistivity with dropping temperature was found (Sarachik et al. 1964). The origin of the resistivity upturn became clear the same year, when Jun Kondo computed the resistivity due to scattering off magnetic impurities (Kondo 1964). These calculations proved that magnetic impurities were responsible for the resistivity minimum, and they further showed that coherent spin-flip scattering was the physical process at work. As temperature was reduced, all excitations in the solid were quenched except the ones due to spin-flip scattering that gave rise to an increased scattering rate of conducting electrons. Due to the interaction between the magnetic moments of the electron and the impurity, the spin of an electron can flip at the same time as the magnetic moment of the impurity. The electron can change the magnetic moment in one unit of spin (two Bohr magnetons in magnetic moment units), and the impurity changes in minus one unit, keeping the magnetic moment constant during the electron-impurity scattering.

The calculation by Kondo was based on the s-d exchange model developed by Zener (1951). This model considered the magnetic exchange interaction between itinerant electrons and a quantum impurity. The itinerant electrons were represented by the s-band of a free-electron-like host. The d electrons took into account the localized spin of the impurity. The calculation by Kondo was based on perturbation theory. The result revealed that the second order of perturbation theory gave rise to a logarithmic term on the temperature. This is troublesome for the perturbation scheme because at low temperatures the second order grows as the temperature is reduced and the perturbation expansion is not converging. Nevertheless, the logarithmic behavior turned out to be correct. Kondo fit the experimental data

using an expansion including logarithms, giving very good agreement with experiments (Kondo 1964). In this way, he made a connection between electron scattering off impurity spins and the resistivity minimum.

Logarithmic scaling with interactions is ubiquitous in Kondo systems. They reflect complex many-body processes that cannot be solved with simple techniques. It basically implies that perturbational approaches will fail evaluating the ground state of the system because the new ground state is fundamentally different from the non-interacting one. It further means that all behavior with temperature will be smooth and slow, giving rise to "crossovers" rather than phase transitions.

The new type of ground state is clearly seen by following a renormalization-group (RG) approach (Bulla et al. 2008). The RG approach is based on the observation that to evaluate the ground state, higher-energy degrees of freedom can be effectively considered by a new Hamiltonian. This new Hamiltonian contains the correct ground state if the interactions are changed while removing the higher-energy states. As the higher-energy degrees of freedom are removed, the interactions are renormalized. The evolution or flow of the interactions as the number of available electronic states is reduced gives clear information on the properties of the ground state.

The conditions to have a Kondo ground state are that the magnetic impurity contains internal degrees of freedom that are degenerate and can be switched by an external electron collision where the spin of the electron changes by 1. This is an elastic process where the electron produces a spin flip, thanks to an "effortless" change in the state of the impurity. Kramers theorem (Kramers 1930) warranties that half-integer spin impurities will present a degenerate ground state. In the presence of a positive uniaxial magnetic anisotropy, the ground state of the impurity will be $\pm 1/2$ and the spin-flip ($\Delta s = \pm 1$) becomes possible.

If the uniaxial magnetic anisotropy is negative, the ground state is $\pm S$ and cannot be connected by a spin flip unless $S = 1/2$. The electrons cannot be correlated via scattering with the impurity. However, there can be spin flips with the excited spin degrees of freedom of the impurity. This gives rise to inelastic spin flips. The STM is an excellent tool to reveal all these spin-flip processes (Heinrich et al. 2004; Ternes et al. 2009; Gauyacq et al. 2012; Ternes 2015, 2017). The two types of spin flips, elastic or Kondo and inelastic, can be seen as rapid variations of the differential conductance measured by an STM. Figure 1 shows a typical example for an STM junction where a spin is changing due to elastic spin flips that show as a peak at zero bias or inelastic spin flips that show as sharp steps symmetrically position with respect to zero bias.

The STM consists of a piezoelectrically controlled electrode (the tip) that can be sharpened to be atomically terminated. It is a two-electrode setup as sketched in Fig. 2a. Due to the atomic termination and the sub-Ångström control of the position of the tip, the STM can drive electric currents within the atoms of the other electrode, in general a flat metallic substrate. At low bias, the current proceeds in the tunneling regime because the electrons cannot overcome the vacuum barrier between the electrodes. This tunneling current is extremely localized, both laterally and vertically, which are key for detecting changes in the current under subatomic
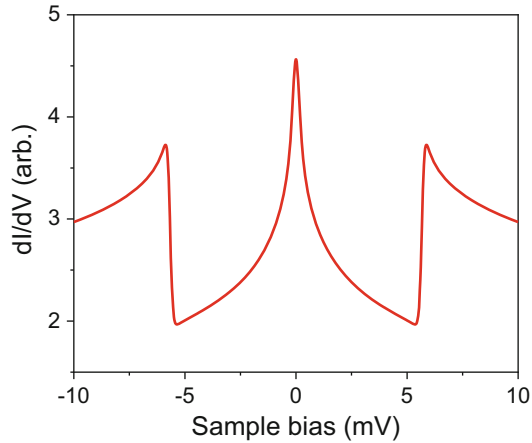
Fig. 1 Simulated differential-conductance spectra of an STM over a magnetic adatom displaying spin-flip processes. At zero bias a clear peak appears due to the existence of a Kondo ground state. The peak signals the presence of elastic spin-flip events that lead to a correlated ground state. Symmetrically at positive and negative bias, two steps appear in the differential conductance signaling the onset of inelastic spin-flip processes. When the spin is excited, new conduction channels abruptly open, increasing the conductance of the junction

motions of the tip. Scanning the surface while keeping the tunneling current constant leads to maps of the tip positions that make the usual STM images. Fixing the tip at a location on the surface and ramping the bias allows to record information on the propagation of electrons through the system. Ideally, the tip is not perturbing the substrate and has either a flat electronic structure or a well-known one. In these cases, the STM gives direct information on how electrons are injected or extracted from the sample. A perturbation-like calculation assuming that only the s-wave of the electronic wavefunctions generating at the tip contributes yields that the conductance is proportional to the local density of states evaluated at some tip "center." The local density of states is the number of states per unit energy weighed by the spatial distribution of each electronic state. This is the Tersoff-Hamann approximation that allows us to interpret STM data as due to the electronic spectra of the surface (Tersoff and Hamann 1985). Under the above conditions, the STM is an excellent tool to measure the spectral properties of a conducting solid surface.

In 1998, the first observations with STM of the Kondo resonance above single magnetic atoms on a metal surface were reported by two groups. Madhavan et al. showed that a narrow asymmetric feature near E_F appeared in the dI/dV of cobalt atoms on a Au(111) surface (Madhavan et al. 1998), while Li et al. observed a characteristic dip around E_F with Ce adatoms on Ag(111) (Li et al. 1998). Both types of spin-flip processes (elastic or Kondo spin-flip and inelastic) can be observed with the STM on a single magnetic impurity. This is due to the peculiar features of the spectral function. In the presence of Kondo physics, the spectral function displays a peak near the Fermi energy (the Abrikosov-Suhl-Kondo or Kondo peak).
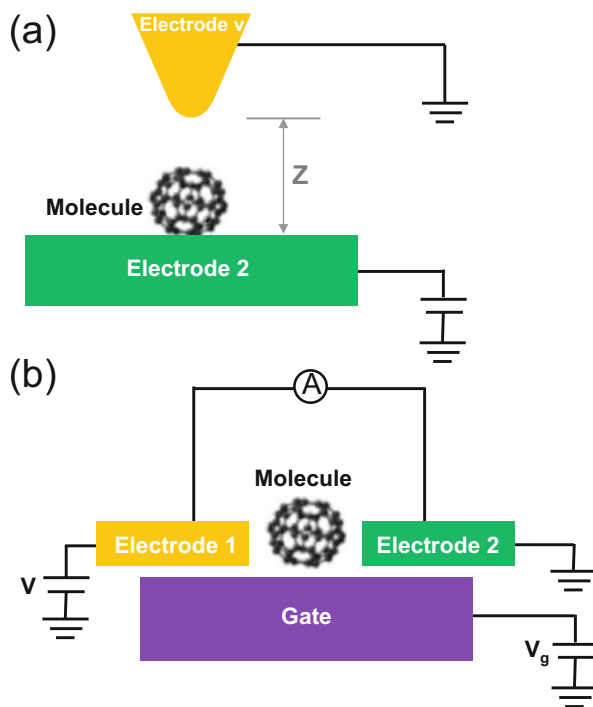
Fig. 2 Sketch of electrodes in a STM setup (**a**) and in a molecular break-junction one (**b**). The limitation of two electrodes in the STM is compensated by the scanning ability of the tip electrode yielding sub-Ångström resolution. Break junctions can change their gaps, but they cannot laterally scan. However they can apply a perpendicular electric field that shifts molecular levels and accordingly change the charge state of the molecule

This corresponds to a feature in the conductance at zero bias, sometimes called a zero-bias anomaly (ZBA). Since the conductance is not uniquely determined by the substrate spectral function, the spectral peak can become a general Fano feature (Fano 1961). In contrast, inelastic peaks generally translate into a step in the spectral function and in the conductance. In simple terms, the step is due to the opening of new transmission channels due to the presence of the excitation. At energies or biases below the excitation threshold, the current proceeds through the elastic channel. Above the threshold, inelastic channels open producing a sudden increase of the conductance (Hipps and Mazur 2006). As in the case of the ZBA, the actual electronic transport process can change the profile with bias producing more complex behavior than just a step in the conductance. In the cases of large conductance (generally not in the tunneling regime), the effect of the excitation will be rather detrimental to transport, and instead of an increase of the conductance at threshold bias, there will be a decrease.

The original Kondo effect implies a reduction of conductance at low temperature. However, the STM Kondo effect can be an increase of conductance. At first

perplexing, the conundrum can be solved by realizing that the conductance of an extended system with many impurities means the averaging of the scattering produced by all the impurities. Hence, the increase of conductance over a single impurity actually means an increased scattering rate that leads to a reduction of conductance over many impurities. The STM gives special insight over the behavior of a single magnetic impurity that can be very different from all other techniques based on averaging procedures.

Before closing this section, it is interesting to mention two different techniques where the excitation of individual magnetic moments has also been achieved by electrical means. One of these techniques is the break-junction technique where a small gap (in the nanometer range or smaller) is filled by adsorbates that contain a magnetic moment (Yanson et al. 1995; Park et al. 2002). These magnetic adsorbates interact with the magnetic moment of passing electrons between the junction electrodes. The other technique is the fabrication of quantum dots (QD), where metallic and semiconducting structures blend, creating an electronic potential landscape that can capture individual electrons and thus originate local magnetic moments. The effect of these local moments on the current can be studied by the conductance signals of the device. The common feature of these techniques is that the experimental setup corresponds to a three-terminal junction as sketched in Fig. 2b. The two electrodes drive a current under the applied bias, and the gate is also electrically polarized but uncoupled via some insulating layers from the molecular junction. Its role is to provide an electric field that shifts the molecular levels without injecting extra charge in the system. This allows to have experiments complementary to the ones with the STM setup. The advantage is the possibility to tune electronic states of the junction, while the STM is more precise in the positioning of the electrodes with the great advantage of being able to scan ample sets of electrode positions.

In 1998, the Kondo effect was observed for the first time in a QD by two different groups. A QD is a nanoscale artificial semiconducting box in contact with metallic leads. A two-dimensional electron gas is confined in the dot and produces a DOS with discrete energy levels. The number of electrons in a QD can be adjusted by applying voltage to a gate electrode. The number of confined electrons in the dot can be odd or even. In the case of an odd number of electrons, the net spin of the dot is nonzero. Goldhaber-Gordon et al. observed a Kondo effect in a single-electron transistor containing a GaAs/AlGaAs heterostructure (Goldhaber-Gordon et al. 1998). When an electric field is originated by the gate electrode, the conductance shows peaks periodic in bias, indicating a quantized number of electrons in the dot. For an odd number of electrons, a Kondo resonance is detected near E_F. This indicates that a localized magnetic moment is present in the dot and that it interacts with the free conduction electrons in the leads. The QD therefore closely mimics the Kondo effect of a magnetic impurity embedded in a metal. Goldhaber-Gordon et al. also showed that the Kondo resonance is suppressed when the temperature is increased and that an external magnetic field can produce a Zeeman splitting of the resonance. As we show here after, these effects are both expected for the Kondo effect. Cronenwett et al. also published the study of the Kondo effect in a

QD (Cronenwett et al. 1998). By adjusting the number of electrons in the QD, they switched the Kondo effect on and off. They studied the temperature and the magnetic field dependence and showed the influence of magnetic anisotropy by rotating the external magnetic field.

Break junctions (Yanson et al. 1995; Park et al. 2002) use an actuator to induce pressure on the gate and mechanically increase the distance between the electrodes (Fig. 2b). Hence, a break-junction experiment can map the behavior of the molecular junction under different charging states, like in the above QD setups. Additionally, they can also change the conductance regime from contact to tunneling by varying the distance between electrodes.

The energy needed to destroy the Kondo ground state is usually given in terms of a temperature, T_K (see Eq. (4)), signaling the temperature above which the Kondo effect disappears. The T_K in QD systems is very low, about a few hundred millikelvin, and as a consequence, to observe the Kondo resonance in a QD requires a higher bias resolution than single-molecule devices where $T_K \sim 10$ K. The reason for the lower T_K in QD as compared to atomic and molecular adsorbates is the lower electron confinement, which leads to very small electronic energies as compared with atomic impurities. In this chapter, we focus on the spin-flip physics of atomic systems. Due to the very different energies involved as compared with quantum dots, typical Kondo energies are much higher and easily resolved by low-temperature scanning tunneling microscopes.

3   The Physics of Spin-Flip Scattering

Let us try to give two physical pictures of the processes at hand. The first one is a ground-state picture that simply acknowledges the fact that the Kondo effect is the creation of a new ground state due to the interactions between electrons and a localized magnetic moment. This ground state is many-body and hence difficult to grasp, but a simplified model can shed some interesting facts easily contrastable with what is known of the Kondo effect. The second picture is of a more dynamical aspect. It treats the Kondo effect as due to coherent spin-flip processes, and it is a local picture, where only the variations of the impurity's magnetic moment are followed during the interaction with the substrate electrons. Finally, we will mention the Kondo peak, as a valuable spectroscopical source.

3.1   The Ground State: Simple Picture of the Kondo Effect

The presence of spin-flip scattering off a magnetic impurity leads to a new ground state of the electronic system. The ground state cannot be approximated as a Fermi sea where one-electron states pile up in energy abiding by the Pauli principle. How can we understand this? A simple cartoon-like model is given by P. Fulde in his book on electron correlations (Fulde 1995).

The model assumes a localized electronic state, a, such as we would expect for an atomic impurity on a solid surface. In the limit of very strong intra-atomic Coulomb

(a) V = 0 (b) V ≠ 0
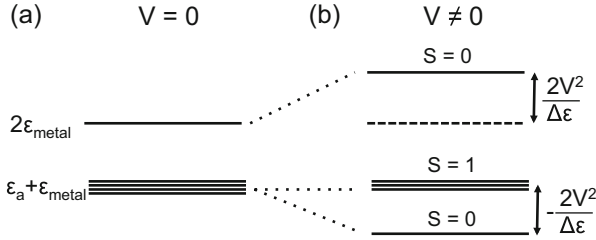
Fig. 3 Energy diagram of the simplified model of Sect. 3.1. Initially we assume just two electrons on two different orbitals. The orbital at energy ε_a is a compact atomic-like orbital, with a large intra-atomic Coulomb repulsion, U, such that double occupancy of the orbital is not permitted. The other orbital corresponds to an extended state, with negligible intra-atomic Coulomb repulsion. When the interaction between the atomic-like and the metallic-like states is zero, $V = 0$, there are four degenerate ground states and one excited state. When a hybridization-like interaction is allowed, $V \neq 0$, a $S = 0$ ground state is generated, leading to the total screening of the otherwise spin-polarized atomic state. The low-energy excitations are magnetic

interaction, $U_a \rightarrow \infty$, this state can only present two occupations, either zero or one electron. This is immediately a magnetic impurity of spin 1/2. To model the metallic surface becomes difficult. Let us assume that it is like an electron in an extended molecular ligand, which is the model that Fulde solves (Fulde 1995). Then our metal will be a single state that is so delocalized that two electrons easily fit in. In other words, $U_{\text{metal}} = 0$. To keep the model simple, we assume we have two electrons in the full system. Furthermore, we assume a nonmagnetic host in the absence of impurities; hence two electrons in the metal will pair antiferromagnetically with spin zero ($S = 0$).

Figure 3a shows the five possible states of these two electrons over the two electronic states (*metal* and *a*) when there is no interaction between impurity and metal. In the absence of impurity-metal hybridization, there are four degenerate states that become the ground state. These corresponds to the four possibilities of putting one electron on the impurity and one electron on the metal. The four possibilities are one singlet state (total spin zero, $S = 0$) and the triplet states (total spin one, $S = 1$). The energy of the ground state is $\varepsilon_a + \varepsilon_{\text{metal}}$ because there is one electron in *a* and one in *metal*. The only excited state is the doubly occupied metal at $2\varepsilon_{\text{metal}}$. The doubly occupied impurity corresponds to an infinite energy ($2\varepsilon_a + U_a$) and is out of our Hilbert space.

When a chemical interaction between impurity and metal is allowed, a one-electron hybridization term allows electrons to flow back and forth the metal from the impurity, which we call $\langle a|V|metal \rangle = V$. This is a one-electron term, so the electron spin is conserved. As a consequence, the new states are still two singlets ($S = 0$) and the three $S = 1$. The interaction couples the two singlets because the hybridization cannot change the spin as we just said. The corresponding Hamiltonian in the $S = 0$ subspace is:

$$\begin{pmatrix} \varepsilon_a + \varepsilon_{metal} & V\sqrt{2} \\ V\sqrt{2} & 2\varepsilon_{metal} \end{pmatrix}. \tag{1}$$

When the hybridization term goes to zero ($V = 0$), we recover the spectrum of the non-interacting case of the $S = 0$ subspace of the previous paragraph. Diagonalizing the new Hamiltonian is simple and yields two states with energies:

$$E_{GS} = \varepsilon_a + \varepsilon_{metal} - \frac{2V^2}{\varepsilon_{metal}}, \qquad (2)$$

and

$$E_{excited} = 2\varepsilon_{metal} + \frac{2V^2}{\varepsilon_{metal}}. \qquad (3)$$

The new spectra are plotted in Fig. 3b, including the invariant $S = 1$ states. We see that the interaction between impurity and metal lifts the degeneracy of the ground state, and the new ground state is stabilized by an energy $-\frac{2V^2}{\varepsilon_{metal}}$. As Fulde shows in his book (Fulde 1995), this simple model contains many features of the complex Kondo physics between an impurity and a true metal substrate.

Features of this simplified model:

- The first startling consequence is magnetic screening of the impurity. Indeed, in the absence of impurity-metal hybridization, the impurity has a local magnetic moment. When the interaction is switched on, the ground state is nonmagnetic and the local moment disappears. This is parallel to the screening of the Kondo effect. In the impurity problem, the magnetic moment of the impurity is screened because the total ground state is nonmagnetic.
- The low-energy excitations are all magnetic. In the simple model, all excitations are singlet-triplet excitations. The excitation energy is $\frac{2V^2}{\varepsilon_{metal}}$. When the system's temperature matches this excitation energy (in the correct units and assuming that we can define thermodynamic quantities), the ground state is destroyed and the system is excited. This temperature would correspond to the Kondo temperature in the correct impurity problem. Obviously the true Kondo temperature is not proportional to the square of the hybridization potential, V^2.
- The ground state is unique, but the first excitation is triply degenerate. Since the ratio $\frac{2V^2}{\varepsilon_{metal}}$ is small because $\varepsilon_{metal} \gg V$, then we have many excited states at very low energies. In the impurity problem, this corresponds to the Kondo peak. Namely, at very low energies (i.e., close to the Fermi energy), there is an enormous density of states due to magnetic excitations. The Kondo peak is a consequence of the low-energy magnetic excitations of the system caused by the hybridization between metal and impurity.
- High-energy excitations correspond to charge transfer from the impurity into the metal. This energy scale is given by ε_{metal}. Excitations are then separated into spin (low-energy, Kondo peak) or charge degrees of freedom.
- As temperature increases, the triplets become populated and the magnetic moment of the impurity is restored.
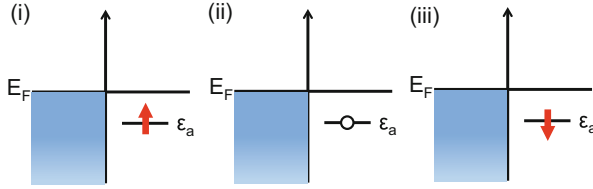
Fig. 4 Local picture of the Kondo spin fluctuations due to the hybridization of an atomic orbital with metallic orbitals. Due to the intra-atomic Coulomb repulsion and the hybridization with the substrate, the atomic charge fluctuates between zero and one electron. As a consequence of the charge fluctuations, the atomic spin changes

An extension of the above features to the impurity problem can be easily done by using a variational approach to the Anderson model (Fulde 1995; Yosida 1966; Varma and Yafet 1976). In the $U \rightarrow \infty$ simplification, the no-hybridization configurations are, see Fig. 4, (i) a Fermi sea (representing the metal) and an empty impurity (excited state), (ii) a Fermi sea with a missing electron (hole) and a filled impurity (excited state), (iii) and a filled impurity with a hole at the Fermi energy (ground state). Minimizing the expected value of the Anderson Hamiltonian with respect to the coefficients of the three configurations when the hybridization term is allowed, we obtain a new ground state that is stabilized by an energy:

$$k_B T_K = D \, \exp\left(-\frac{\pi |\varepsilon_a|}{\Gamma_a}\right). \tag{4}$$

The electronic bands of the metallic surface are represented by half the bandwidth D. Here Γ_a is the width of the impurity level due to the hybridization with the substrate ($\Gamma_a = 2\pi \sum_k V_k^2 \delta(\varepsilon_a - \varepsilon_k)$). The impurity level energy, ε_a, is now measured from the Fermi energy, which is ε_{metal} in the above simplified model. Then we see that as above, the ingredients determining the strength of the Kondo effect (the ground-state energy) are ε_a and Γ_a.

The behavior of ε_a and Γ_a is actually intrinsically related with the occupation of the impurity level on the metal surface. In the case that $|\varepsilon_a| \gg \Gamma_a$, the level is completely occupied, and the Kondo temperature, Eq. (4), tends to zero. Hence, even if there is a magnetic moment on the impurity with a sizeable coupling to the substrate (large Γ_a) when the level is always full ($|\varepsilon_a| \gg \Gamma_a$), there is no Kondo effect, and the magnetic moment of the impurity remains frozen. This is the local-moment regime. On the other limit, if $|\varepsilon_a| < \Gamma_a$ the level is basically half-filled (or closed to half-filling). This corresponds to a very large Kondo temperature. As a consequence the peak of spin excitations (the Kondo peak) will merge with the peak of charge fluctuations (the resonance derived from the ε_a level). Indeed the charge fluctuations will win over the spin ones, and the Kondo effect will not be well defined. This is the mixed-valence regime. Mixed valence means that the charge in the impurity is fluctuating among several valence states of the impurity. Finally, the

Kondo regime takes place in between, for impurity levels that are largely occupied but allowing charge fluctuations.

The study of charge fluctuations is clearly linked to the study of Kondo physics. We briefly review this in the next section.

3.2   The Dynamical Spin-Flip Picture

Adopting a local picture immediately leads us to charge fluctuations and their role in the Kondo effect. P. W. Anderson (1961) soon realized that the $s - d$ exchange model of Zener and others that Kondo adopted (1964) could be explained in terms of charge fluctuations of the impurity (Schrieffer and Wolff 1966). The occupation of the impurity level will change over time as a consequence of the hybridization with the substrate and the large intra-atomic repulsion ($U \rightarrow \infty$ in the above discussion) if the impurity level is within Γ_a of the Fermi energy following the preceding discussion. If the level is very far from the Fermi energy, the charge state will be frozen and no Kondo effect is possible. But if the average charge state is not zero or one, the level occupation fluctuates between zero and one electron.

Figure 4 shows a simplified scheme of this system. During charge fluctuation the spin is initially in one state, say up. After a certain time, the impurity gets empty. Then it captures another electron with spin up or down with equal probability for nonmagnetic substrates. And this process is repeated over time. As a consequence, the spin on the impurity changes randomly over time. If the average charge state is large, say 90%, then the impurity is 90% of the time occupied, and for a brief 10% of the time, it is empty. An observer would see an impurity that is occupied but that changes randomly its spin due to its hybridization with the metal substrate. Effectively this is a $s - d$ exchange interaction: the spin on the impurity flips by flipping back a spin in the metal. P. W. Anderson (1961) showed that this effective interaction is of the antiferromagnetic exchange type (trending to antialign spins in the impurity and in the metal). Later on Schrieffer and Wolff (Schrieffer and Wolff 1966; Hewson 1993) set a rigorous framework to transform between the Anderson Hamiltonian (describing the impurity hybridization and charge fluctuations) and the $s - d$ exchange or Kondo Hamiltonian.

The difference between the spin-flip picture and the previous ground-state calculation is that the spin-flip picture locally evaluates the physics of the impurity. As a consequence, the construction of the ground state of the full system is simplified to the local fluctuations of the impurity state. The physics is the same, but they offer complementary pictures that help understanding the Kondo effect.

3.3   The Spectroscopic Signature: The Kondo Resonance

The spectroscopic signature of the Kondo effect is the Kondo peak. In the STM this shows as a zero-bias anomaly in the conductance. In photoemission, it is a peak in the spectra. The peak takes place very close to the Fermi energy because it

comes from very low-energy excitations of magnetic origin. When the intra-atomic Coulomb repulsion is finite, hole-like excitations are also allowed and the Kondo peak actually extends below the Fermi energy. In a symmetric case (symmetric Anderson model, $|\varepsilon_a| = U/2$), particle and hole excitations weigh the same in the spectra, and the Kondo peak is centered on the Fermi energy.

As we show in the simplified model above, not only the ground state is shifted due to the hybridization but also the metallic states. If we evaluate the square of the overlap of the metallic state before and after switching on the hybridization, we find that it is $1 - n_a$ for metallic states close to the Fermi energy, where n_a is the occupation of the impurity level. In the Kondo regime, n_a approaches one. This means that the metal states at the Fermi energy before and after connecting the hybridization are virtually orthogonal. The metal states become quasiparticles that actually behave like non-interacting fermions (Hewson 1993).

Actual STM measurements are quite involved, and the complete characterization of the Kondo peak requires several measurements. However, the Kondo peak has some unique features that make it totally different from any other source of ZBA. The Kondo effect disappears with increasing temperature by developing a particular broadening of the Kondo resonance. The broadening can be shown by measuring the corresponding decrease of the conductance amplitude at resonance. Changes in amplitude with temperature can be expressed as (Pustilnik and Glazman 2001):

$$G(T) = \frac{2e^2}{h} t \left[1 + \frac{T^2}{T_K^2} (2^{1/s} - 1) \right]^{-s} + G_{el} \qquad (5)$$

with $s = 0.22$, while G_{el} accounts for a constant background. The free parameters T_K, t, and G_{el} are found by fitting the experiment. As for the broadening, assuming a Lorentzian resonance, the full width at half maximum Δ_K is approximately (Nagaoka et al. 2002; Ternes et al. 2009):

$$\Delta_K = \sqrt{(\alpha k_B T)^2 + (2k_B T_K)^2} \qquad (6)$$

where k_B is the Boltzmann constant, T is the working temperature, and $\alpha = 5.4$ (Ternes et al. 2009). When $T \ll T_K$, then $\Delta_K \simeq 2k_B T_K$.

A second sign that experiments are dealing with the Kondo effect is the peak dependence with external magnetic field, $\vec{B}$. An external magnetic field induces a Zeeman splitting of the impurity's magnetic levels; hence spin-up scattering is not any longer degenerate with spin-down scattering, and the Kondo effect is destroyed. However, its destruction is not complete, because electrons of higher energy can interfere with opposite spin electrons if their energy difference is exactly the Zeeman splitting. Indeed, we are describing an inelastic effect, and at the energy threshold of this inelastic effect, the electrons can spin flip coherently, recovering a Kondo peak, but instead of being centered near the Fermi energy, the new peak is centered at the energy of the excitation. Since electrons above the Fermi energy and below (holes) can produce this inelastic effect, there are two new peaks that

appear, replacing the previous Kondo peak when the magnetic field is applied. For very large fields, coherence cannot be achieved and all peaks disappear, or in other words, the impurity's magnetic moment cannot be screened out at large external magnetic fields.

Both the temperature and magnetic field behaviors permit us to characterize a zero-bias anomaly as due to the Kondo effect. However, experiments are further complicated by the fact that these anomalies are rarely pure peaks. Indeed, they present Fano line shapes that are more difficult to analyze. The origin of the Fano line shape comes from the contribution of many electronic states to the STM conductance other than the impurity's ones. In STM, the line shape of the Kondo resonance depends on the quantum interference, or Fano interference, of the tunneling channels (Fano 1961; Madhavan et al. 2001). One of the tunneling channels takes place between the tip and the Kondo resonance. The second channel is tunneling between the tip and the metal surface through a non-Kondo process. The third channel is directly tunneling into the d-level of the impurity. In this case, the electron eventually has its spin flipped and is located at the Fermi energy in its final state. Channels 1 and 2 preserve the spin orientation and therefore can interfere. This interference is the Fano interference of a localized state with a continuum. Depending on the relative phases of both processes, the interference produces the asymmetric line shape of the ZBA. Several workers have devoted their attention to the appearance of these Fano profiles (Madhavan et al. 2001; Plihal and Gadzuk 2001; Merino and Gunnarsson 2004; Baruselli et al. 2015; Jacob 2015; Frank and Jacob 2015; Fernández et al. 2016). They all model the previous interference mechanism but using different numerical tools. Frank and Jacob (2015) simplify their approach writing a 2×2 matrix for the Hamiltonian, including the above principal scattering channels, and give a very physical account of the leading terms of the interference for the case of Co adatoms. Previously, similar calculations (Merino and Gunnarsson 2004) had shown the importance of particular electronic structure on the continuum states, such as surface states; recently Aligia and co-workers explain the experimental Fano line shape of a Co atom on confined surface states following the same rational (Fernández et al. 2016).

4   Inelastic Spin-Flip Spectroscopy

The excitation of spins using the STM is a mature technique with many well-tested examples and a thorough understanding of its principles. The first IETS was performed by Heinrich and co-workers (Heinrich et al. 2004) studying Mn atoms on an $Al_2O_3/NiAl$ substrate. The oxide layer was proven to be fundamental to increase the lifetime of the excitation and in this way resolve it in the spectra. The spectra were a set of steps that split in smaller steps under an external magnetic field, showing the magnetic origin of the inelastic effect. The steps in the spectra originate when the applied bias is larger than the inelastic thresholds. At this moment a new inelastic channel opens, suddenly increasing the conductance. The reason why

the rise is so abrupt comes from the sharp edge of the Fermi distribution at low temperatures. Indeed, during tunneling the inelastic effect is a close channel if the final electron states correspond to an occupied state or, in other words, to an energy below the Fermi energy. Hence, the Fermi distribution function has a fundamental role in determining when the inelastic effect takes place; as temperature increases, the Fermi distribution function smears out, eventually rendering the opening of the channel, and hence the conductance step, invisible.

Other experiments followed using IETS to characterize magnetic adsorbates and magnetic interactions at surfaces. The assembly of atomic spin chains of Mn on Cu_2N/Cu(100) was an astonishing example of the capability of the STM to create new physical objects not explored before (Hirjibehedin et al. 2006). The atoms were aligned leaving a distance of 3.5 Å between atoms, and they were shown to interact via a spin-spin Heisenberg interaction of the form $J\vec{S}_i \cdot \vec{S}_{i+1}$ between the spins of atoms i and $i + 1$. The experimental data nicely followed the excitation thresholds of the above Heisenberg chain, showing the spectroscopic signature of the spin excitations of each chain.

Single Fe atoms on the same substrate also showed tantalizing magnetic spectra that were correctly identified as due to the magnetic anisotropy created by the surface on the Fe atom (Hirjibehedin et al. 2007). These experiments showed that the strength of the signal, namely, the size of the conductance steps at the inelastic thresholds, followed the modulus square of the matrix element between initial and final spin of the total spin vector. This empirical result was theoretically justified by Mats Persson (2009) and soon after identified with the strong-coupling co-tunneling arguments of Lorente and Gauyacq (Lorente and Gauyacq 2009; Gauyacq et al. 2012). This last work emphasized the experimental fact that the inelastic changes in conductance were outstandingly large, easily becoming larger than the elastic contribution, in contrast to vibrational IETS (Ho 2002; Komeda 2005; Morgenstern et al. 2013) that rarely yields changes above 10% of the elastic conductance. The authors justified the efficiency of the inelastic transitions using the following physical picture (Lorente and Gauyacq 2009; Gauyacq et al. 2012). The tunneling of electrons leads to the temporary formation of a negative-like (or positive-like or both) intermediate ion. When the electron leaves the ion, the final state can have a large probability to be different from the initial one. This is because the co-tunneling process (an electron jumping in and out of the atom) changes the spin by one unit, and the overlap of this state with one of the excited states plus the spin of the outgoing electron is related to Clebsch-Gordan coefficients that are large numbers, many times larger than one. As a consequence, the transition probability can become very large. Indeed, the physical process is the one of angular momentum compositions between the impurity spin and the tunneling electron spin. Angular momentum composition can lead to large spin transfer and hence to an efficient spin torque (Delgado et al. 2010; Novaes et al. 2010; Gauyacq et al. 2012).

The $\vec{S}^2$ theory for spin excitation is simple and permits a quick rationalization of experimental spectra. It has become a standard for inelastic spin transition using ab initio (Hurley et al. 2011a, b) or empirical approaches (Ternes 2015).

Recent experiments (Ormaza et al. 2017) performed on a nickelocene molecule show extremely large changes in conductance. Indeed, the step size is more than six times the elastic conductance. This means that when a current is driven through this type of molecules, the current is basically inelastic, and any type of rationalization of the electronic transport of the molecule in terms of elastic currents (i.e., Landauer approach) will fail. Nickelocene is a $S = 1$ molecule, with an easy plane anisotropy due to its internal ligand field. The large inelastic efficiency is a consequence of the efficient spin transitions due to the actual energy positioning of the conducting molecular orbitals with respect to the Fermi energy (Ormaza et al. 2017).

Many other experiments have been performed together with theoretical work. These works have been recently summarized in different review articles (Gauyacq et al. 2012; Ternes 2015, 2017).

5   Magnetic Molecules in Break-Junction Experiments

A Kondo peak is clearly seen in the differential conductance at different temperatures in a Au-C_{60}-Au mechanical controllable break junction, where the Kondo effect is carried by the C_{60} molecule (Parks et al. 2007). Despite the very nonmagnetic nature of fullerenes, a large fraction of contacted molecules showed Kondo peaks in the conductance measurements. Obviously the difference among the sets of measurements was the way molecules were contacted by the electrodes. Fullerenes are known to display large electron affinities, and in the presence of an electron-rich environment such as metallic electrodes, charge transfer is natural. The experiments can be largely explained by the transfer of a single electron, in the cases where the Kondo peak was seen. The implication of these findings is that the electrode-molecule interface is of great importance determining charge transfer, the chemical stability, and ultimately the system's electronic ground state.

Molecules of increasing sophistication have been also studied in break-junction setups (Gaudenzi et al. 2017). Radical molecules present unpaired electrons which is a nonstandard molecular electronic structure. As such the spin leaves in a molecular orbital associated with a group of atoms, contrary to the usual case of local magnetic moments associated to single magnetic atoms that generally are transition-metal or rare-earth atoms. The experiment uses the gating electrode to change the charge state of the molecule, which is beautifully demonstrated by the magnetic IETS at the different molecular charges. A radical electron in a molecule is very unstable in front of charge transfer from the electrodes, and hence we generally do not expect to find any Kondo evidence in organic molecules without metallic centers. Surprisingly, Kondo peaks in the break-junction traces of these molecules are found. The conclusion is that these molecules are special radicals that preserve their open-shell electronic structures despite of not being in solution and in contact with electrodes (Gaudenzi et al. 2017).

Park et al. (2002) studied the Kondo effect of a molecular break-junction device. They showed that a Co ion bonded to polypyridyl ligands with different lengths of insulating tethers, $[Co(tpy-(CH_2)_5-SH)_2]^{2+}$ and $[Co(tpy-SH)_2]^{2+}$, presented a

Zeeman-split Kondo resonance with an external magnetic field in the dI/dV. The splitting amplitude may be tuned by changing the magnetic moment of the molecule via the gate electrode, while changes of the splitting with external magnetic field yield a gyromagnetic factor close to 2 as expected for a free electron, showing that the intrinsic spin-orbit coupling of Co is small.

Break junctions and molecular systems have a rich set of parameters to tune different Kondo regimes. Work performed in Cornell with theoretical insight from Bariloche (Parks et al. 2010) showed that different regimes of the Kondo ground state of a magnetic molecule can be achieved by playing with the mechanical degrees of freedom of the molecule. Stretching or contracting the molecule shifts electronic states which change the magnetic moment and the coupling of the molecule with the electrodes. The consequence is that the molecule can be driven in a state where the spin is too large for the available number of hybridization channels with the substrate. The substrate electrons cannot screen the local magnetic moment completely, and the Kondo effect takes place in the presence of a surviving local magnetic moment. This underscreened Kondo regime is clearly seen in the splitting of the Kondo peak. Contrary to the screened Kondo that needs an external critical magnetic field to split, the underscreened Kondo displays a split peak as soon as the magnetic field is different from zero (Parks et al. 2010).

6 The Kondo Effect and Interimpurity Interactions

When several local magnetic moments interact on a substrate, they can compose, creating a new object with different magnetic moment, different spin dynamics, and different surface spin screening. Their properties can be utterly different from the initial constituents on the same surface. Here, we briefly review the work performed with the STM, revealing these interesting phenomena.

6.1 The Two-Impurity Problem

Wahl et al. (2007) used atom manipulation to vary the distance between two Co adatoms on Cu(100) and recorded dI/dV spectra. For interatomic distances beyond 6 Å, each Co adatom presented individual and unperturbed Kondo resonances. Below 6 Å, the Kondo ground state of the two Co atoms was perturbed through an exchange interaction. The sign of the exchange interaction depended on the Co-Co distance. At 5.12 Å, the exchange coupling was AFM, and the Kondo resonance splits apart. The resonance splitting was, however, not sufficiently strong to be detected in the dI/dV spectrum. Rather, a broadened line width was observed. At 2.54 Å, the two Co atoms coupled instead ferromagnetically. Since no feature appeared in the dI/dV spectrum, this meant that $T_K^N \ll T$, i.e., the new Kondo ground state was lost at the working temperature T. In a different study by Bork et al. 2011, they picked up a Co atom with the tip and approached the tip into contact

with a Co adatom on a Au(111) surface. By varying the interimpurity distance, the authors were able to map the phase space of different ground states as the impurity-impurity interaction changed the Kondo screening.

Another experiment consisted of $CoCu_nCo$ chains adjusting the distance between two Co atoms (Néel et al. 2011). Depending on the length of the chain, a different exchange interaction is produced. For $n = 1$, the interaction is ferromagnetic; for $n = 2$ the interaction is antiferromagnetic. In this last case, no resonance splitting is observed, but only a broadened feature. An oscillation for T_K was evidenced with n ≥ 3 due to an RKKY interaction mediated by the Cu atoms between two Co atoms.

Electrons in different levels of a nano-object such as a cluster of molecules can also present a complex Kondo effect. This is the case of copper phthalocyanine (CuPc) molecules on Ag (100) (Mugarza et al. 2010, 2012; Korytár and Lorente 2011). As in the case of point-contact and break-junction setups, the presence of the surface can induce charge transfer in the adsorbed systems. In CuPc, experimental and theoretical evidence (Mugarza et al. 2010, 2012; Korytár and Lorente 2011) show that an electron can be captured by the large organic ligands of the molecule. The magnetic situation becomes very interesting; there are two different 1/2 spins, one in the ligand and one in the d^9 configuration of the Cu ion, that are weakly coupled through the molecular structure. Electronic-structure calculations show that the coupling is ferromagnetic, and this is fully consistent with the Kondo structure found in the experimental dI/dV. In the experiment a clear Kondo peak is found, very similar to the one that is found for NiPc. Additionally, the spectra show an electron satellite as well as a hole one around the main Kondo peak; these two smaller peaks are absent from the NiPc case. NiPc basically shares the ligand structure with CuPc, in agreement with the idea that an extra electron is being captured there and this is the origin of the Kondo peak. However in CuPc there is an extra d electron; the ferromagnetic coupling with the ligand electron leads to an overall $S = 1$ that presents an excitation to $S = 0$. Then above the excitation energy of roughly 25 meV, the injected electrons can excite the molecular singlet and end up at the Fermi energy of the substrate that holds the Kondo peak. The corresponding large Kondo density of states leads to a peak at the excitation energy. If the electron is extracted from the sample, a similar peak is detected at -25 meV. The physical picture is of a two-impurity problem where one of the spins is strongly screened by the substrate (Kondo effect), while the other spin is in a compact d orbital with small interaction with the substrate and, thus, no Kondo screening. This is the scenario of an underscreened $S = 1$ Kondo system at present measuring temperatures (few Kelvins).

The combination of molecules in chains also gives rise to interesting Kondo physics. In DiLullo et al. (2012), they show the modulation of the Kondo temperature along the molecular spin chain.

This type of physics turns out to be ubiquitous in molecular adsorbates. The double-decker $TbPc_2$ also shows a prominent Kondo feature (Komeda et al. 2011) due to a ligand electron and so does YPc_2 (Robles et al. 2012). The large magnetic moment of the core f-electron of $TbPc_2$ is largely intact. As a consequence the system is again an underscreened Kondo one.

6.2   Kondo Lattices

When Kondo impurities interact, arranged in an infinite array, new physics takes place (Hewson 1993). Below a coherence temperature, T^*, the Kondo impurities are coherent and create a full new ground state of the many-electron system. It can be proved that the new state behaves like a fermion. Given that this new fermionic state is the result of weakly hybridizing, well-defined Kondo states, we would expect to find a dispersless state. Indeed this is the case and the lack of dispersion of the new fermion can be translated in a very large effective mass, typically thousands of times the electronic mass. Kondo lattices are at the origin of heavy-fermion states.

These states have been achieved (Tsukahara et al. 2011) in arrays of iron phthalocyanine (FePc) molecules on Au (111). Their Kondo states hybridized, giving rise to a small gap at the Fermi energy with two dispersless states. But heavy-fermion compounds are known for a long time (Hewson 1993). The adsorption of undissociated oxygen molecules with their large spins (Alducin et al. 2010) on Ag (110) and Au (110) is suspected to give rise to a Kondo lattice and hence to a heavy-fermion phase as well (Jiang et al. 2011). Another example of a 2-D Kondo lattice has been reported on fluorinated FePc molecules on Au (111) (Girovsky et al. 2017).

A recent review by Morr (2017) shows the extraordinary progress in classical 3-D heavy-fermion systems, thanks to the study of quasiparticle-interference (QPI) patterns with the STM. QPI is very sensitive to the small-energy scale of heavy-fermion excitations about the Fermi energy. Since these excitations originate in the substrate-mediated coupling of Kondo peaks, they carry all relevant information on the heavy-fermion structure. The basic principle in QPI is the interference caused in the extended Fermi states by scattering off defects on the surface of the solid. This permits a mapping of the Fermi surface (without defects but thanks to the defects) following the q-vectors of the surface Brillouin zone. Heavy-fermion superconductivity in compounds, such as $CeCoIn_5$, has been largely attributed to spin fluctuations, thanks to these studies (Morr 2017).

Recent STM experiments on $CeCoIn_5$ achieve outstanding resolution of Co and Ce atoms on cleaved surfaces of this material (Kim et al. 2017). The tunneling measurements show that the STM is only sensitive to the d electrons of the Co atoms; but when the STM reaches close to the surface, the d electrons of the Ce atoms become visible, leading to characteristic patterns on the surface. This study permits to identify the role of the different atoms of the material and give experimental evidence of the effect of spin fluctuations driving superconductivity.

6.3   Impurity Entanglement and New Quantum Objects

The above cases studied the interaction between Kondo systems. The initial magnetic moments, despite of being Kondo screened, managed to interact either via direct exchange interactions or via substrate-mediated interactions. But the Kondo effect was individually operative for each impurity. Here, we mention new

experiments where non-Kondo impurities become a new system that can be Kondo screened. Surprisingly, the individual constituents of the nanostructure are not Kondo screened.

The experiments (Choi et al. 2017b) consisted in assembling a heterogeneous atomic chain on a Cu_2N/Cu (100) substrate via atom manipulation. On this surface, Mn and Fe atoms cannot be Kondo screened at accessible temperatures. The reason of this lies on the anisotropy of these spin systems. Fe is a $S = 2$ system on this substrate with a strong easy-axis anisotropy that leads to a nondegenerate ground state, and Mn is 5/2, also easy-axis (albeit a different direction), and its ground state is degenerate but very inefficiently coupled via spin flips (indeed the weight of allowed $\Delta S = \pm 1$ transitions is tiny). These two types of atoms do not present measurable Kondo peaks. But when they are forced to interact on a surface, they develop the characteristic spectra of a Kondo system. Their Kondo peak is visible from the FeMn system, as soon as two inelastic shoulders also appear in the spectra (Fig. 5). These inelastic shoulders come from the collective spin excitations of the two-atom system and are a direct measurement of the strength of their coupling (Hirjibehedin et al. 2006; Choi et al. 2016b, 2017b). The compound system has a not well-defined total spin, but their ground state is doubly degenerate and can be coupled via a spin flip. However, the S_z component of the spin is a good quantum number, and the ground state is $S_z = \pm 1/2$. This can be flipped by a substrate's electron. As a consequence, Kondo physics develops (Fig. 5a). A simple-minded S_z calculation of the FeMn systems easily yields a $S_z = \pm 1/2$ ground
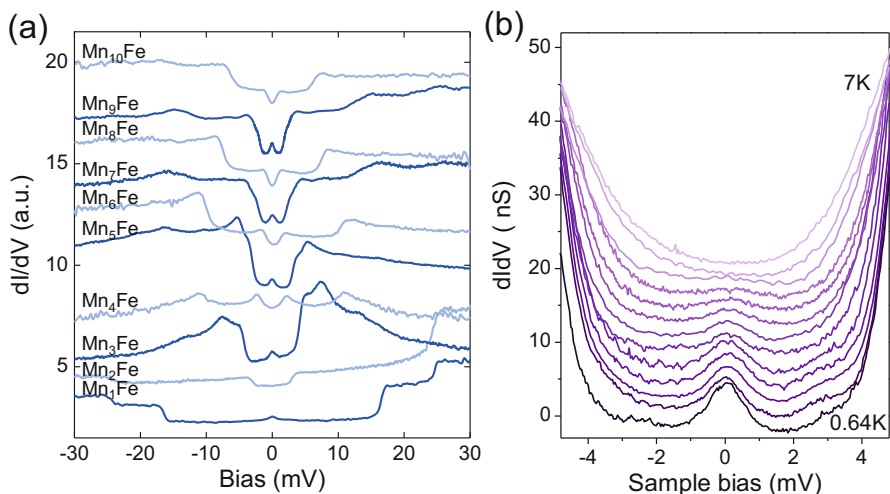
Fig. 5 Differential conductance (**a**) as a function of the applied bias for $Mn_x Fe$ chains for x going from 1 to 10. The spectra are radically different as the number of Mn atoms, x, increases. However, some common features can be remarked if x is even or odd. For even x no Kondo peak is observed, while it is clear and increasing for increasing odd x. The Kondo peak presents a strong temperature dependence as shown in (**b**) for the Mn_3Fe case

states because the atoms couple antiferromagnetically on this surface (Choi et al. 2016b, 2017b). The ZBA are really Kondo peaks as tested by their behavior with temperature, following Eqs. (5) and (6); see Choi et al. (2017b).

Two very different evolutions of the spectra follow as more Fe or Mn atoms are added to the FeMn system (Fig. 5a). Only if FeMn$_x$ spin chains are assembled with x an odd number, Kondo peaks are experimentally found. The parity of x can be understood from the need to have $S_z = \pm 1/2$ ground states. However, MnFe$_x$ do not show Kondo peaks. The reason for this last behavior is more subtle and has to do with the increasing anisotropy of the spin chain as the number of large-anisotropy Fe atoms increases (Choi et al. 2016b). A larger anisotropy implies less weight of terms that favor spin flip in the ground-state composition. In the limit of very large anisotropy, an Ising-like solution with Néel states is favored, but despite their $S_z = \pm 1/2$ character, these states cannot be toggled by a spin flip. The reason is simple; Néel states are a collection of spin up and down; if we allow one spin flip of the $S_z = 1/2$ Néel ground state, we change one only spin flip of the chain. The resulting state is an excitation orthogonal to the $S_z = -1/2$ Néel ground state. As a result, the excited state cannot de-excite into the "other" Néel ground state without higher-order interactions.

This shows that the actual composition in terms of spin configurations of the ground state is very important. The number of spin configurations in the ground state is a measurement of the entanglement of the system. In a bipartite system (e.g., the Fe atom with the Mn$_x$ chain), the entanglement can be determined using the von Neumann entropy that is an efficient way of computing the amounts of bipartite configurations entering the ground state. In Choi et al. (2017b) a direct correspondence is shown between the von Neumann entropy and the strength of the Kondo peak giving an experimental way of evidencing entanglement in a composite system of spins.

7 Point Contact and High-Conductance Regime

To study the conductance of atomic-sized contacts, a variety of experiments using the tip of STM were performed (Pascual et al. 1993; Agraït et al. 1993; Olesen et al. 1994; Gai et al. 1996). These experiments were carried by bringing the tip into the pristine surface forming a contact. In these experiments, contact formation is monitored by recording the current at a fixed junction voltage. Instead of limiting investigations to tip-surface point contacts, tip contacts can be also performed with single atoms or molecules adsorbed on a metal surface as shown in some STM experiments (Limot et al. 2005; Néel et al. 2007a, b, c, 2008a, b; Kröger et al. 2007; Jensen et al. 2007; Kröger 2008). Tip-adsorbate contacts turn out to be highly reproducible and stable contrary to tip-surface contacts. Single-atom or single-molecule junctions can be formed in this way with an increased control compared to other experimental techniques. The electrodes (the tip and the surface) and the contact constriction (the atom or the molecule) can be in fact characterized prior to and after the contact.

Conductance measurements of a point contact formed by an STM tip on a single atom show surprising stability of the Kondo resonance. The Kondo peak is little affected by structural changes induced by the tip-atom contact (Néel et al. 2007c; Vitali et al. 2008; Bork et al. 2011). However, more recent measurements show that atomic relaxations produced by the tip displacement continuously affect both the Kondo resonance and the conductance prior to and after tip contact (Choi et al. 2012). The experimental results show a reproducible exponential variation of the Kondo width as a function of conductance. The findings demonstrate that the Kondo effect of single atoms adsorbed on a metal surface, and more generally their magnetism, may be tuned via the ballistic conductance of the junction.

Recent experiments using the STM to contact a Co atom on Cu (100) with a sharp nickel tip (Choi et al. 2017b) show a splitting of the ZBA. The ZBA splitting is a sign of strong magnetic polarization of the tip. The authors argue that in spite of the strong polarization of the tip, the atom remains largely unpolarized due to the coupling to the nonmagnetic surface. This leads to the survival of the Kondo effect despite the injection of spin-polarized electrons from the ferromagnetic tip. However, the Kondo effect is influenced by the exchange coupling with the tip, acting like a local magnetic field that effectively splits the Kondo peak. These results are in line with previous studies of the Kondo effect on ferromagnetic point contacts that show the survival of Kondo correlations in the presence of spin-polarized currents (Calvo et al. 2009).

7.1   Non-equilibrium Kondo

Further studies of Co on Cu (100) have been scrutinizing the behavior of the Kondo effect under non-equilibrium (Choi et al. 2017a). This is typically the case when a difference of bias is established, and the drop of bias is not complete at the impurity. In the case of contact experiments, the bias is continuously dropping along the contact region where the impurity is localized. Two major effects signal the presence of non-equilibrium processes in Kondo: (i) there are two nearly equivalent Fermi energies due to the two contacted electrodes; this naturally gives rise to some double peak structure akin to producing a Kondo peak per electrode; (ii) due to the partial drop, the Kondo effect can be established with one electrode producing a Kondo peak well above the other Fermi energy; as a consequence, the scattering electrons can continue into the second electrode, producing some sort of leakage from the Kondo effect; this leads to enhanced decoherence of the Kondo spin flips.

Using DFT and non-equilibrium Green's functions transport calculations, the different channels were identified as well as the magnetic structure of the atomic junction (Choi et al. 2017a). Cobalt maintains its magnetic moment of $2 \mu_B$ regardless of the tip-substrate distance, but the transport properties change as the tip approaches the sample. In tunneling, transport proceeds by the sp electron structure of the adsorbate. When the couplings of the atom to tip and substrate equilibrate, new conduction channels open due to the hybridization of the Co d-electrons to
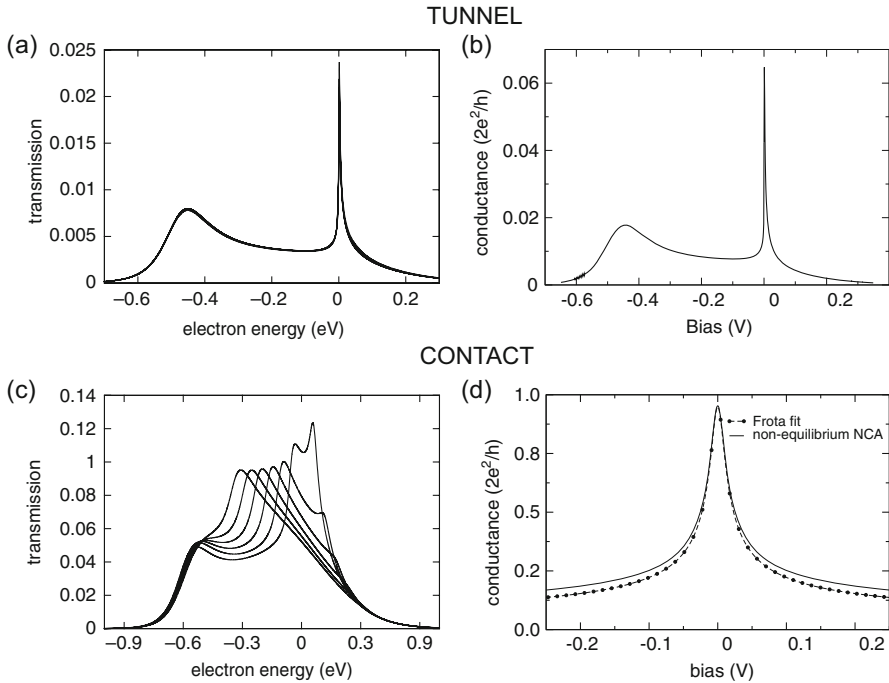
Fig. 6 Comparison between the transmission function (**a**) and (**c**) (proportional to the spectral function in the present calculations) and the differential conductance (**b**) and (**d**) in tunneling and in contact, respectively. We see that only in tunnel can the differential conductance be directly compared with the spectral function that is independent of the applied bias. In (**a**) ten different values of the bias are plotted from -0.6 to 0.2 V; no difference can be seen. At contact, (**c**), non-equilibrium effects become clearly distinguished in the spectral function, such as splitting of the Kondo peak and increased broadening due to bias-induced decoherence. The spectral function is very different for the six plotted values (from 0.1 to 0.6 V). These effects are however difficult to glean from the differential conductance where the electron transmission contributes for a small energy window for each bias. (Figure modified and adapted from Choi et al. 2017a)

the tip. At contact one single channel dominates with d_{z^2} character. This orbital possesses the largest overlap with the holding substrate, making it responsible for the Kondo signature. Indeed, despite being a $S = 1$ system, Co on Cu(100) is basically a $S = 1/2$ Kondo system due to the single d_{z^2} spin (Choi et al. 2016a; Baruselli et al. 2015; Jacob 2015).

To treat the non-equilibrium Kondo physics when the tip drives the system into the contact regime, Choi et al. (2017a) use the non-equilibrium non-crossing approximation (Hettler et al. 1998). In tunneling, the computed conductance is shown to be proportional to the tunneling spectral function. Figures 6a, c depict the electron transmission (proportional to the spectral function for the plotted case) as a function of electron energy and b and d the differential conductance. Clearly the spectral function is proportional to the differential conductance in tunneling, in

agreement with the Tersoff-Hamann picture (Tersoff and Hamann 1985). This is not true in contact, where the spectral function develops two Kondo peaks when the applied bias is larger than the Kondo temperature and the peaks decay by bias-induced decoherence, cases c and d in Fig. 6.

Despite this very different regime, the width of the ZBA conductance is fixed by the equilibrium spectral function, reflecting the increase of the Kondo temperature due to the larger couplings to electrodes in the contact regime. Hence, the experimentally obtained widths are indeed proportional to the Kondo temperature of the system as the point contact increases its atomic coordination (Choi et al. 2017a).

8 Electron Paramagnetic Resonance

Manassen (1989) and Komeda and Manassen (2008) used GHz excitation frequencies to study single magnetic moments in Si. More recently, other experiments also using STM (Müllegger et al. 2014) showed data that molecular spins could also be addressed in a GHz pulsed STM.

New highly reproducible data with quantitative measurements and explanation recently came available in a low-temperature STM setup (Baumann et al. 2015). Figure 7 shows a simplified scheme of the experiment. Basically, an oscillatory bias is set between tip and sample that matches the energy between two magnetic states of the magnetic atom under study. If the current is spin polarized, they can
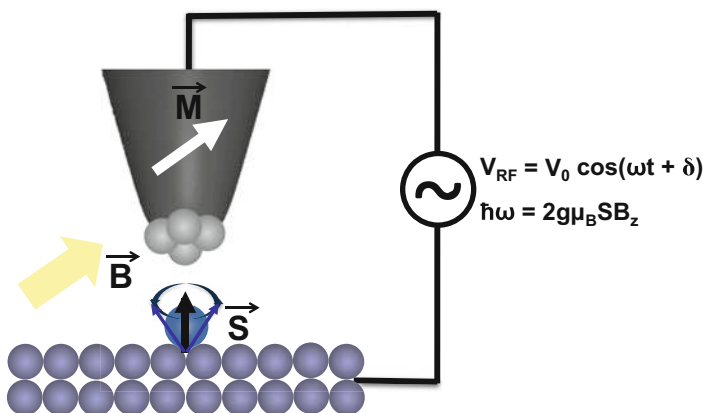
Fig. 7 Scheme with the experimental setup to record EPR spectra over a localized magnetic moment. An STM tip is magnetized with magnetic moment $\vec{M}$ following the direction of the external magnetic field $\vec{B}$. An alternating bias is set between tip and substrate, $V_{RF} = V_0\cos(\omega t + \delta)$, with $\omega = 2\pi f$. When the applied frequency, $\hbar\omega$, matches twice the Zeeman energy of the z-component of the B-field, the resonance condition is attained between the two low-energy levels of the local magnetic spin $\vec{S}$ under the influence of the external field. This translates into a change in the tunneling current as a function of the applied frequency ω, roughly following a Lorentzian shape. Fitting the expression Eq. (8), all the parameters characterizing the EPR signal can be extracted

see a change in the current with the frequency of the alternating bias. They showed compelling evidence that Rabi oscillations were established between two Zeeman-split spin states of an iron atom on an MgO surface when a GHz modulation signal was injected on the electron current. Evidence was given that time-dependent electrons can drive the magnetic oscillations. This is a startling fact that is still under debate: what is the actual physical process that makes currents and spins interact? Baumann and co-workers (Baumann et al. 2015) assume that it is the mechanical deformation of the adsorption configuration that leads to the coupling as induced by the pulsed electric field between tip and sample. Lado et al (2017) have argued that it is rather the long-distance dipole-dipole interaction of the tip with the magnetic moment that drives the EPR signal. Recently, Reina et al. (2018, unpublished) have produced a model that explains many of the experimental features of EPR excitation, and it is based on the shifting of the electronic levels of the adsorbate by the oscillating field. When electrons tunnel in and out of the adsorbate, the levels are subjected to the oscillating electric field of the tip. This physical mechanism is the Anderson mechanism for producing the Kondo antiferromagnetic coupling of the s-d exchange Hamiltonian. It is then very efficient in coupling currents and spins.

The density matrix formalism is particularly well adapted for the study of EPR (Delgado and Fernández-Rossier 2017). Let us assume that we have two clear spin states that can be connected via an oscillating driving term (the electric field between tip and sample in the STM setup). Then the density matrix contains three independent terms, namely, the population of each state and a cross term between the two states. This cross term is called *the coherence* because it is a direct measurement of how coherent the two spin states are. When this term goes to zero, the two states are totally disconnected, and there is no connection between the states. The states are then incoherent and their respective phases randomly change in time. If the phase is constant (or almost constant) in time, the density-matrix cross term is not zero, and the two spin states are coherent in time. They will interfere and commonly evolve in time.

Under this situation there are clearly three time-related quantities that contain all the physics of the coherent excitation of spin via an oscillatory driving stimulus. The lifetime of each state is clearly one of the time scales. If we assume we only have one excited state (as we said above), then there is only one lifetime and it is called T_1. The other obvious lifetime is the pure dephasing time, or the time when the relative phase between the two spin states is constant. This is generally called T_2^* (please be aware that in NV centers a different notation is followed). However, the phase can also disappear because the excited state has disappeared. Then the time regulating the evolution will rather be the complete decoherence time, or T_2. Since we are studying the limitation of the phase due to either pure dephasing or population decay, the simple approach is to assume that the inverse times or rates add up linearly:

$$\frac{1}{T_2} = \frac{1}{T_1} + \frac{1}{T_2^*}.$$

(7)

This implies that the rate of disappearance of the phase is the sum of the rates of disappearance due to the destruction of the state T_1 and the destruction of the relative phase T_2^*.

The third magnitude that enters the density matrix dynamics is the Rabi oscillation frequency, Ω. This frequency is a direct measuring of the strength of the coupling between the external driving force and the reaction of the system to that external field. There is a fourth time that we have not mentioned before, because it is self-evident; this is the time associated to the energy scale of the system given by the energy of the excited spin state, f_0 (such that the excitation energy is $E = hf_0$ where h is Planck's constant).

All of the above terms give rise to a particular evolution of the density matrix. The experiments are performed with spin-polarized currents, so the STM can read the change of spin state. Hence, we can assume that the change in current due to the resonant excitation of the spin state will be proportional to the change of density matrix from the equilibrium situation where the spin state is not excited. In this case, the classical Bloch equations for magnetic moments are retrieved, and the line shape with oscillation frequency f, found in the experiments, is reproduced (Delgado and Fernández-Rossier 2017):

$$\frac{\Delta I}{I}(f) = -\frac{\Omega^2 T_1 T_2}{1 + (f - f_0)^2 T_2^2 + \Omega^2 T_1 T_2}. \tag{8}$$

The experimental data are being produced at fast pace and with outstanding quality. Measurement of Cu adsorbates on MgO shows the distinct hyperfine structure of the Cu nuclear magnetic moment. The STM is now capable of looking into the magnetic structure of a single atomic nucleus (Heinrich, 2017, personal communication). More new exciting data are sure to be produced as we write these lines.

9   Conclusions

The study of individual magnetic impurities on surfaces is revolutionizing the field of electron correlations in magnetic impurities, thanks to the advent of scanning probes. The STM permits a complete characterization of the Kondo effect, giving relevant and complementary information on the spectral features of the Kondo resonance, as well as its spatial distribution. The STM is flexible enough to be able to tackle different conductance regimes, permitting us to have new insight on the non-equilibrium Kondo effect. Together with the atomic manipulation abilities of the STM, the Kondo effect is suggesting new possibilities in the creation of nanostructures with specific electronic properties. In this chapter, we have seen how by creating nanostructures with different species we can tune the degree of spin entanglement. This permits us to reach very different Kondo regimes as shown by the tunneling differential conductance.

New techniques are quickly evolving. Recently, the time domain is yielding new capabilities to the STM. As a consequence, we have now EPR signals on localized magnetic moments that will complete our knowledge on the impurity problem. Moreover, pump-and-probe techniques, together with photon excitation, will be very interesting techniques to probe the extraordinary correlations due to magnetic moments.

References

Agraït N, Rodrigo JG, Vieira S (1993) Conductance steps and quantization in atomic-size contacts. Phys Rev B 47(18):12345. https://doi.org/10.1103/PhysRevB.47.12345

Alducin M, Sánchez-Portal D, Arnau A, Lorente N (2010) Mixed-valency signature in vibrational inelastic electron tunneling spectroscopy. Phys Rev Lett 104:136101. https://link.aps.org/doi/10.1103/PhysRevLett.104.136101

Anderson PW (1961) Localized magnetic states in metals. Phys Rev 124:41–53. https://link.aps.org/doi/10.1103/PhysRev.124.41

Balatsky AV, Nishijima M, Manassen Y (2012) Electron spin resonance-scanning tunneling microscopy. Adv Phys 61(2):117–152. https://doi.org/10.1080/00018732.2012.668775

Baruselli PP, Requist R, Smogunov A, Fabrizio M, Tosatti E (2015) Co adatoms on cu surfaces: ballistic conductance and Kondo temperature. Phys Rev B 92:045119. http://link.aps.org/doi/10.1103/PhysRevB.92.045119

Baumann S, Paul W, Choi T, Lutz CP, Ardavan A, Heinrich AJ (2015) Electron paramagnetic resonance of individual atoms on a surface. Science 350(6259):417–420. https://doi.org/10.1126/science.aac8703. http://science.sciencemag.org/content/350/6259/417

Bork J, hui Zhang Y, Diekhöner L, Borda L, Simon P, Kroha J, Wahl P, Kern K (2011) A tunable two-impurity Kondo system in an atomic point contact. Nat Phys 7:901–906

Bulla R, Costi TA, Pruschke T (2008) Numerical renormalization group method for quantum impurity systems. Rev Mod Phys 80:395–450. https://link.aps.org/doi/10.1103/RevModPhys.80.395

Calvo MR, Fernandez-Rossier J, Palacios JJ, Jacob D, Natelson D, Untiedt C (2009) The Kondo effect in ferromagnetic atomic contacts. Nature 458(7242):1150–1153. https://doi.org/10.1038/nature07878

Choi DJ, Rastei MV, Simon P, Limot L (2012) Conductance-driven change of the Kondo effect in a single cobalt atom. Phys Rev Lett 108:266803. https://link.aps.org/doi/10.1103/PhysRevLett.108.266803

Choi DJ, Guissart S, Ormaza M, Bachellier N, Bengone O, Simon P, Limot L (2016a) Kondo resonance of a Co atom exchange coupled to a ferromagnetic tip. Nano Lett 16:6298. https://doi.org/10.1021/acs.nanolett.6b02617

Choi DJ, Robles R, Gauyacq JP, Ternes M, Loth S, Lorente N (2016b) Structural and magnetic properties of FeMn$_x$ chains ($x = 1$–6) supported on Cu$_2$N/Cu (100). Phys Rev B 94:085406. https://link.aps.org/doi/10.1103/PhysRevB.94.085406

Choi DJ, Abufager P, Limot L, Lorente N (2017a) From tunneling to contact in a magnetic atom: the non-equilibrium Kondo effect. J Chem Phys 146(9):092309. https://doi.org/10.1063/1.4972874

Choi DJ, Robles R, Yan S, Burgess JAJ, Rolf-Pissarczyk S, Gauyacq JP, Lorente N, Ternes M, Loth S (2017b) Building complex Kondo impurities by manipulating entangled spin chains. Nano Lett 17(10):6203–6209. https://doi.org/10.1021/acs.nanolett.7b02882, pMID:28872317

Cronenwett SM, Oosterkamp TH, Kouwenhoven LP (1998) A tunable Kondo effect in quantum dots. Science 281(5376):540–544. http://www.sciencemag.org/cgi/content/abstract/281/5376/540

Delgado F, Fernández-Rossier J (2017) Spin decoherence of magnetic atoms on surfaces. Prog Surf Sci 92(1):40–82. https://doi.org/10.1016/j.progsurf.2016.12.001, http://www.sciencedirect.com/science/article/pii/S0079681616300351

Delgado F, Palacios JJ, Fernández-Rossier J (2010) Spin-transfer torque on a single magnetic adatom. Phys Rev Lett 104:026601. https://link.aps.org/doi/10.1103/PhysRevLett.104.026601

DiLullo A, Chang SH, Baadji N, Clark K, Klckner JP, Prosenc MH, Sanvito S, Wiesendanger R, Hoffmann G, Hla SW (2012) Molecular Kondo chain. Nano Lett 12(6):3174–3179. https://doi.org/10.1021/nl301149d

Fano U (1961) Effects of configuration interaction on intensities and phase shifts. Phys Rev 124:1866–1878. https://link.aps.org/doi/10.1103/PhysRev.124.1866

Fernández J, Moro-Lagares M, Serrate D, Aligia AA (2016) Manipulation of the surface density of states of ag(111) by means of resonators: experiment and theory. Phys Rev B 94:075408. https://link.aps.org/doi/10.1103/PhysRevB.94.075408

Frank S, Jacob D (2015) Orbital signatures of Fano-Kondo line shapes in STM adatom spectroscopy. Phys Rev B 92:235127. http://link.aps.org/doi/10.1103/PhysRevB.92.235127

Fulde P (1995) Electron correlations in molecules and solids. Springer, Berlin

Gai Z, He Y, Yu H, Yang WS (1996) Observation of conductance quantization of ballistic metallic point contacts at room temperature. Phys Rev B 53(3):1042. https://doi.org/10.1103/PhysRevB.53.1042

Gaudenzi R, de Bruijckere J, Reta D, Moreira IdPR, Rovira C, Veciana J, van der Zant HSJ, Burzurí E (2017) Redox-induced gating of the exchange interactions in a single organic diradical. ACS Nano 11(6):5879–5883. https://doi.org/10.1021/acsnano.7b01578, pMID:28494146

Gauyacq JP, Lorente N, Novaes FD (2012) Excitation of local magnetic moments by tunneling electrons. Prog Surf Sci 87(5–8):63–107. https://doi.org/10.1016/j.progsurf.2012.05.003, http://linkinghub.elsevier.com/retrieve/pii/S0079681612000068

Girovsky J, Nowakowski J, Ali ME, Baljozovic M, Rossmann HR, Nijs T, Aeby EA, Nowakowska S, Siewert D, Srivastava G, Wäckerlin C, Dreiser J, Decurtins S, Liu SX, Oppeneer PM, Jung TA, Ballav N (2017) Long-range ferrimagnetic order in a two-dimensional supramolecular Kondo lattice. Nat Commun 8:15388. https://doi.org/10.1038/ncomms15388

Goldhaber-Gordon D, Shtrikman H, Mahalu D, Abusch-Magder D, Meirav U, Kastner MA (1998) Kondo effect in a single-electron transistor. Nature 391(6663):156–159. https://doi.org/10.1038/34373

Heinrich AJ, Gupta JA, Lutz CP, Eigler DM (2004) Single-atom spin-flip spectroscopy. Science 306(5695):466–469. https://doi.org/10.1126/science.1101077, http://science.sciencemag.org/content/306/5695/466

Hettler MH, Kroha J, Hershfield S (1998) Nonequilibrium dynamics of the anderson impurity model. Phys Rev B 58:5649–5664. http://link.aps.org/doi/10.1103/PhysRevB.58.5649

Hewson A (1993) The Kondo problem to heavy fermions. Cambridge University Press, Cambridge

Hipps KW, Mazur U (2006) Inelastic electron tunneling spectroscopy. https://onlinelibrary.wiley.com/doi/abs/10.1002/0470027320.s0605

Hirjibehedin CF, Lutz CP, Heinrich AJ (2006) Spin coupling in engineered atomic structures. Science 312(5776):1021–1024

Hirjibehedin CF, Lin CY, Otte AF, Ternes M, Lutz CP, Jones BA, Heinrich AJ (2007) Large magnetic anisotropy of a single atomic spin embedded in a surface molecular network. Science 317(5842):1199–1203. https://doi.org/10.1126/science.1146110, http://www.sciencemag.org/cgi/doi/10.1126/science.1146110

Ho W (2002) Single-molecule chemistry. J Chem Phys 117(24):11033–11061. https://doi.org/10.1063/1.1521153

Hurley A, Baadji N, Sanvito S (2011a) Perturbative approach to the Kondo effect in magnetic atoms on nonmagnetic substrates. Phys Rev B 84:115435. https://link.aps.org/doi/10.1103/PhysRevB.84.115435

Hurley A, Baadji N, Sanvito S (2011b) Spin-flip inelastic electron tunneling spectroscopy in atomic chains. Phys Rev B 84:035427. https://link.aps.org/doi/10.1103/PhysRevB.84.035427

Jacob D (2015) Towards a full ab initio theory of strong electronic correlations in nanoscale devices. J Phys Condens Matter 27(24):245606. https://stacks.iop.org/0953-8984/27/i=24/a=245606

Jensen H, Kröger J, Néel N, Berndt R (2007) Silver oligomer and single fullerene electronic properties revealed by a scanning tunnelling microscope. Eur Phys J D 45(3):465–469. https://doi.org/10.1140/epjd/e2007-00157-x

Jiang Y, Zhang YN, Cao JX, Wu RQ, Ho W (2011) Real-space imaging of Kondo screening in a two-dimensional o2 lattice. Science 333(6040):324–328. https://doi.org/10.1126/science.1205785, http://science.sciencemag.org/content/333/6040/324,

Kim H, Yoshida Y, Lee CC, Chang TR, Jeng HT, Lin H, Haga Y, Fisk Z, Hasegawa Y (2017) Atomic-scale visualization of surface-assisted orbital order. Sci Adv 3(9). https://doi.org/10.1126/sciadv.aao0362, http://advances.sciencemag.org/content/3/9/eaao0362

Komeda T (2005) Chemical identification and manipulation of molecules by vibrational excitation via inelastic tunneling process with scanning tunneling microscopy. Prog Surf Sci 78(2):41–85. https://doi.org/10.1016/j.progsurf.2005.05.001, http://www.sciencedirect.com/science/article/pii/S0079681605000262

Komeda T, Manassen Y (2008) Distribution of frequencies of a single precessing spin detected by scanning tunneling microscope. Appl Phys Lett 92(21):212506. https://doi.org/10.1063/1.2937940

Komeda T, Isshiki H, Liu J, Zhang YF, Lorente N, Katoh K, Breedlove BK, Yamashita M (2011) Observation and electric current control of a local spin in a single-molecule magnet. Nat Commun 2:217. https://doi.org/10.1038/ncomms1210, article

Kondo J (1964) Resistance minimum in dilute magnetic alloys. Prog Theor Phys 32(1):37–49. https://doi.org/10.1143/PTP.32.37

Kondo J (2005) Sticking to my bush. J Phys Soc Japan 74(1):1–3. https://doi.org/10.1143/JPSJ.74.1

Korytár R, Lorente N (2011) Multi-orbital non-crossing approximation from maximally localized Wannier functions: the Kondo signature of copper phthalocyanine on Ag(100). J Phys Condens Matter 23(35):355009. http://stacks.iop.org/0953-8984/23/i=35/a=355009

Kramers HA (1930) Théorie générale de la rotation paramagnétique dans les cristaux. Proceedings Koninklijke Akademie van Wetenschappen 33:959–972

Kröger J (2008) Nonadiabatic effects on surfaces: Kohn anomaly, electronic damping of adsorbate vibrations, and local heating of single molecules. J Phys Condens Matter 20(22):224015. https://doi.org/10.1088/0953-8984/20/22/224015

Kröger J, Jensen H, Berndt R (2007) Conductance of tip-surface and tip-atom junction on Au(111) explored by a scanning tunnelling microscope. New J Phys 9(5):153. https://doi.org/10.1088/1367-2630/9/5/153

Lado JL, Ferrón A, Fernández-Rossier J (2017) Exchange mechanism for electron paramagnetic resonance of individual adatoms. Phys Rev B 96:205420. https://link.aps.org/doi/10.1103/PhysRevB.96.205420

Li J, Schneider WD, Berndt R, Delley B (1998) Kondo scattering observed at a single magnetic impurity. Phys Rev Lett 80:2893–2896. https://link.aps.org/doi/10.1103/PhysRevLett.80.2893

Limot L, Kröger J, Berndt R, Garcia-Lekue A, Hofer WA (2005) Atom transfer and single-adatom contacts. Phys Rev Lett 94(12):126102. https://doi.org/10.1103/PhysRevLett.94.126102

Lorente N, Gauyacq JP (2009) Efficient spin transitions in inelastic electron tunneling spectroscopy. Phys Rev Lett 103:176601. https://link.aps.org/doi/10.1103/PhysRevLett.103.176601

Loth S, Baumann S, Lutz CP, Eigler DM, Heinrich AJ (2012) Bistability in atomic-scale antiferromagnets. Science 335:196–199

Madhavan V, Chen W, Jamneala T, Crommie MF, Wingreen NS (1998) Tunneling into a single magnetic atom: spectroscopic evidence of the Kondo resonance. Science 280(5363):567–569. http://www.sciencemag.org/cgi/content/abstract/280/5363/567

Madhavan V, Chen W, Jamneala T, Crommie MF, Wingreen NS (2001) Local spectroscopy of a Kondo impurity: Co on Au(111). Phys Rev B 64(16):165412. https://doi.org/10.1103/PhysRevB.64.165412

Manassen Y, Hamers RJ, Demuth JE, Castellano Jr AJ (1989) Direct observation of the precession of individual paramagnetic spins on oxidized silicon surfaces. Phys Rev Lett 62:2531–2534. https://link.aps.org/doi/10.1103/PhysRevLett.62.2531

Meissner W, Voigt B (1930) Messungen mit hilfe von flüssigem helium xi widerstand der reinen metalle in tiefen temperaturen. Annalen der Physik 7:761–798

Merino J, Gunnarsson O (2004) Simple model for scanning tunneling spectroscopy of noble metal surfaces with adsorbed Kondo impurities. Phys Rev B 69:115404. https://link.aps.org/doi/10.1103/PhysRevB.69.115404

Morgenstern K, Lorente N, Rieder KH (2013) Controlled manipulation of single atoms and small molecules using the scanning tunnelling microscope. Physica Status Solidi (b) 250(9):1671–1751. https://doi.org/10.1002/pssb.201370558

Morr DK (2017) Theory of scanning tunneling spectroscopy: from Kondo impurities to heavy fermion materials. Rep Prog Phys 80(1):014502. http://stacks.iop.org/0034-4885/80/i=1/a=014502

Mugarza A, Lorente N, Ordejón P, Krull C, Stepanow S, Bocquet ML, Fraxedas J, Ceballos G, Gambardella P (2010) Orbital specific chirality and homochiral self-assembly of achiral molecules induced by charge transfer and spontaneous symmetry breaking. Phys Rev Lett 105:115702. https://link.aps.org/doi/10.1103/PhysRevLett.105.115702

Mugarza A, Robles R, Krull C, Korytár R, Lorente N, Gambardella P (2012) Electronic and magnetic properties of molecule-metal interfaces: transition-metal phthalocyanines adsorbed on Ag(100). Phys Rev B 85:155437. https://link.aps.org/doi/10.1103/PhysRevB.85.155437

Müllegger S, Tebi S, Das AK, Schöfberger W, Faschinger F, Koch R (2014) Radio frequency scanning tunneling spectroscopy for single-molecule spin resonance. Phys Rev Lett 113:133001. https://link.aps.org/doi/10.1103/PhysRevLett.113.133001

Nagaoka K, Jamneala T, Grobis M, Crommie MF (2002) Temperature dependence of a single Kondo impurity. Phys Rev Lett 88:077205. https://link.aps.org/doi/10.1103/PhysRevLett.88.077205

Néel N, Kröger J, Limot L, Berndt R (2007a) Conductance of single atoms and molecules studied with a scanning tunnelling microscope. Nanotechnology 18(4):044027. https://doi.org/10.1088/0957-4484/18/4/044027

Néel N, Kröger J, Limot L, Frederiksen T, Brandbyge M, Berndt R (2007b) Controlled contact to a C_{60} molecule. Phys Rev Lett 98(6):065502. https://doi.org/10.1103/PhysRevLett.98.065502

Néel N, Kröger J, Limot L, Palotas K, Hofer WA, Berndt R (2007c) Conductance and Kondo effect in a controlled single-atom contact. Phys Rev Lett 98(1):016801. https://doi.org/10.1103/PhysRevLett.98.016801

Néel N, Kröger J, Limot L, Berndt R (2008a) Conductance of oriented C_{60} molecules. Nano Lett 8(5):1291. https://doi.org/10.1021/nl073074i

Néel N, Limot L, Kröger J, Berndt R (2008b) Rotation of C_{60} in a single-molecule contact. Phys Rev B 77(12):125431. https://doi.org/10.1103/PhysRevB.77.125431

Néel N, Berndt R, Kröger J, Wehling TO, Lichtenstein AI, Katsnelson MI (2011) Two-site Kondo effect in atomic chains. Phys Rev Lett 107(10):106804. https://doi.org/10.1103/PhysRevLett.107.106804

Novaes FD, Lorente N, Gauyacq JP (2010) Quenching of magnetic excitations in single adsorbates at surfaces: Mn on cun/cu(100). Phys Rev B 82:155401. https://link.aps.org/doi/10.1103/PhysRevB.82.155401

Olesen L, Lægsgaard E, Stensgaard I, Besenbacher F, Schiøtz J, Stoltze P, Jacobsen KW, Nørskov JK (1994) Quantized conductance in an atom-sized point contact. Phys Rev Lett 72(14):2251. https://doi.org/10.1103/PhysRevLett.72.2251

Ormaza M, Bachellier N, Faraggi MN, Verlhac B, Abufager P, Ohresser P, Joly L, Romeo M, Scheurer F, Bocquet ML, Lorente N, Limot L (2017) Efficient spin-flip excitation of a nickelocene molecule. Nano Lett 17(3):1877–1882. https://doi.org/10.1021/acs.nanolett.6b05204, pMID:28199115

Park J, Pasupathy AN, Goldsmith JI, Chang C, Yaish Y, Petta JR, Rinkoski M, Sethna JP, Abruňa HD, McEuen PL, Ralph DC (2002) Coulomb blockade and the Kondo effect in single-atom transistors. Nature 417:722–725. https://doi.org/10.1038/nature00791

Parks JJ, Champagne AR, Hutchison GR, Flores-Torres S, Abruňa HD, Ralph DC (2007) Tuning the Kondo effect with a mechanically controllable break junction. Phys Rev Lett 99:026601. https://link.aps.org/doi/10.1103/PhysRevLett.99.026601

Parks JJ, Champagne AR, Costi TA, Shum WW, Pasupathy AN, Neuscamman E, Flores-Torres S, Cornaglia PS, Aligia AA, Balseiro CA, Chan GKL, Abruňa HD, Ralph DC (2010) Mechanical control of spin states in spin-1 molecules and the underscreened Kondo effect. Science 328(5984):1370. https://doi.org/10.1126/science.1186874

Pascual JI, Méndez J, Gómez-Herrero J, Baró AM, García N, Binh VT (1993) Quantum contact in gold nanostructures by scanning tunneling microscopy. Phys Rev Lett 71(12):1852. https://doi.org/10.1103/PhysRevLett.71.1852

Persson M (2009) Theory of inelastic electron tunneling from a localized spin in the impulsive approximation. Phys Rev Lett 103:050801. https://link.aps.org/doi/10.1103/PhysRevLett.103.050801

Pla JJ, Tan KJ, Dehollain Jp, Lim WH, Morton JJL, Jamieson DN, Dzurak AS, Morello A (2012) A single-atom electron spin qubit in silicon. Nature 489:541–545

Plihal M, Gadzuk JW (2001) Nonequilibrium theory of scanning tunneling spectroscopy via adsorbate resonances: nonmagnetic and Kondo impurities. Phys Rev B 63:085404. https://link.aps.org/doi/10.1103/PhysRevB.63.085404

Pustilnik M, Glazman LI (2001) Kondo effect in real quantum dots. Phys Rev Lett 87:216601. https://link.aps.org/doi/10.1103/PhysRevLett.87.216601

Robles R, Lorente N, Isshiki H, Liu J, Katoh K, Breedlove BK, Yamashita M, Komeda T (2012) Spin doping of individual molecules by using single-atom manipulation. Nano Lett 12(7):3609–3612. https://doi.org/10.1021/nl301301e, pMID:22642450

Sarachik MP, Corenzwit E, Longinotti LD (1964) Resistivity of mo-nb and mo-re alloys containing 1% fe. Phys Rev 135:A1041–A1045. https://link.aps.org/doi/10.1103/PhysRev.135.A1041

Schrieffer JR, Wolff PA (1966) Relation between the anderson and kondo hamiltonians. Phys Rev 149:491–492. https://link.aps.org/doi/10.1103/PhysRev.149.491

Spinelli A, Bryant B, Delgado F, Fernández-Rossier J, Otte AF (2014) Imaging of spin waves in atomically designed nanomagnets. Nat Mater 13(8):782–785. https://doi.org/10.1038/nmat4018

Ternes M (2015) Spin excitations and correlations in scanning tunneling spectroscopy. New J Phys 17(6):063016. http://stacks.iop.org/1367-2630/17/i=6/a=063016

Ternes M (2017) Probing magnetic excitations and correlations in single and coupled spin systems with scanning tunneling spectroscopy. Prog Surf Sci 92(1):83–115. https://doi.org/10.1016/j.progsurf.2017.01.001, http://www.sciencedirect.com/science/article/pii/S0079681617300011

Ternes M, Heinrich AJ, Schneider WD (2009) Spectroscopic manifestations of the Kondo effect on single adatoms. J Phys Condens Matt 21(5):053001. http://iopscience.iop.org/0953-8984/21/5/053001

Tersoff J, Hamann DR (1985) Theory of the scanning tunneling microscope. Phys Rev B 31:805–813. https://link.aps.org/doi/10.1103/PhysRevB.31.805

Toskovic R, van den Berg R, Spinelli A, Eliens IS, van den Toorn B, Bryant B, Caux JS, Otte AF (2016) Atomic spin-chain realization of a model for quantum criticality. Nat Phys 6(3):187–191. https://doi.org/10.1038/nphys3722

Tsukahara N, Shiraki S, Itou S, Ohta N, Takagi N, Kawai M (2011) Evolution of kondo resonance from a single impurity molecule to the two-dimensional lattice. Phys Rev Lett 106:187201. https://link.aps.org/doi/10.1103/PhysRevLett.106.187201

van Sciver SW (2012) Helium cryogenics. Springer, New York

Varma CM, Yafet Y (1976) Magnetic susceptibility of mixed-valence rare-earth compounds. Phys Rev B 13:2950–2954. https://link.aps.org/doi/10.1103/PhysRevB.13.2950

Vitali L, Ohmann R, Stepanow S, Gambardella P, Tao K, Huang R, Stepanyuk VS, Bruno P, Kern K (2008) Kondo effect in single atom contacts: the importance of the atomic geometry. Phys Rev Lett 101(21):216802. https://doi.org/10.1103/PhysRevLett.101.216802

Wahl P, Simon P, Diekhöner L, Stepanyuk VS, Bruno P, Schneider MA, Kern K (2007) Exchange interaction between single magnetic adatoms. Phys Rev Lett 98(5):056601. http://link.aps.org/abstract/PRL/v98/e056601

Yanson IK, Fisun VV, Hesper R, Khotkevich AV, Krans JM, Mydosh JA, van Ruitenbeek JM (1995) Size dependence of Kondo scattering in point contacts. Phys Rev Lett 74:302–305. https://link.aps.org/doi/10.1103/PhysRevLett.74.302

Yosida K (1966) Bound state due to the $s-d$ exchange interaction. Phys Rev 147:223–227. https://link.aps.org/doi/10.1103/PhysRev.147.223

Zener C (1951) Interaction between the d shells in the transition metals. Phys Rev 81:440–444. https://link.aps.org/doi/10.1103/PhysRev.81.440

First-Principles Quantum Transport Modeling of Spin-Transfer and Spin-Orbit Torques in Magnetic Multilayers

23

Branislav K. Nikolić, Kapildeb Dolui, Marko D. Petrović, Petr Plecháč, Troels Markussen, and Kurt Stokbro

Contents

Abstract

A unified approach for computing (*i*) spin-transfer torque in magnetic trilayers like spin valve and magnetic tunnel junction, where injected charge current flows perpendicularly to interfaces, and (*ii*) spin-orbit torque in magnetic bilayers of the type ferromagnet/spin-orbit-coupled material, where injected charge current flows parallel to the interface, was reviewed. The experimentally explored and technologically relevant spin-orbit-coupled materials include 5*d* heavy metals, topological insulators, Weyl semimetals, and transition metal dichalcogenides. This approach requires to construct the torque operator for a given Hamiltonian of the device and the steady-state nonequilibrium density matrix, where the latter is expressed in terms of the nonequilibrium Green's functions and split into three

B. K. Nikolić (✉) · K. Dolui
Department of Physics and Astronomy, University of Delaware, Newark, DE, USA
e-mail: bnikolic@udel.edu

M. D. Petrović · P. Plecháč
Department of Mathematical Sciences, University of Delaware, Newark, DE, USA

T. Markussen · K. Stokbro
Synopsys QuantumWise, Copenhagen, Denmark

contributions. Tracing these contributions with the torque operator automatically yields field-like and damping-like components of spin-transfer torque or spin-orbit torque vector, which is particularly advantageous for spin-orbit torque where the direction of these components depends on the unknown-in-advance orientation of the current-driven nonequilibrium spin density in the presence of spin-orbit coupling. Illustrative examples are provided by computing spin-transfer torque in a one-dimensional toy model of a magnetic tunnel junction and realistic Co/Cu/Co spin valve, both of which are described by first-principles Hamiltonians obtained from noncollinear density functional theory calculations, as well as by computing spin-orbit torque in a ferromagnetic layer described by a tight-binding Hamiltonian which includes spin-orbit proximity effect within ferromagnetic monolayers assumed to be generated by the adjacent monolayer transition metal dichalcogenide. In addition, it is shown here how spin-orbit proximity effect, quantified by computing (via first-principles retarded Green's function) spectral functions and spin textures on monolayers of realistic ferromagnetic material like Co in contact with heavy metal or monolayer transition metal dichalcogenide, can be tailored to enhance the magnitude of spin-orbit torque. Errors made in the calculation of spin-transfer torque are quantified when using Hamiltonian from collinear density functional theory, with rigidly rotated magnetic moments to create noncollinear magnetization configurations, instead of proper (but computationally more expensive) self-consistent Hamiltonian obtained from noncollinear density functional theory.

1   What Is Spin Torque and Why Is It Useful?

The spin-transfer torque (STT) (Ralph and Stiles 2008; Locatelli et al. 2014; Slonczewski 1996; Berger 1996) is a phenomenon in which a spin current of sufficiently large density ($\sim 10^7$ A/cm^2) injected into a ferromagnetic metal (FM) (For easy navigation, provided is a list of abbreviations used throughout the chapter: 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional; BZ, Brillouin zone; CD, current-driven; DFT, density functional theory; DL, damping-like; FL, field-like; FM, ferromagnetic metal; GF, Green's function; HM, heavy metal; I, insulator; KS, Kohn-Sham; LCAO, linear combination of atomic orbitals; LLG, Landau-Lifshitz-Gilbert; ML, monolayer; MLWF, maximally localized Wannier function; MRAM, magnetic random access memory; MTJ, magnetic tunnel junction; ncDFT, noncollinear DFT; NEGF, nonequilibrium Green's function; NM, normal metal; PBE, Perdew-Burke-Ernzerhof; scf, self-consistent field; SHE, spin Hall effect; SOC, spin-orbit coupling; SOT, spin-orbit torque; STT, spin-transfer torque; TBH, tight-binding Hamiltonian; TI, topological insulator; TMD, transition metal dichalcogenide; WSM, Weyl semimetal; XC, exchange-correlation.) either switches its magnetization from one static configuration to another or generates a dynamical situation with steady-state precessing magnetization (Locatelli et al. 2014). The origin of STT is the absorption of the itinerant flow of spin angular momentum component normal to the magnetization direction. Figure 1a illustrates a setup with
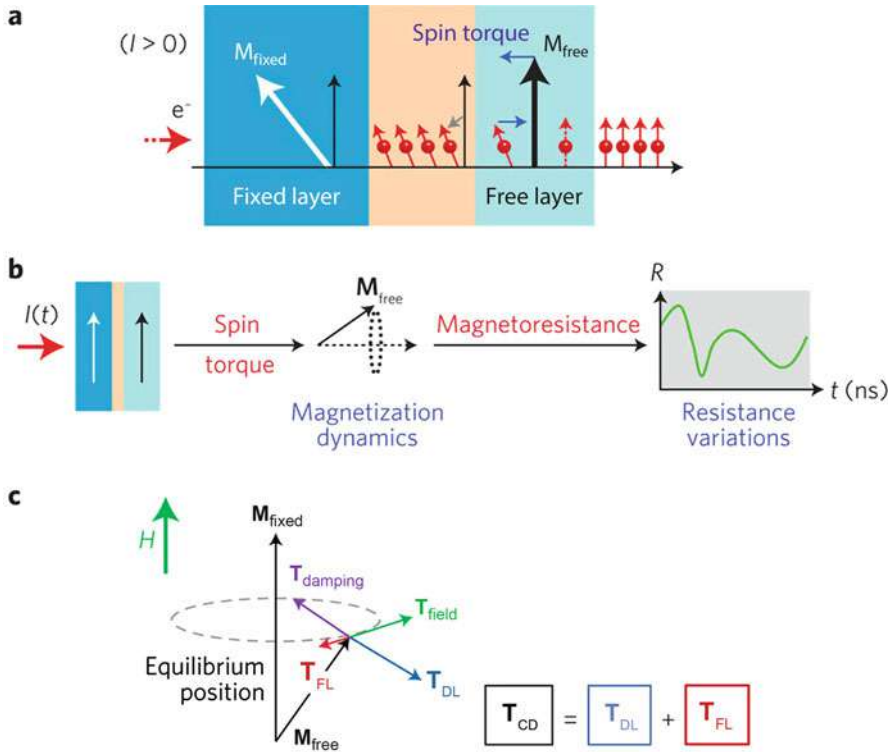
Fig. 1 Spin-transfer torque basics: (**a**) in a ferromagnet/nonmagnetic-spacer/ferromagnet setup, with noncollinear magnetizations $\mathbf{M}_{\mathrm{fixed}}$ of the fixed FM layer and $\mathbf{M}_{\mathrm{free}}$ of the free FM layer, the transverse spin component of the conduction electrons (red) polarized in the direction of $\mathbf{M}_{\mathrm{fixed}}$ is absorbed as they pass through the free layer, thereby generating a torque on $\mathbf{M}_{\mathrm{free}}$; (**b**) device applications of STT in (**a**) are based on torque-induced magnetization dynamics that is converted into resistance variations via the magnetoresistive effects; (**c**) torques on $\mathbf{M}_{\mathrm{free}}$ where field-like component of current-driven STT, $\mathbf{T}_{\mathrm{FL}}$, is orthogonal to the plane spanned by $\mathbf{M}_{\mathrm{fixed}}$ and $\mathbf{M}_{\mathrm{free}}$ and competes with the effective-field torque $\mathbf{T}_{\mathrm{field}}$ (present also in equilibrium), while damping-like component of current-driven STT, $\mathbf{T}_{\mathrm{DL}}$, is parallel or antiparallel (depending on the current direction) to Gilbert damping torque $\mathbf{T}_{\mathrm{damping}}$ (present also in equilibrium). The illustration in (**c**) assumes particular case where $\mathbf{M}_{\mathrm{fixed}}$ and the effective magnetic field are aligned. (Adapted from Locatelli et al. 2014)

two noncollinear magnetizations which generates STT. This setup can be realized as FM/NM/FM (NM, normal metal) spin valve, exemplified by Co/Cu/Co trilayer in Fig. 2a and employed in early experiments (Tsoi et al. 1998; Myers et al. 1999; Katine et al. 2000), or FM/I/FM (I, insulator) magnetic tunnel junctions (MTJs), exemplified by Fe/MgO/Fe trilayer and employed in later experiments (Sankey et al. 2008; Kubota et al. 2008; Wang et al. 2011) and recent applications (Locatelli et al. 2014; Kent and Worledge 2015). In such magnetic multilayers, injected unpolarized charge current passes through the first thin FM layer to become spin-polarized
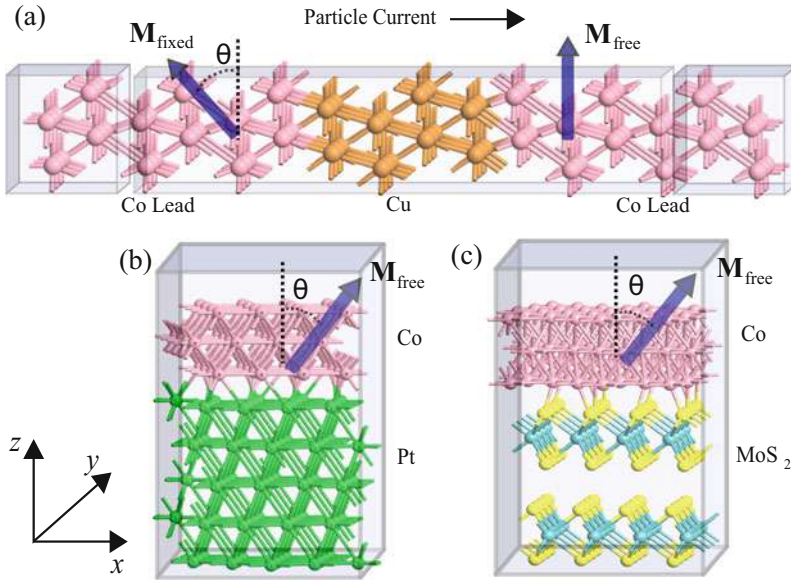
Fig. 2 Schematic view of (**a**) FM/NM/FM trilayer for calculations of STT in spin valves, (**b**) FM/HM bilayer for calculations of SOT in the presence of the spin Hall current along the z-axis generated by the HM layer, and (**c**) FM/monolayer-TMD for calculations of SOT in the absence of any spin Hall current. The semi-infinite FM layers in (**a**) are chosen as Co(0001), and the spacer in between consists of l ($l = 4$ in the illustration) monolayers of Cu(111). The trilayer in (**a**) is assumed to be infinite in the transverse directions to the current flow, so that the depicted supercell is periodically repeated within the yz-plane. The bilayer in (**b**) consists of Co(0001) and Pt(111), and in (**c**) it consists of Co(0001) and monolayer MoS_2. The bilayers in (**b**) and (**c**) are assumed to be infinite in the xy-plane but of finite thickness along the z-axis. Small bias voltage V_b is applied to inject electrons along the positive x-axis, so that particle current is perpendicular to interfaces in (**a**) and parallel to the interface in (**b**) and (**c**)

in the direction of its fixed magnetization along the unit vector $\mathbf{M}_{fixed}$, and it is directed into the second thin FM layer with magnetization along the unit vector $\mathbf{M}_{free}$ where transverse (to $\mathbf{M}_{free}$) component of flowing spins is absorbed. The STT-induced magnetization dynamics is converted into resistance variations via the magnetoresistive effect, as illustrated in Fig. 1b, which is much larger in MTJ than in spin valves. The rich nonequilibrium physics arising in the interplay of spin currents carried by fast conduction electrons, described quantum mechanically, and slow collective magnetization dynamics, described by the Landau-Lifshitz-Gilbert (LLG) equation which models magnetization as a classical vector subject to thermal fluctuations (Berkov and Miltat 2008; Evans et al. 2014; Petrović et al. 2018), is also of great fundamental interest. Note that at cryogenic temperatures, where thermal fluctuations are suppressed, quantum-mechanical effects in STT-driven magnetization dynamics can also be observed (Zholud et al. 2017; Mahfouzi and Kioussis 2017; Mahfouzi et al. 2017).

Another setup exhibiting current-induced magnetization dynamics is illustrated in Fig. 2b, c. It utilizes a single FM layer, so that the role of polarizing FM layer with M_{fixed} in Figs. 1a and 2a is taken over by strong spin-orbit coupling (SOC) introduced by heavy metals (HMs) (Miron et al. 2011; Liu et al. 2012) (such as $5d$ metals Pt, W, and Ta) as in Fig. 2b, topological insulators (TIs) (Mellnik et al. 2014; Fan et al. 2014; Han et al. 2017; Wang et al. 2017), Weyl semimetals (WSMs) (MacNeill et al. 2017a, b), and atomically thin transition metal dichalcogenides (TMDs) (Sklenar et al. 2016; Shao et al. 2016; Guimarães et al. 2018; Lv et al. 2018). The TMDs are compounds of the type MX_2 (M = Mo, W, Nb; X = S, Se, Te) where one layer of M atoms is sandwiched between two layers of X atoms, as illustrated by monolayer MoS_2 in Fig. 2c. The SOC is capable of converting charge into spin currents (Vignale 2010; Sinova et al. 2015; Soumyanarayanan et al. 2016), so that their absorption by the FM layer in Fig. 2b, c leads to the so-called spin-orbit torque (SOT) (Manchon et al. 2019) on its free magnetization M_{free}.

The current-driven (CD) STT and SOT vectors are analyzed by decomposing them into two contributions, $T_{CD} = T_{DL} + T_{FL}$, commonly termed (Ralph and Stiles 2008; Manchon et al. 2019) damping-like (DL) and field-like (FL) torque based on how they enter into the LLG equation describing the classical dynamics of magnetization. As illustrated in Fig. 1c, these two torque components provide two different handles to manipulate the dynamics of M_{free}. In the absence of current, displacing M_{free} out of its equilibrium position leads to the effective-field torque T_{field} which drives M_{free} into precession around the effective magnetic field, while Gilbert damping $T_{damping}$ acts to bring it back to its equilibrium position. Under nonequilibrium conditions, brought by injecting steady-state or pulse current (Baumgartner et al. 2017), T_{DL} acts opposite to $T_{damping}$ for "fixed-to-free" current direction, and it enhances $T_{damping}$ for "free-to-fixed" current directions in Fig. 1a. Thus, the former (latter) acts as antidamping (overdamping) torque trying to bring M_{free} antiparallel (parallel) to M_{fixed} (note that at cryogenic temperatures, one finds apparently only antidamping action of T_{DL} for both current directions (Zholud et al. 2017)). The T_{FL} component induces magnetization precession and modifies the energy landscape seen by M_{free}. Although $|T_{FL}|$ is minuscule in metallic spin valves (Wang et al. 2008), it can reach 30–40% of $|T_{DL}|$ in MTJs (Sankey et al. 2008; Kubota et al. 2008), and it can become several times larger than $|T_{DL}|$ in FM/HM bilayers (Kim et al. 2013; Yoon et al. 2017). Thus, T_{FL} component of SOT can play a crucial role (Baumgartner et al. 2017; Yoon et al. 2017) in triggering the reversal process of M_{free} and in enhancing the switching efficiency. In concerted action with T_{DL} and possible other effects brought by interfacial SOC, such as the Dzyaloshinskii-Moriya interaction (Perez et al. 2014), this can also lead to complex inhomogeneous magnetization switching patterns observed in SOT-operated devices (Baumgartner et al. 2017; Yoon et al. 2017; Perez et al. 2014).

By adjusting the ratio $|T_{DL}|/|T_{FL}|$ (Timopheev et al. 2015) via tailoring of material properties and device shape, as well as by tuning the amplitude and duration of the injected pulse current (Baumgartner et al. 2017), both STT- and SOT-operated devices can implement variety of functionalities, such as nonvolatile magnetic random access memories (MRAMs) of almost unlimited endurance,

microwave oscillators, microwave detectors, spin-wave emitters, memristors, and artificial neural networks (Locatelli et al. 2014; Kent and Worledge 2015; Borders et al. 2017). The key goal in all such applications is to actively manipulate magnetization dynamics, without the need for external magnetic fields that are incompatible with downscaling of the device size, while using the smallest possible current [e.g., writing currents $\lesssim 20\,\mu A$ would enable multigigabit MRAM (Kent and Worledge 2015)] and energy consumption. For example, recent experiments (Wang et al. 2017) have demonstrated current-driven magnetization switching at room temperature in FM/TI bilayers using current density $\sim 10^5$ A/cm^2, which is two orders of magnitude smaller than for STT-induced magnetization switching in MTJs or one to two orders of magnitude smaller than for SOT-induced magnetization switching in FM/HM bilayers. The SOT-MRAM is expected to be less affected by damping, which offers flexibility for choosing the FM layer, while it eliminates insulating barrier in the writing process and its possible dielectric breakdown in STT-MRAM based on MTJs (Kent and Worledge 2015). Also, symmetric switching profile of SOT-MRAM evades the asymmetric switching issues in STT-MRAM otherwise requiring additional device/circuit engineering. On the other hand, SOT-MRAM has a disadvantage of being a three-terminal device.

2   How to Model Spin Torque Using Nonequilibrium Density Matrix Combined with Density Functional Theory Calculations

The absorption of the component of flowing spin angular momentum that is transverse to $\mathbf{M}_{\text{free}}$, as illustrated in Fig. 1a, occurs (Stiles and Zangwill 2002; Wang et al. 2008) within a few ferromagnetic monolayers (MLs) near NM/FM or I/FM interface. Since the thickness of this interfacial region is typically shorter (Wang et al. 2008) than any charge or spin dephasing length that would make electronic transport semiclassical, STT *requires* quantum transport modeling (Brataas et al. 2006). The essence of STT can be understood using simple one-dimensional (1D) models solved by matching spin-dependent wave function across the junction, akin to elementary quantum mechanics problems of transmission and reflection through a barrier, as provided in Ralph and Stiles (2008), Manchon et al. (2008) and Xiao et al. (2008). However, to describe details of experiments, such as bias voltage dependence of STT in MTJs (Kubota et al. 2008; Sankey et al. 2008) or complex angular dependence of SOT in FM/HM bilayers (Garello et al. 2013), more involved calculations are needed employing tight-binding or first-principles Hamiltonian as an input. For example, simplistic tight-binding Hamiltonians (TBHs) with single orbital per site have been coupled (Theodonis et al. 2006) to nonequilibrium Green's function (NEGF) formalism (Stefanucci and van Leeuwen 2013) to compute SOT in FM/HM bilayers (Kalitsov et al. 2017) or bias voltage dependence of DL and FL components of STT in MTJs which can describe some features of the experiments by adjusting the tight-binding parameters (Kubota et al. 2008).

However, not all features of STT experiments on MTJs (Wang et al. 2011) can be captured by such NEGF+TBH approach. Furthermore, due to spin-orbit proximity effect, driven by hybridization of wave functions from FM and HM layers (Dolui and Nikolić 2017) or FM and metallic surfaces of three-dimensional (3D) TIs (Marmolejo-Tejada et al. 2017), simplistic Hamiltonians like the Rashba ferromagnetic model (Manchon and Zhang 2008; Haney et al. 2013; Lee et al. 2015; Li et al. 2015; Pesin and MacDonald 2012a; Ado et al. 2017; Kalitsov et al. 2017) or the gapped Dirac model (Ndiaye et al. 2017) are highly inadequate to describe realistic bilayers employed in SOT experiments. This is emphasized by Fig. 3 which shows how spectral function and spin texture on the surface of semi-infinite Co layer can change dramatically as we change the adjacent layer. For example, nonzero in-plane spin texture on the surface of semi-infinite Co layer in contact with vacuum is found in Fig. 3e, despite Co magnetization being perpendicular to that surface. This is the consequence of the Rashba SOC enabled by inversion symmetry breaking (Chantis et al. 2007) where an electrostatic potential gradient can be created by the charge distribution at the metal/vacuum interface to confine wave functions into a Rashba spin-split quasi two-dimensional (2D) electron gas (Bahramy et al. 2012). The surface of semi-infinite Co within the layer embedded into Co/Cu(9 ML)/Co junction illustrated in Fig. 2a does not have any in-plane spin texture in Fig. 3f since the structure is inversion symmetric, but its spectral function in Fig. 3b is quite different from the one on the surface of an isolated semi-infinite Co layer in Fig. 3a. Bringing semi-infinite Co layer in contact with 5 MLs of Pt or 1 ML of MoS_2 transforms its spectral function from Fig. 3a to the ones in Fig. 3c, d, respectively, while inducing the corresponding spin textures in Fig. 3g, h due to spin-orbit proximity effect signified by the "leakage" of SOC from HM or TMD layer into the FM layer.

Note that the spectral function and spin texture at the Co/Pt interface are quite different from those of the ferromagnetic Rashba Hamiltonian in 2D often employed (Manchon and Zhang 2008; Haney et al. 2013; Lee et al. 2015; Li et al. 2015; Pesin and MacDonald 2012a; Ado et al. 2017; Kalitsov et al. 2017) in the calculations of SOT as the putative simplistic description of the FM/HM interface. When charge current flows within FM monolayer hosting spin textures – such as the ones displayed in Fig. 3e, g, h – more forward-going electron states will be occupied and less the backward-going ones which due to spin-momentum locking leads to nonequilibrium spin density (Edelstein 1990; Aronov and Lyanda-Geller 1989) as one of the principal mechanisms behind SOT (Manchon et al. 2019). The direction of the nonequilibrium spin density is easily identified in the case of simple spin textures, such as the one in Fig. 3e or those associated with simplistic models (Pesin and MacDonald 2012b) like the Rashba Hamiltonian or the Dirac Hamiltonian discussed in Sect. 4. Conversely, for complex spin textures within heterostructures of realistic materials, as exemplified by those in Fig. 3g, h, one needs first principles coupled with electronic transport calculations (Chang et al. 2015; Johansson et al. 2018).

Thus, capturing properties of realistic junctions illustrated by Fig. 3 requires first-principles Hamiltonian as offered by the density functional theory (DFT). In the linear-response regime, appropriate for spin valves or SOT-operated bilayers in
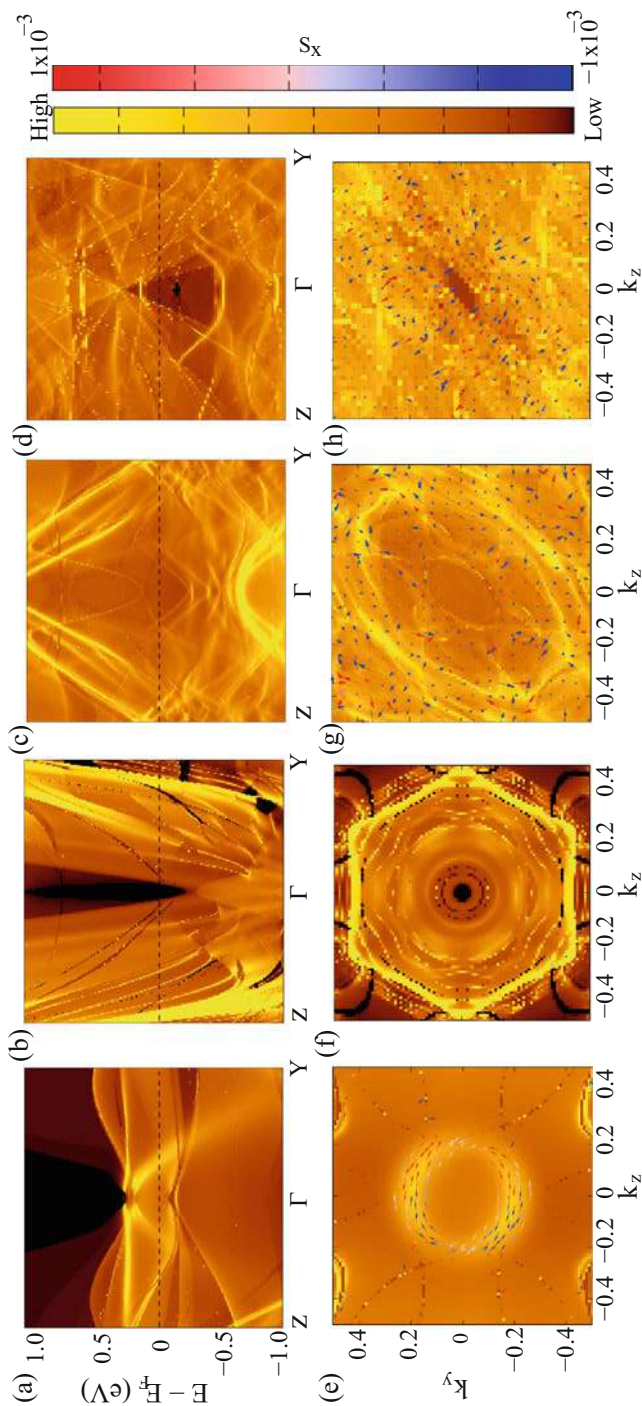
Fig. 3 Spectral function $A(E; \mathbf{k}_{\parallel}, x \in \text{ML of Co})$, defined in Eq. (6), plotted along high-symmetry k-path Z-Γ-Y at (**a**) the surface of semi-infinite Co(0001) in contact with vacuum, (**b**) ML of semi-infinite Co(0001) in contact with 9 MLs of Cu(111) within Co/Cu(9 ML)/Co spin valve, (**c**) ML of semi-infinite Co(0001) in contact with 5 MLs of Pt(111), and (**d**) ML of semi-infinite Co(0001) in contact with 1 ML of MoS$_2$. Panels (**e**)–(**h**) plot constant energy contours of $A(E - E_F = 0; \mathbf{k}_{\parallel}, x \in \text{ML of Co})$ and the corresponding spin textures where the out-of-plane S_x component of spin is indicated in color (red for positive and blue for negative). The magnetization of Co is along the x-axis which is perpendicular to the ML of Co over which the spectral functions and spin textures are scanned. The units for k_y and k_z are $2\pi/a$ and $2\pi/b$ where a and b are the lattice constants along the y- and the z-axis, respectively. The horizontal dashed black line in panels (**a**)–(**d**) denotes the position of the Fermi energy E_F. Panels (**a**) and (**e**) are adapted from Marmolejo-Tejada et al. (2017), and panels (**c**) and (**g**) are adapted from Dolui and Nikolić (2017)

Fig. 2, one can also employ first-principles-derived TBH as offered by transforming the DFT Hamiltonian to a basis of orthogonal maximally localized Wannier functions (MLWFs) in a selected energy window around the Fermi energy E_F. This procedure retains faithfully the overlap matrix elements and their phases, orbital character of the bands, and the accuracy of the original DFT calculations (Marzari et al. 2012). Although Wannier TBH has been used to describe infinite-FM-on-infinite-HM bilayers (Freimuth et al. 2014; Mahfouzi and Kioussis (2018)), its accuracy can be compromised by complicated band entanglement in hybridized metallic systems (Marzari et al. 2012). It is also cumbersome to construct Wannier TBH for junctions in other geometries, like the spin valve in Fig. 2a, or when FM/HM bilayer is attached to leads made of different NM material. In such cases, one needs to perform multiple calculations (Shelley et al. 2011; Thygesen and Jacobsen 2005) (such as on periodic leads, supercell composed of the central region of interest attached to buffer layers of the lead material on both sides, etc.) where one can encounter different MLWFs for two similar but nonidentical systems (Thygesen and Jacobsen 2005), nonorthogonal MLWFs belonging to two different regions (Thygesen and Jacobsen 2005), and Fermi energies of distinct calculations that have to be aligned (Shelley et al. 2011). Also, to compute the current or STT in MTJs at finite bias voltage, one needs to recalculate Hamiltonian in order to take into account self-consistent charge redistribution and the corresponding electrostatic potential in the presence of current flow. Otherwise, without computing them across the device, the current-voltage characteristics violates (Christen and Büttiker 1996; Hernández and Lewenkopf 2013) gauge invariance, i.e., invariance with respect to the global shift of electric potential by a constant, $V \to V + V_0$.

The noncollinear DFT (ncDFT) (Capelle et al. 2001; Eich and Gross 2013; Eich et al. 2013; Bulik et al. 2013) coupled to nonequilibrium density matrix (Stefanucci and van Leeuwen 2013) offers an algorithm to compute spin torque in arbitrary device geometry at vanishing or finite bias voltage. The single-particle spin-dependent Kohn-Sham (KS) Hamiltonian in ncDFT takes the form

$$\hat{H}_{\text{KS}} = -\frac{\hbar^2 \nabla^2}{2m} + V_{\text{H}}(\mathbf{r}) + V_{\text{XC}}(\mathbf{r}) + V_{\text{ext}}(\mathbf{r}) - \boldsymbol{\sigma} \cdot \mathbf{B}_{\text{XC}}(\mathbf{r}), \tag{1}$$

where $V_H(\mathbf{r})$, $V_{\text{ext}}(\mathbf{r})$, and $V_{\text{XC}}(\mathbf{r}) = E_{\text{XC}}[n(\mathbf{r}), \mathbf{m}(\mathbf{r})]/\delta n(\mathbf{r})$ are the Hartree, external, and exchange-correlation (XC) potentials, respectively, and $\boldsymbol{\sigma} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$ is the vector of the Pauli matrices. The extension of DFT to the case of spin-polarized systems is formally derived in terms of total electron density $n(\mathbf{r})$ and vector magnetization density $\mathbf{m}(\mathbf{r})$. In the collinear DFT, $\mathbf{m}(\mathbf{r})$ points in the same direction at all points in space, which is insufficient to study magnetic systems where the direction of the local magnetization is not constrained to a particular axis or systems with SOC. In ncDFT (Capelle et al. 2001), XC functional $E_{\text{XC}}[n(\mathbf{r}), \mathbf{m}(\mathbf{r})]$ depends on $\mathbf{m}(\mathbf{r})$ pointing in arbitrary direction. The XC magnetic field is then given by $\mathbf{B}_{\text{XC}}(\mathbf{r}) = \delta E_{\text{XC}}[n(\mathbf{r}), \mathbf{m}(\mathbf{r})]/\delta \mathbf{m}(\mathbf{r})$.

Once the Hamiltonian of the device is selected, it has to be passed into the formalism of nonequilibrium quantum statistical mechanics. Its central concept is

the density matrix $\boldsymbol{\rho}$ describing quantum many-particle system at finite temperature in equilibrium, or in the presence of external static or time-dependent fields which drive the system out of equilibrium. The knowledge of $\boldsymbol{\rho}$ makes it possible to compute the expectation value of any observable

$$O = \mathrm{Tr}\,[\boldsymbol{\rho}\mathbf{O}], \tag{2}$$

such as the charge density, charge current, spin current, and spin density of interest to spin torque modeling. These require to insert their operators (in some matrix representation) as $\mathbf{O}$ into Eq. (2), where a notation in which bold letters denote matrix representation of an operator in a chosen basis is used. For the KS Hamiltonian in ncDFT in Eq. (1), the torque operator is given by the time derivative of the electronic spin operator (Haney et al. 2007; Carva and Turek 2009)

$$\mathbf{T} = \frac{d\mathbf{S}}{dt} = \frac{1}{2i}[\boldsymbol{\sigma}, \mathbf{H}_{\mathrm{KS}}] = \boldsymbol{\sigma} \times \mathbf{B}_{\mathrm{XC}}. \tag{3}$$

Its trace with $\boldsymbol{\rho}$ yields the spin torque vector while concurrently offering a microscopic picture (Haney et al. 2007) for the origin of torque – misalignment of the nonequilibrium spin density of current carrying quasiparticles with respect to the spins of electrons comprising the magnetic condensate responsible for nonzero $\mathbf{B}_{\mathrm{XC}}$. This causes local torque on individual atoms, which is summed by performing trace in Eq. (2) to find the net effect on the total magnetization $\mathbf{M}_{\mathrm{free}}$ of the free FM layer. Examples of how to evaluate such trace, while using $\mathbf{O} \mapsto \mathbf{T}$ in Eq. (2) in different matrix representations, are given as Eqs. (26) and (27) in Sect. 3.

In equilibrium, $\boldsymbol{\rho}_{\mathrm{eq}}$ is fixed by the Boltzmann-Gibbs prescription, such as $\boldsymbol{\rho}_{\mathrm{eq}} = \sum_n f(E)|\Psi_n\rangle\langle\Psi_n|$ in grand canonical ensemble describing electrons with the Fermi distribution function $f(E)$ due to contact with a macroscopic reservoir at chemical potential μ and temperature T, where E_n and $|\Psi_n\rangle$ are eigenenergies and eigenstates of the Hamiltonian, respectively. Out of equilibrium, the construction of $\boldsymbol{\rho}_{\mathrm{neq}}$ is complicated by the variety of possible driving fields and open nature of a driven quantum system. For example, the Kubo linear-response theory has been used to obtain $\boldsymbol{\rho}_{\mathrm{neq}}$ for small applied electric field in infinite-FM-on-infinite-HM bilayer geometry (Freimuth et al. 2014; Mahfouzi and Kioussis (2018)). However, for arbitrary junction geometry and magnitude of the applied bias voltage V_b or injected pulse current, the most advantageous is to employ the NEGF formalism (Stefanucci and van Leeuwen 2013). This requires to evaluate its two fundamental objects – the retarded GF, $G_{\mathbf{nn}'}^{\sigma\sigma'}(t, t') = -i\Theta(t - t')\langle\{\hat{c}_{\mathbf{n}\sigma}(t), \hat{c}_{\mathbf{n}'\sigma'}^{\dagger}(t')\}\rangle$, and the lesser GF, $G_{\mathbf{nn}'}^{<,\sigma\sigma'}(t, t') = i\langle\hat{c}_{\mathbf{n}'\sigma'}^{\dagger}(t')\hat{c}_{\mathbf{n}\sigma}(t)\rangle$ – describing the density of available quantum states and how electrons occupy those states, respectively. The operator $\hat{c}_{\mathbf{n}\sigma}^{\dagger}$ ($\hat{c}_{\mathbf{n}\sigma}$) creates (annihilates) electron with spin σ at site $\mathbf{n}$ (another index would be required to label more than one orbital present at the site), and $\langle\ldots\rangle$ denotes the nonequilibrium statistical average (Stefanucci and van Leeuwen 2013).

In time-dependent situations, the nonequilibrium density matrix is given by (Petrović et al. 2018; Stefanucci and van Leeuwen 2013)

$$\boldsymbol{\rho}_{\text{neq}}(t) = \mathbf{G}^{<}(t, t)/i. \tag{4}$$

In stationary problems $\mathbf{G}$ and $\mathbf{G}^{<}$ depend only on the time difference $t - t'$ and can, therefore, be Fourier transformed to depend on energy E instead of $t - t'$. The retarded GF in stationary situations is then given by

$$\mathbf{G}(E) = [E - \mathbf{H} - \boldsymbol{\Sigma}_L(E, V_b) - \boldsymbol{\Sigma}_R(E, V_b)]^{-1}, \tag{5}$$

assuming representation in the basis of orthogonal orbitals. In the case of nonorthogonal basis set $|\phi_n\rangle$, one should make a replacement $E \mapsto E\mathbf{D}$ where $\mathbf{D}$ is the overlap matrix composed of elements $\langle \phi_n | \phi_m \rangle$. The self-energies (Velev and Butler 2004; Rungger and Sanvito 2008) $\boldsymbol{\Sigma}_{L,R}(E, V_b)$ describe the semi-infinite leads which guarantee continuous energy spectrum of devices in Fig. 2 required to reach the steady-state transport regime. The leads terminate at infinity into the left (L) and right (R) macroscopic reservoirs with different electrochemical potentials, $\mu_L - \mu_R = eV_L - eV_R = eV_b$. The usual assumption about the leads is that the applied bias voltage V_b induces a rigid shift in their electronic structure(Brandbyge et al. 2002), so that $\boldsymbol{\Sigma}_{L,R}(E, V_b) = \boldsymbol{\Sigma}_{L,R}(E - eV_{L,R})$.

In equilibrium or near equilibrium (i.e., in the linear-response transport regime at small $eV_b \ll E_F$), one needs $\mathbf{G}_0(E)$ obtained from Eq. (5) by setting $V_L = V_R = 0$. The spectral functions shown in Fig. 3a–d can be computed at an arbitrary plane at position x within the junction in Fig. 2a using $\mathbf{G}_0(E)$

$$A(E; \mathbf{k}_{\parallel}, x) = -\frac{1}{\pi} \text{Im} \left[G_0(E; \mathbf{k}_{\parallel}; x, x) \right], \tag{6}$$

where the diagonal matrix elements $G_0(E; \mathbf{k}_{\parallel}; x, x)$ are obtained by transforming the retarded GF from a local orbital to a real-space representation. The spin textures in Fig. 3e–h within the constant energy contours are computed from the spin-resolved spectral function. The equilibrium density matrix can also be expressed in terms of $\mathbf{G}_0(E)$

$$\boldsymbol{\rho}_{\text{eq}} = -\frac{1}{\pi} \int_{-\infty}^{+\infty} dE \, \text{Im} \, \mathbf{G}_0(E) f(E), \tag{7}$$

where $\text{Im} \, \mathbf{O} = (\mathbf{O} - \mathbf{O}^{\dagger})/2i$.

The nonequilibrium density matrix is determined by the lesser GF

$$\boldsymbol{\rho}_{\text{neq}} = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} dE \, \mathbf{G}^{<}(E). \tag{8}$$

In general, if a quantity has nonzero expectation values in equilibrium, that one must be subtracted from the final result since it is *unobservable* in transport experiments. This is exemplified by spin current density in time-reversal invariant systems (Nikolić et al. 2006); spin density, diamagnetic circulating currents, and circulating heat currents in the presence of external magnetic field or spontaneous magnetization breaking time-reversal invariance; and FL component of STT (Theodonis et al. 2006). Thus, the current-driven part of the nonequilibrium density matrix is defined as

$$\boldsymbol{\rho}_{\mathrm{CD}} = \boldsymbol{\rho}_{\mathrm{neq}} - \boldsymbol{\rho}_{\mathrm{eq}}. \tag{9}$$

Although the NEGF formalism can include many-body interactions, such as electron-magnon scattering (Mahfouzi and Nikolić 2014) that can affect STT (Zholud et al. 2017; Levy and Fert 2006; Manchon et al. 2010) and SOT (Yasuda et al. 2017; Okuma and Nomura 2017), here the focus is on the usually considered and conceptually simpler elastic transport regime where the lesser GF of a two-terminal junction

$$\mathbf{G}^{<}(E) = \mathbf{G}(E)\left[if_L(E)\boldsymbol{\Gamma}_L(E) + if_R(E)\boldsymbol{\Gamma}_R(E)\right]\mathbf{G}^{\dagger}(E), \tag{10}$$

is expressed solely in terms of the retarded GF, the level broadening matrices $\boldsymbol{\Gamma}_{L,R}(E) = i[\boldsymbol{\Sigma}_{L,R}(E) - \boldsymbol{\Sigma}_{L,R}^{\dagger}(E)]$ determining the escape rates of electrons into the semi-infinite leads and shifted Fermi functions $f_{L,R}(E) = f(E - eV_{L,R})$.

For purely computational purposes, the integration in Eq. (8) is typically separated [non-uniquely (Xie et al. 2016)] into the apparent "equilibrium" and current-driven "nonequilibrium" terms (Brandbyge et al. 2002; Sanvito 2011)

$$
\begin{aligned}
\boldsymbol{\rho}_{\mathrm{neq}} = &-\frac{1}{\pi} \int\limits_{-\infty}^{+\infty} dE \operatorname{Im} \mathbf{G}(E) f(E - eV_R) \\
&+ \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} dE \, \mathbf{G}(E) \cdot \boldsymbol{\Gamma}_L(E - eV_L) \cdot \mathbf{G}^{\dagger}(E) \left[f_L(E) - f_R(E)\right].
\end{aligned} \tag{11}
$$

The first "equilibrium" term contains integrand which is analytic in the upper complex plane and can be computed via contour integration (Brandbyge et al. 2002; Areshkin and Nikolić 2010; Ozaki 2007; Karrasch et al. 2010), while the integrand in the second "current-driven" term is a nonanalytic function in the entire complex energy plane so that its integration has to be performed directly along the real axis (Sanvito 2011) between the limits set by the window of nonzero values of $f_L(E) - f_R(E)$. Although the second term in Eq. (11) contains information about the bias voltage [through the difference $f_L(E) - f_R(E)$] and about the lead assumed to be injecting electrons into the device (through $\boldsymbol{\Gamma}_L$), it cannot (Xie et al. 2016; Mahfouzi and Nikolić 2013) be used as the proper $\boldsymbol{\rho}_{\mathrm{CD}}$ defined in Eq. (9). This is due

to the fact that the second term in Eq. (9), expressed in terms of the retarded GF via Eq. (7), does not cancel the gauge-noninvariant first term in Eq. (11) which depends explicitly [through $f(E-eV_R)$] on the arbitrarily chosen reference potential V_R and implicitly on the voltages applied to both reservoirs [through $\mathbf{G}(E)$]. Nevertheless, the second term in Eq. (11), written in the linear-response and zero-temperature limit,

$$\boldsymbol{\rho}_{\mathrm{CD}} \stackrel{?}{=} \frac{eV_b}{2\pi} \mathbf{G}_0(E_F) \cdot \boldsymbol{\Gamma}_L(E_F) \cdot \mathbf{G}_0^{\dagger}(E_F), \qquad (12)$$

has often been used in STT literature (Haney et al. 2007; Heiliger et al. 2008) as the putative but *improper* (due to being gauge-noninvariant, which is marked by "?" on the top of the equality sign) expression for $\boldsymbol{\rho}_{\mathrm{CD}}$. Its usage leads to ambiguous (i.e., dependent on arbitrarily chosen V_R) nonequilibrium expectation values.

The *proper* gauge-invariant expression was derived in Mahfouzi and Nikolić (2013)

$$\boldsymbol{\rho}_{\mathrm{CD}} = \frac{eV_b}{2\pi} \mathbf{G}_0(E_F) \cdot \boldsymbol{\Gamma}_L(E_F) \cdot \mathbf{G}_0^{\dagger}(E_F) - \frac{eV_R}{\pi} \mathrm{Im}\, \mathbf{G}_0(E_F)$$

$$- \frac{1}{\pi} \int\limits_{-\infty}^{+\infty} dE\, \mathrm{Im} \left[\mathbf{G}_0 \left(eU_{\mathbf{n}} - eV_L \frac{\partial \boldsymbol{\Sigma}_L}{\partial E} - eV_R \frac{\partial \boldsymbol{\Sigma}_R}{\partial E} \right) \mathbf{G}_0 \right] f(E), \qquad (13)$$

which is given here at zero temperature so that it can be contrasted with Eq. (12). The second and third term in Eq. (13), whose purpose is to subtract any nonzero expectation value that exists in thermodynamic equilibrium, make it quite different from Eq. (12) while requiring to include also electrostatic potential profile $U_{\mathbf{n}}$ across the active region of the device interpolating between V_L and V_R. For example, the second term in Eq. (13) traced with an operator gives equilibrium expectation value governed by the states at E_F which must be removed. The third term in Eq. (13) ensures the gauge invariance of the nonequilibrium expectation values while making the whole expression non-Fermi-surface property. The third term also renders the usage of Eq. (13) computationally demanding due to the requirement to perform integration from the bottom of the band up to E_F together with sampling of $\mathbf{k}_{\parallel}$ points for the junctions in Fig. 2.

Figure 4 shows an example of a left-right *asymmetric* MTJ, FM/I/FM′, whose FM and FM′ layers are assumed to be made of the same material but have different thicknesses. This setup allows us to demonstrate how application of improper $\boldsymbol{\rho}_{\mathrm{CD}}$ in Eq. (12) yields linear-response $\mathbf{T}_{\mathrm{FL}} \propto V_b$ in Fig. 4b that is *incorrectly* an order of magnitude smaller than the correct result in Fig. 4a. This is due to the fact that $\mathbf{T}_{\mathrm{FL}}$ in MTJs possess both the nonequilibrium CD contribution due to spin reorientation at interfaces, where net spin created at one interface is reflected at the second interface where it briefly precesses in the exchange field of the free FM layer, and equilibrium contribution due to interlayer exchange coupling (Theodonis et al. 2006; Yang et al. 2010). The ambiguity in Fig. 4 arises when this equilibrium contribution
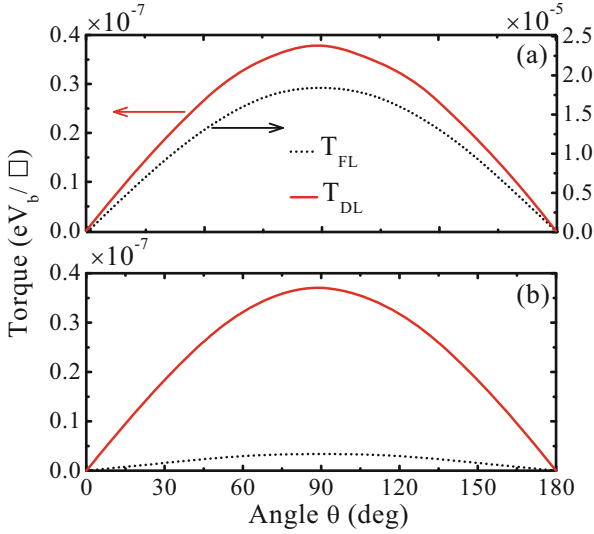
Fig. 4 The angular dependence of the damping-like (i.e., parallel), $|\mathbf{T}_{DL}|$, and field-like (i.e., perpendicular), $|\mathbf{T}_{FL}|$, components of the STT vector in Fig. 1c in an asymmetric (Oh et al. 2009) FM/I/FM′ MTJ computed in the linear-response regime at zero temperature using (**a**) the proper gauge-invariant expression Eq. (13) for ρ_{CD} and (**b**) the improper gauge-noninvariant expression Eq. (12) for ρ_{CD}. The FM/I/FM′ trilayer in Fig. 1a is modeled by TBH (Theodonis et al. 2006) defined on an infinite cubic lattice with a single orbital per site and lattice spacing a. Its insulating barrier has thickness 5 MLs with on-site potential $\varepsilon_{\mathbf{n}} = 6.0$ eV; the left FM layer is semi-infinite, and the right FM′ layer is 20 MLs thick. Both FM layers have the same exchange field $J = 0.5$ eV. The symbol $\square = a^2$ denotes unit interfacial area. (Adapted from Mahfouzi and Nikolić 2013)

is improperly subtracted, so that current-driven $\mathbf{T}_{FL}$ in Fig. 4b is contaminated by a portion of equilibrium contribution added to it when using improper ρ_{CD} in Eq. (12). On the other hand, since $\mathbf{T}_{DL}$ has a zero expectation value in equilibrium, both the proper and improper expressions for ρ_{CD} give the same result in Fig. 4.

Note that in the left-right *symmetric* MTJs, $\mathbf{T}_{FL} \propto V_b$ vanishes. Since this is a rather general result which holds for both MTJs and spin valves in the linear-response regime (Theodonis et al. 2006; Xiao et al. 2008; Heiliger and Stiles 2008) and it has been confirmed in numerous experiments (Wang et al. 2011; Oh et al. 2009), one can use it as a validation test of the computational scheme. For example, the usage of improper ρ_{CD} in Eq. (12), or the proper one in Eq. (13) but with possible software bug, would give nonzero $\mathbf{T}_{FL} \neq 0$ in symmetric junctions at small applied V_b which contradicts experiments (Wang et al. 2011; Oh et al. 2009). In the particular case of symmetric junction, one can actually employ a simpler expression (Mahfouzi and Nikolić 2013; Stamenova et al. 2017) than Eq. (13) which guarantees $\mathbf{T}_{FL} \equiv 0$

$$\rho_{CD} = \frac{eV_b}{4\pi}\mathbf{G}_0(E_F) \cdot [\mathbf{\Gamma}_L(E_F) - \mathbf{\Gamma}_R(E_F)] \cdot \mathbf{G}_0^\dagger(E_F). \tag{14}$$

This expressions is obtained by assuming (Mahfouzi and Nikolić 2013) the particular gauge $V_L = -V_b/2 = -V_R$. Such special gauges and the corresponding Fermi surface expressions for $\boldsymbol{\rho}_{\mathrm{CD}} \propto V_b$ in the linear-response regime do exist also for asymmetric junctions, but one does not know them in advance except for the special case of symmetric junctions (Mahfouzi and Nikolić 2013).

In the calculations in Fig. 4, $(T^x_{\mathrm{CD}}, T^y_{\mathrm{CD}}, T^z_{\mathrm{CD}}) = \mathrm{Tr}\,[\boldsymbol{\rho}_{\mathrm{CD}}\mathbf{T}]$ was first computed using the torque operator $\mathbf{T}$ akin to Eq. (3) but determined by the TBH of the free FM layer. These three numbers are then used to obtain FL (or perpendicular) torque component, $T_{\mathrm{FL}} = T^y_{\mathrm{CD}}$ along the direction $\mathbf{M}_{\mathrm{free}} \times \mathbf{M}_{\mathrm{fixed}}$, and DL (or parallel) torque component, $T_{\mathrm{DL}} = \sqrt{(T^x_{\mathrm{CD}})^2 + (T^z_{\mathrm{CD}})^2}$ in the direction $\mathbf{M}_{\mathrm{free}} \times (\mathbf{M}_{\mathrm{free}} \times \mathbf{M}_{\mathrm{fixed}})$. In MTJs angular dependence of STT components stems only from the cross product, so that $\propto \sin\theta$ dependence (Theodonis et al. 2006; Xiao et al. 2008) for both FL and DL components is obtained in Figs. 4 and 5.
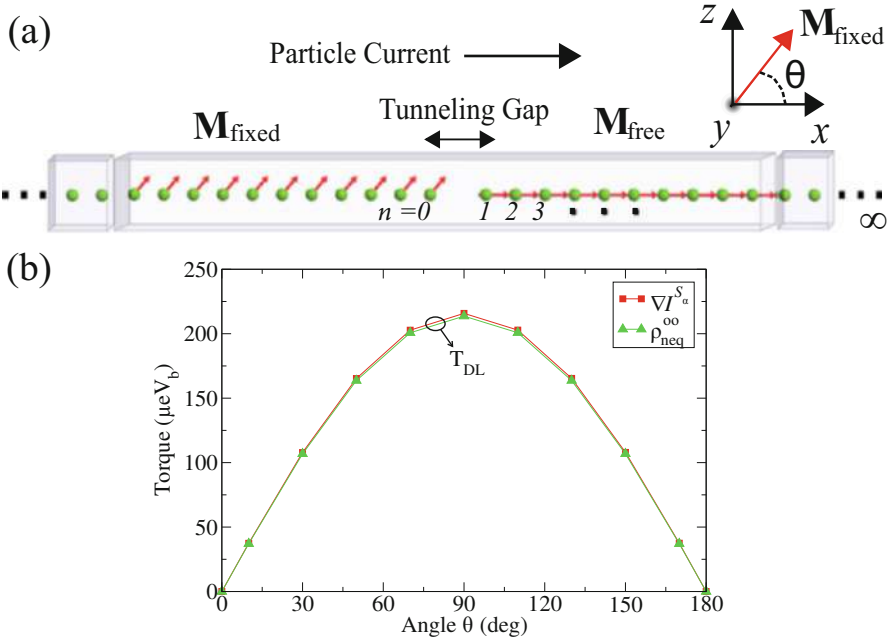
Fig. 5 (**a**) Schematic view of a 1D toy model of MTJ consisting of the left and the right semi-infinite chains of carbon atoms separated by a vacuum gap. (**b**) Comparison of DL component of STT in this MTJ computed via the spin current divergence algorithm (Theodonis et al. 2006; Wang et al. 2008) in Eq. (21) and by using decomposition of the nonequilibrium density matrix into $\boldsymbol{\rho}^{\mu\nu}_{\mathrm{neq}}$ contributions in Eqs. (15), (16), (18), (19), and (20). The DL component of STT, computed in both algorithms in the linear-response regime, acts on the right carbon chain whose magnetic moments comprise the free magnetization $\mathbf{M}_{\mathrm{free}}$ rotated by an angle θ with respect to the fixed magnetization $\mathbf{M}_{\mathrm{fixed}}$ of the left carbon chain. Since MTJ is left-right symmetric, the FL component of the STT vector is zero in the linear-response regime (Wang et al. 2011; Theodonis et al. 2006; Xiao et al. 2008; Oh et al. 2009; Heiliger and Stiles 2008)

In the case of SOT, $\mathbf{T}_{\mathrm{DL}} \propto \mathbf{M}_{\mathrm{free}} \times \mathbf{f}$ and $\mathbf{T}_{\mathrm{DL}} \propto \mathbf{M}_{\mathrm{free}} \times (\mathbf{M}_{\mathrm{free}} \times \mathbf{f})$, where the direction specified by the unit vector $\mathbf{f}$ is determined dynamically once the current flows in the presence of SOC. Therefore, $\mathbf{f}$ is not known in advance (aside from simplistic models like the Rashba ferromagnetic one where $\mathbf{f}$ is along the y-axis for charge current flowing along the x-axis, as illustrated in Fig. 9). Thus, it would be advantageous to decompose $\boldsymbol{\rho}_{\mathrm{CD}}$ into contributions whose trace with the torque operator in Eq. (3) directly yields $\mathbf{T}_{\mathrm{DL}}$ and $\mathbf{T}_{\mathrm{FL}}$. Such decomposition was achieved in Mahfouzi et al. (2016), using adiabatic expansion of Eq. (4) in the powers of $d\mathbf{M}_{\mathrm{free}}/dt$ and symmetry arguments, where $\boldsymbol{\rho}_{\mathrm{neq}} = \boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{oo}} + \boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{oe}} + \boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{eo}} + \boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{ee}}$ is the sum of the following terms

$$\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{oo}} = \frac{1}{8\pi} \int\limits_{-\infty}^{+\infty} dE \, [f_L(E) - f_R(E)] \left(\mathbf{G}\boldsymbol{\Gamma}_L\mathbf{G}^\dagger - \mathbf{G}^\dagger\boldsymbol{\Gamma}_L\mathbf{G} \right.$$

$$\left. - \mathbf{G}\boldsymbol{\Gamma}_R\mathbf{G}^\dagger + \mathbf{G}^\dagger\boldsymbol{\Gamma}_R\mathbf{G} \right), \tag{15}$$

$$\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{oe}} = \frac{1}{8\pi} \int\limits_{-\infty}^{+\infty} dE \, [f_L(E) - f_R(E)] \left(\mathbf{G}\boldsymbol{\Gamma}_L\mathbf{G}^\dagger + \mathbf{G}^\dagger\boldsymbol{\Gamma}_L\mathbf{G} - \mathbf{G}\boldsymbol{\Gamma}_R\mathbf{G}^\dagger - \mathbf{G}^\dagger\boldsymbol{\Gamma}_R\mathbf{G} \right),$$

$$\tag{16}$$

$$\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{eo}} = \frac{1}{8\pi} \int\limits_{-\infty}^{+\infty} dE \, [f_L(E) + f_R(E)] \left(\mathbf{G}\boldsymbol{\Gamma}_L\mathbf{G}^\dagger - \mathbf{G}^\dagger\boldsymbol{\Gamma}_L\mathbf{G} \right. \tag{17}$$

$$\left. + \mathbf{G}\boldsymbol{\Gamma}_R\mathbf{G}^\dagger - \mathbf{G}^\dagger\boldsymbol{\Gamma}_R\mathbf{G} \right) \equiv 0, \tag{18}$$

$$\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{ee}} = \frac{1}{8\pi} \int\limits_{-\infty}^{+\infty} dE \, [f_L(E) + f_R(E)] \left(\mathbf{G}\boldsymbol{\Gamma}_L\mathbf{G}^\dagger + \mathbf{G}^\dagger\boldsymbol{\Gamma}_L\mathbf{G} + \mathbf{G}\boldsymbol{\Gamma}_R\mathbf{G}^\dagger + \mathbf{G}^\dagger\boldsymbol{\Gamma}_R\mathbf{G} \right).$$

$$\tag{19}$$

The four terms are labeled by being odd (o) or even (e) under inverting bias polarity (first index) or time (second index) (Mahfouzi et al. 2016). The terms $\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{oo}}$ and $\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{oe}}$ depend on $f_L(E) - f_R(E)$ and, therefore, are nonzero only in nonequilibrium generated by the bias voltage $V_b \neq 0$ which drives the steady-state current. Using an identity from the NEGF formalism (Stefanucci and van Leeuwen 2013), $\mathbf{G}(\boldsymbol{\Gamma}_L + \boldsymbol{\Gamma}_R)\mathbf{G}^\dagger = i(\mathbf{G} - \mathbf{G}^\dagger)$, reveals that $\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{eo}} \equiv 0$ and

$$\boldsymbol{\rho}_{\mathrm{neq}}^{\mathrm{ee}} = -\frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} dE \, [f_L(E) + f_R(E)] \mathrm{Im}\, \mathbf{G}. \tag{20}$$

Thus, $\rho_{\text{neq}}^{\text{ee}}$ term is nonzero even in equilibrium where it becomes identical to the equilibrium density matrix in Eq. (7), $V_b = 0 \Rightarrow \rho_{\text{neq}}^{\text{ee}} \equiv \rho_{\text{eq}}$. Since $\rho_{\text{neq}}^{\text{oo}}$ is odd under time reversal, its trace with the torque operator in Eq. (3) yields DL component of STT (which depends on three magnetization vectors and it is, therefore, also odd) and FL component of SOT (which depends on one magnetization vector and it is, therefore, also odd). Similarly trace of $\rho_{\text{neq}}^{\text{oe}}$ with the torque operator in Eq. (3) yields FL component of STT and DL component of SOT (Mahfouzi et al. 2016).

In the linear-response regime, pertinent to calculations of STT in spin valves and SOT in FM/spin-orbit-coupled-material bilayers, $f_L(E) - f_R(E) \rightarrow (-\partial f/\partial E) e V_b$. This confines integration in $\rho_{\text{neq}}^{\text{oo}}$ and $\rho_{\text{neq}}^{\text{oe}}$ expressions to a shell of few $k_B T$ around the Fermi energy, or at zero temperature these are just matrix products evaluated at the Fermi energy, akin to Eqs. (12), (13), and (14). Nevertheless, to compute $\rho_{\text{CD}} = \rho_{\text{neq}}^{\text{oo}} + \rho_{\text{neq}}^{\text{oe}} + \rho_{\text{neq}}^{\text{eo}} + \rho_{\text{neq}}^{\text{ee}} - \rho_{\text{eq}}$, one still needs to perform the integration over the Fermi sea in order to obtain $\rho_{\text{neq}}^{\text{ee}} - \rho_{\text{eq}}$, akin to Eq. (13), which can be equivalently computed as $[\rho_{\text{neq}}^{\text{ee}}(V_b) - \rho_{\text{neq}}^{\text{ee}}(-V_b)]/2$ using some small V_b.

To evade singularities on the real axis caused by the poles of the retarded GF in the matrix integral of the type $\int_{-\infty}^{+\infty} dE\, \mathbf{G} f_p(E)$ appearing in Eqs. (7), (13), and (20), such integration can be performed along the contour in the upper half of the complex plane where the retarded GF is analytic. The widely used contour (Brandbyge et al. 2002) consists of a semicircle, a semi-infinite line segment, and a finite number of poles of the Fermi function $f_p(E)$. This contour should be positioned sufficiently far away from the real axis, so that $\mathbf{G}$ is smooth over both of these two segments, while also requiring to select the minimum energy E_{min} (as the starting point of semicircular path) below the bottom of the band edge which is not known in advance in DFT calculations. That is, in self-consistent calculations, incorrectly selected minimum energy causes the charge to erroneously disappear from the system with convergence trivially achieved but to physically incorrect solution. By choosing different types of contours (Areshkin and Nikolić 2010; Ozaki 2007; Karrasch et al. 2010) [such as the "Ozaki contour" (Ozaki 2007; Karrasch et al. 2010) employed in the calculations in Fig. 9] where residue theorem leads to just a sum over a finite set of complex energies, proper positioning of E_{min} and convergence in the number of Fermi poles, as well as selection of sufficient number of contour points along the semicircle and contour points on the line segment, are completely bypassed.

The algorithm based on Eqs. (15), (16), (18), (19), and (20) is compared to often employed spin current divergence algorithm (Theodonis et al. 2006; Wang et al. 2008; Manchon et al. 2008) using a toy model of 1D MTJ, prior to its application to STT calculations in Sect. 3 and SOT calculations in Sect. 4. The model, illustrated in Fig. 5a where the left and right semi-infinite chains of carbon atoms are separated by a vacuum gap, is described by the collinear DFT Hamiltonian implemented in ATK package (Atomistix Toolkit 2017) using single-zeta polarized (Junquera et al. 2001) orbitals on each atom, Ceperley-Alder (Ceperley and Alder 1980) parametrization

of the local spin density approximation for the XC functional (Ceperley and Alder 1980), and norm-conserving pseudopotentials accounting for electron-core interactions. In the absence of spin-flip processes by impurities and magnons or SOC, the STT vector at site n within the right chain can be computed from the divergence (in discrete form) of spin current (Theodonis et al. 2006), $T_n^\alpha = -\nabla I^{S_\alpha} = I_{n-1,n}^{S_\alpha} - I_{n,n+1}^{S_\alpha}$. Its sum over the whole free FM layer gives the total STT as

$$T_{\text{CD}}^\alpha = \sum_{n=1}^{N_{\text{free}}} (I_{n-1,n}^{S_\alpha} - I_{n,n+1}^{S_\alpha}) = I_{0,1}^{S_\alpha} - I_{N_{\text{free}},N_{\text{free}}+1}^{S_\alpha}. \tag{21}$$

Here $I_{0,1}^{S_\alpha}$ is the local spin current, carrying spins pointing in the direction $\alpha \in \{x, y, z\}$, from the last site inside the barrier (which is the last site of the left carbon chain in Fig. 5a) toward the first site of the free FM layer (which is the first site of the right carbon chain in Fig. 5a). Similarly, $I_{N_{\text{free}},N_{\text{free}}+1}$ is the local spin current from the last site inside the free FM layer and the first site of the right lead. Thus, Eq. (21) expresses STT on the free FM layer composed of N_{free} sites as the difference (Wang et al. 2008) between spin currents entering through its left and exiting through its right interface. In the case of semi-infinite free FM layer, $N_{\text{free}} \rightarrow \infty$ and $I_{N_{\text{free}},N_{\text{free}}+1} \rightarrow 0$. The nonequilibrium local spin current can be computed in different ways (Wang et al. 2008), one of which utilizes NEGF expression for ρ_{CD}

$$I_{n,n+1}^{S_\alpha} = \frac{i}{2} \text{Tr} \left[\sigma_\alpha \left(\mathbf{H}_{n,n+1} \rho_{\text{CD}}^{n+1,n} - \rho_{\text{CD}}^{n,n+1} \mathbf{H}_{n+1,n} \right) \right]. \tag{22}$$

Here $\mathbf{H}_{n,n+1}$ and $\rho_{\text{CD}}^{n,n+1}$ are the submatrices of the Hamiltonian and the current-driven part of the nonequilibrium density matrix, respectively, of the size $2N_{\text{orbital}} \times 2N_{\text{orbital}}$ (2 is for spin and N_{orbital} is for the number of orbitals per each atom) which connect sites n and $n + 1$.

Combining Eqs. (14), (21), and (22) yields T_{CD}^x and T_{CD}^z from which we obtain $T_{\text{DL}} = \sqrt{(T_{\text{CD}}^x)^2 + (T_{\text{CD}}^z)^2}$ in the linear-response regime plotted in Fig. 5b as a function of angle θ between $\mathbf{M}_{\text{free}}$ and $\mathbf{M}_{\text{fixed}}$. Alternatively, evaluating the trace of the product of $\rho_{\text{neq}}^{\text{oo}}$ and the torque operator in Eq. (3) yields a vector with two nonzero components, which turn out to be identical to T_{CD}^x and T_{CD}^z computed from the spin current divergence algorithm, as demonstrated in Fig. 5b. The trace of $\rho_{\text{neq}}^{\text{oe}}$ with the torque operator gives a vector with zero x- and z-components and nonzero y-component which, however, is canceled by adding the trace of $\rho_{\text{neq}}^{\text{ee}} - \rho_{\text{eq}}$ with the torque operator to finally produce zero FL component of the STT vector. This is expected because MTJ in Fig. 5a is left-right symmetric.

We emphasize that the algorithm based on the trace of the torque operator with the current-driven part of the nonequilibrium density matrix ρ_{CD} is a *more general approach* than the spin current divergence algorithm since it is valid even in the

presence of spin-flip processes by impurities and magnons or SOC (Haney et al. 2007). In particular, it can be employed to compute SOT (Freimuth et al. 2014; Mahfouzi and Kioussis 2018) in FM/spin-orbit-coupled-material bilayers where spin torque *cannot* (Haney and Stiles 2010) be expressed any more as in Eq. (21).

3 Example: Spin-Transfer Torque in FM/NM/FM Trilayer Spin-Valves

First-principles quantum transport modeling of STT in spin valves (Haney et al. 2007; Wang et al. 2008) and MTJs (Stamenova et al. 2017; Heiliger and Stiles 2008; Jia et al. 2011; Ellis et al. 2017) is typically conducted using an assumption that greatly simplifies computation – noncollinear spins in such systems are described in a *rigid* approximation where one starts from the collinear DFT Hamiltonian and then rotates magnetic moments of either fixed or free FM layer in the spin space in order to generate the relative angle between $\mathbf{M}_{\text{fixed}}$ and $\mathbf{M}_{\text{free}}$ (as it was also done in the calculations of STT in 1D toy model of MTJ in Fig. 5). On the other hand, obtaining true ground state of such system requires noncollinear XC functionals (Capelle et al. 2001; Eich and Gross 2013; Eich et al. 2013; Bulik et al. 2013) and the corresponding self-consistent XC magnetic field $\mathbf{B}_{\text{xc}}$ introduced in Eq. (1). For a given self-consistently converged ncDFT Hamiltonian represented in the linear combination of atomic orbitals (LCAO) basis, we can extract the matrix representation of $\mathbf{B}_{\text{xc}}^{\alpha}$ in the same basis using

$$\mathbf{B}_{\text{XC}}^{x} = 2 \cdot \text{Re} \, \mathbf{H}^{\uparrow\downarrow}, \tag{23}$$

$$\mathbf{B}_{\text{XC}}^{y} = -2 \cdot \text{Im} \, \mathbf{H}^{\uparrow\downarrow}, \tag{24}$$

$$\mathbf{B}_{\text{XC}}^{z} = \mathbf{H}^{\uparrow\uparrow} - \mathbf{H}^{\downarrow\downarrow}. \tag{25}$$

Since LCAO basis sets $|\phi_n\rangle$ are typically nonorthogonal (as is the case of the basis sets (Junquera et al. 2001; Ozaki 2003; Schlipf and Gygi 2015) implemented in ATK (Atomistix Toolkit 2017) and OpenMX (Openmx) packages employed in the calculations of Figs. 5, 6, 7, and 8), the trace leading to the spin torque vector

$$\mathbf{T}_{\text{CD}} = \text{Tr} \, [\boldsymbol{\rho}_{\text{CD}} \boldsymbol{\sigma} \times \mathbf{D}^{-1} \mathbf{B}_{\text{XC}}], \tag{26}$$

requires to use the identity operator $\mathbf{1} = \sum_{ij} |\phi_i\rangle D_{ij}^{-1} \langle \phi_j|$ which inserts $\mathbf{D}^{-1}$ matrix into Eq. (26) where all matrices inside the trace are representations in the LCAO basis. In the real-space basis spanned by the eigenstates $|\mathbf{r}\rangle$ of the position operator, the same trace in Eq. (26) becomes

$$\mathbf{T}_{\text{CD}} = \int\limits_{\text{free FM}} d^3\mathbf{r} \, \mathbf{m}_{\text{CD}}(\mathbf{r}) \times \mathbf{B}_{\text{XC}}(\mathbf{r}). \tag{27}$$
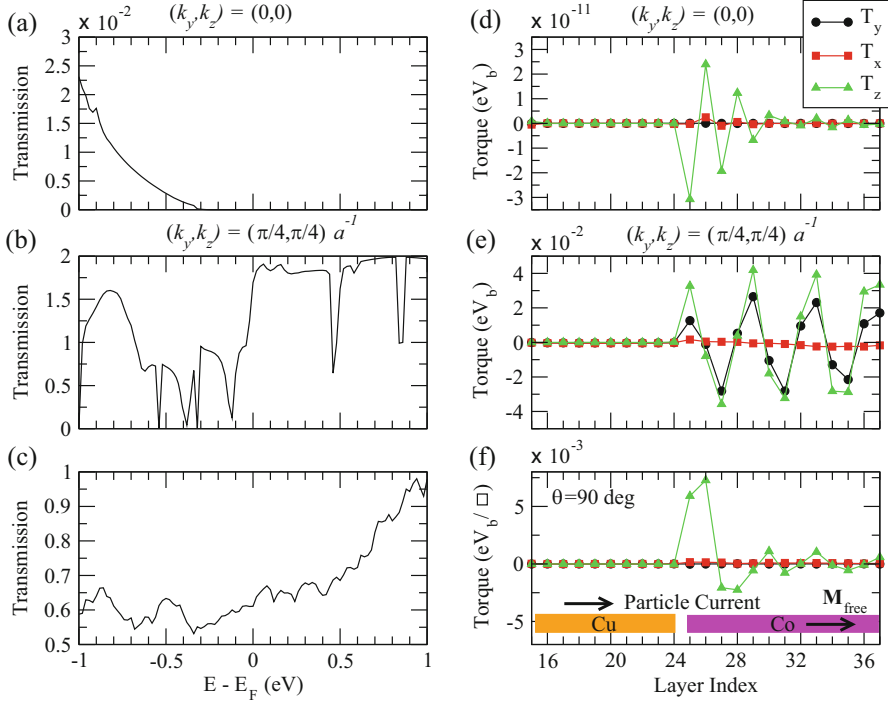
Fig. 6 (**a**), (**b**) Transmission function, defined in Eq. (28), for Co/Cu(9 ML)/Co spin valve illustrated in Fig. 2a at two selected transverse wave vectors $\mathbf{k}_\parallel$, as well as (**c**) summed, as in Eq. (29), over all $\mathbf{k}_\parallel$ vectors within the 2D BZ. (**d**)–(**f**) The corresponding layer-resolved Cartesian components of the STT vector at an angle $\theta = 90°$ between the magnetizations $\mathbf{M}_{\text{fixed}}$ of the left Co layer acting as a polarizer and $\mathbf{M}_{\text{free}}$ of the right Co layer receiving the torque. The mesh of $\mathbf{k}_\parallel$ equally spaced points employed in (**c**) is 51×51, and in (**f**) it is 201×201. The symbol $\square = a^2$ denotes unit interfacial area, where $a = 2.52$ Å is the lattice constant of a common 2D hexagonal unit cell of Co/Cu bilayer (with a lattice mismatch of about 0.65%)

This is a nonequilibrium generalization of the equilibrium torque expression found in ncDFT (Capelle et al. 2001) where $\mathbf{m}_{\text{eq}}(\mathbf{r}) = \langle \mathbf{r} | \boldsymbol{\rho}_{\text{eq}} \boldsymbol{\sigma} | \mathbf{r} \rangle$ in ncDFT is replaced by $\mathbf{m}_{\text{CD}}(\mathbf{r}) = \langle \mathbf{r} | \boldsymbol{\rho}_{\text{CD}} \boldsymbol{\sigma} | \mathbf{r} \rangle$. Note that in thermodynamic equilibrium the integral in Eq. (27) must be zero when integration is performed over all space, which is denoted as "zero-torque theorem" (Capelle et al. 2001), but $\mathbf{m}_{\text{eq}} \times \mathbf{B}_{\text{XC}}(\mathbf{r})$ can be nonzero locally which gives rise to equilibrium torque on the free FM layer that has to be removed by subtracting $\boldsymbol{\rho}_{\text{eq}}$ to obtain $\boldsymbol{\rho}_{\text{CD}}$ in Eq. (9) and plug it into Eq. (27).

We employ ATK package to compute STT in Co/Cu(9 ML)/Co spin valve illustrated in Fig. 2a using ncDFT Hamiltonian combined with Eq. (26). Prior to DFT calculations, the interface builder in the VNL package (Virtual Nanolab 2017) is employed to construct a common unit cell for Co/Cu bilayer. In order to determine the interlayer distance and relaxed atomic coordinates, DFT calculations using VASP (Vienna; Kresse and Hafner 1993; Kresse and Furthmüller
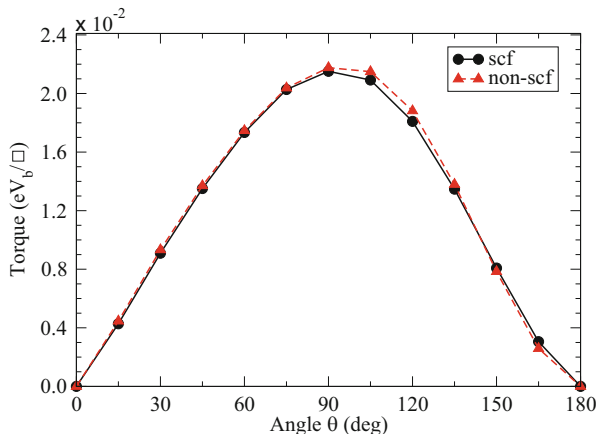
Fig. 7 The DL component of STT vector in symmetric Co/Cu(9 ML)/Co spin valve in Fig. 2a driven by current at small bias voltage V_b as a function of angle θ between the magnetizations $\mathbf{M}_{\text{fixed}}$ of the left Co layer and $\mathbf{M}_{\text{free}}$ of the right Co layer. Its value at $\theta = 90°$ is the sum of layer-resolved STT shown in Fig. 6f. The non-scf curve is computed from Eq. (26) using the rigid approximation (Haney et al. 2007; Wang et al. 2008; Heiliger and Stiles 2008; Jia et al. 2011), where the spin valve is described by collinear DFT Hamiltonian and magnetic moments of the fixed Co layer are rotated by an angle θ with respect to magnetic moments in the free Co layer. Conversely, in the computation of scf curve, the DFT Hamiltonian from the rigid approximation serves as the first iteration leading toward converged self-consistent ncDFT Hamiltonian of the spin valve

1996) with Perdew-Burke-Ernzerhof (PBE) parametrization (Perdew et al. 1996) of the generalized gradient approximation for the XC functional and projected augmented wave (Blöchl 1994; Kresse and Joubert 1999) description of electron-core interactions are performed. The cutoff energy for the plane wave basis set is chosen as 600 eV, while k-points were sampled on a 11×11 surface mesh. In ATK calculations, PBE XC functional, norm-conserving pseudopotentials for describing electron-core interactions and SG15 (medium) LCAO basis set (Schlipf and Gygi 2015) are employed. The energy mesh cutoff for the real-space grid is chosen as 100 Hartree.

The layer-resolved Cartesian components of STT vector within the free Co layer are shown in Fig. 6d–f. The contribution from a propagating state oscillates as a function of position without decaying in Fig. 6e with a spatial period $2\pi/(k_\zeta^\downarrow - k_\eta^\uparrow)$ where ζ (η) denotes different sheets (Wang et al. 2008) of the Fermi surface for minority (majority) spin. This is due to the fact that noncollinear spin in Fig. 2a entering the right Co layer is not an eigenstate of the spin part of the Hamiltonian determined by $\mathbf{M}_{\text{free}}$, and it is, therefore, forced into precession. However, since the shapes of the Fermi surface for majority and minority spin in Co are quite different from each other (Wang et al. 2008), the spatial periods of precession can vary rapidly for different $\mathbf{k}_\parallel$ within the 2D Brillouin zone (BZ). Thus, summation of their contributions leads to cancelation and, therefore, fast decay of STT away
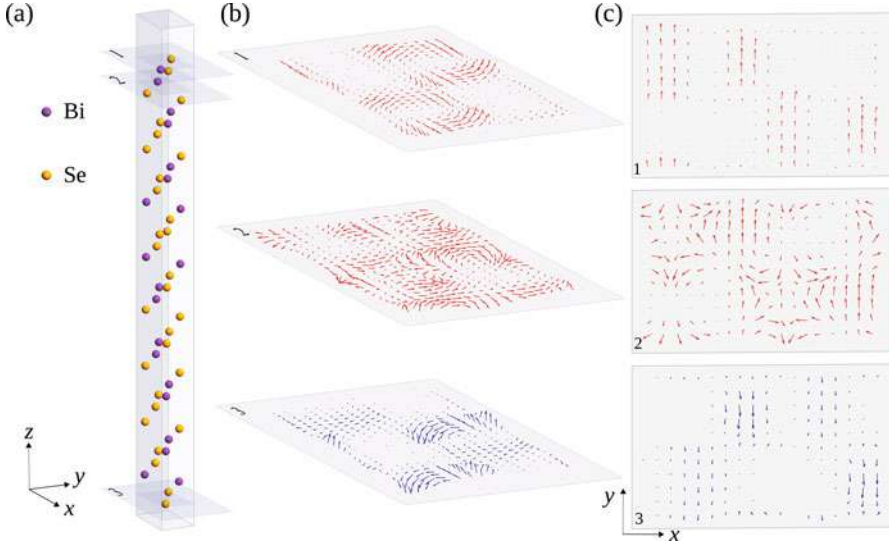
Fig. 8 (**a**) The arrangement of Bi and Se atoms within a supercell of Bi$_2$Se$_3$ thin film (infinite in the x- and the y-directions) consisting of five quintuple layers (one such layer contains three Se layers strongly bonded to two Bi layers in between) and total thickness ≈ 5 nm along the z-axis. (**b**) The vector field of current-driven nonequilibrium spin density $\mathbf{S}_{CD}(\mathbf{r})$ within selected planes shown in (**a**), generated by injection of unpolarized charge current along the x-axis. The planes 1 and 3 correspond to the top and bottom metallic surfaces of Bi$_2$Se$_3$ thin film, while plane 2 resides in the bulk at a distance $d \approx 0.164$ nm away from plane 1. (**c**) The vector fields in (**b**) projected onto each of the selected planes in (**a**). The real-space grid of points in panels (**b**) and (**c**) has spacing $\simeq 0.4$ Å. (Adapted from Chang et al. 2015)

from the interface (Stiles and Zangwill 2002; Wang et al. 2008), as demonstrated by plotting such sum to obtain the total STT per ML of Co in Fig. 6f.

The propagating vs. evanescent states are identified by finite (Fig. 6b) vs. vanishing (Fig. 6a), respectively, $\mathbf{k}_\parallel$-resolved transmission function obtained from the Landauer formula in terms of NEGFs (Stefanucci and van Leeuwen 2013)

$$T(E, \mathbf{k}_\parallel) = \text{Tr}\,[\boldsymbol{\Gamma}_R(E, \mathbf{k}_\parallel)\mathbf{G}_0(E, \mathbf{k}_\parallel)\boldsymbol{\Gamma}_L(E, \mathbf{k}_\parallel)\mathbf{G}_0^\dagger(E, \mathbf{k}_\parallel)], \qquad (28)$$

where the transverse wave vector $\mathbf{k}_\parallel$ is conserved in the absence of disorder. The total transmission function per unit interfacial area is then evaluated using (Ω_{BZ} is the area of sampled 2D BZ)

$$T(E) = \frac{1}{\Omega_{BZ}} \int\limits_{2D\ BZ} T(E, \mathbf{k}_\parallel), \qquad (29)$$

as shown in Fig. 6c. The transmission function in Fig. 6a at $\mathbf{k}_\parallel = (0, 0)$ vanishes at the Fermi energy, signifying evanescent state which *cannot* carry any current across

the junction. Nonetheless, such states can contribute (Ralph and Stiles 2008; Stiles and Zangwill 2002; Wang et al. 2008) to STT vector, as shown in Fig. 6d. Thus, the decay of STT away from Cu/Co interface in Fig. 6f arises both from the cancelation among contributions from propagating states with different $\mathbf{k}_\parallel$ and the decay of contributions from each evanescent state, where the latter are estimated (Wang et al. 2008) to generate $\simeq 10\%$ of the total torque on the ML of free Co layer that is closest to the Cu/Co interface in Fig. 2a.

Since the considered Co/Cu(9 ML)/Co spin valve is left-right symmetric, the FL component of the STT vector is zero. The DL component, as the sum of all layer-resolved torques in Fig. 6f, is plotted as a function of the relative angle θ between $\mathbf{M}_{\text{fixed}}$ and $\mathbf{M}_{\text{free}}$ in Fig. 7. The angular dependence of STT in spin valves does not follow $\propto \sin\theta$ dependence found in the case of MTJs in Figs. 4 and 5.

Although similar analyses have been performed before using collinear DFT Hamiltonian and rigid rotation of magnetic moments in fixed Co layer (Haney et al. 2007; Wang et al. 2008), in Fig. 7 an error made in this approximation was additionally quantified by computing torque using ncDFT Hamiltonian. The rigid approximation is then just the first iteration of the full self-consistent field (scf) calculations leading to the converged ncDFT Hamiltonian. The difference between scf and non-scf calculations in Fig. 7 is rather small due to large number of spacer MLs of Cu, but it could become sizable for small number of spacer MLs enabling coupling between two FM layers.

4 Example: Spin-Orbit Torque in FM/Monolayer-TMD Heterostructures

The calculation of SOT driven by injection of unpolarized charge current into bilayers of the type FM/HM shown in Fig. 2b, FM/monolayer-TMD shown in Fig. 2c, FM/TI, or FM/WSM can be performed using the same NEGF+ncDFT framework combining the torque operator $\mathbf{T}$, $\boldsymbol{\rho}_{\text{CD}}$ expressed in terms of NEGFs and ncDFT Hamiltonian that was delineated in Sect. 2 and applied in Sect. 3 to compute the STT vector in spin valves. Such first-principles quantum transport approach can also easily accommodate possible third capping insulating layer (such as MgO or AlO_x) employed experimentally to increase (Kim et al. 2013) the perpendicular magnetic anisotropy which tilts the magnetization out of the plane of the interface. However, the results of such calculations are not as easy to interpret as in the case of a transparent picture (Stiles and Zangwill 2002; Wang et al. 2008) in Fig. 6d–f explaining how spin angular momentum gets absorbed close to the interface in junctions which exhibit conventional STT. This is due to the fact that several microscopic mechanisms can contribute to SOT, such as the spin Hall effect (SHE) (Vignale 2010; Sinova et al. 2015) within the HM layer (Freimuth et al. 2014; Mahfouzi and Kioussis 2018) with strong bulk SOC and around FM/HM interface (Wang et al. 2016); current-driven nonequilibrium spin density – the so-called Edelstein effect (Edelstein 1990; Aronov and Lyanda-Geller 1989) – due to strong interfacial SOC; spin currents generated in transmission and reflection from

SO-coupled interfaces within 3D transport geometry (Zhang et al. 2015; Kim et al. 2017); and spin-dependent scattering of impurities (Pesin and MacDonald 2012a; Ado et al. 2017) or boundaries (Mahfouzi et al. 2016) in the presence of SOC within FM monolayers. This makes it difficult to understand how to optimize SOT by tailoring materials combination or device geometry to enhance one or more of these mechanisms.

An example of first-principles quantum transport modeling of the Edelstein effect is shown in Fig. 8 for the case of the metallic surface of Bi_2Se_3 as the prototypical 3D TI (Bansil et al. 2016). Such materials possess a usual bandgap in the bulk, akin to conventional topologically trivial insulators, but they also host metallic surfaces whose low-energy quasiparticles behave as massless Dirac fermions. The spins of such fermions are perfectly locked to their momenta by strong SOC, thereby forming spin textures in the reciprocal space (Bansil et al. 2016). In general, when charge current flows through a surface or interface with SOC, the presence of SOC-generated spin texture in the reciprocal space, such as those shown in Figs. 3e, g, and h, will generate nonequilibrium spin density which can be computed using

$$\mathbf{S}_{\text{CD}} = \frac{\hbar}{2}\text{Tr}\,[\boldsymbol{\rho}_{\text{CD}}\boldsymbol{\sigma}]. \tag{30}$$

In the case of simplistic Hamiltonians – such as the Rashba one for 2D electron gas (Winkler 2003), $\hat{H}_{\text{Rashba}} = (\hat{p}_x^2 + \hat{p}_y^2)/2m^* + \frac{\alpha_{\text{SO}}}{\hbar}(\hat{\sigma}_x\hat{p}_y - \hat{\sigma}_y\hat{p}_x)$, or the Dirac one for the metallic surface of 3D TI, $\hat{H}_{\text{Dirac}} = v_F(\hat{\sigma}_x\hat{p}_y - \hat{\sigma}_y\hat{p}_x)$ – the direction and the magnitude of $\mathbf{S}_{\text{CD}}$ are easily determined by back-of-the-envelope calculations (Pesin and MacDonald 2012b). For example, spin texture (i.e., expectation value of the spin operator in the eigenstates of a Hamiltonian) associated with $\hat{H}_{\text{Rashba}}$ consists of spin vectors locked to momentum vector along the two Fermi circles formed in the reciprocal space at the intersection of the Rashba energy-momentum dispersion (Winkler 2003) and the Fermi energy plane. Thus, current flow will disturb balance of momenta to produce $\mathbf{S}_{\text{CD}}$ in the direction transverse to current flow. The same effect is substantially enhanced (Pesin and MacDonald 2012b), by a factor $\hbar v_F/\alpha_{\text{SO}} \gg 1$ where v_F is the Fermi velocity in TI and α_{SO} is the strength of the Rashba SOC, because spin texture associated with $\hat{H}_{\text{Dirac}}$ consists of spin vectors locked to momentum vector along a *single* Fermi circle formed in the reciprocal space at the intersection of the Dirac cone energy-momentum dispersion (Bansil et al. 2016) and the Fermi energy plane. This eliminates the compensating effect of the spins along the second circle in the case of $\hat{H}_{\text{Rashba}}$. Note that nonzero total $\mathbf{S}_{\text{CD}} \propto V_b$ generated by the Edelstein effect is allowed only in nonequilibrium since in equilibrium $\mathbf{S}$ changes sign under time reversal and, therefore, has to vanish (assuming the absence of external magnetic field or magnetization).

In the case of a thin film of Bi_2Se_3 described by ncDFT Hamiltonian, unpolarized charge current injected along the x-axis generates $\mathbf{S}_{\text{CD}} = (0, S_{\text{CD}}^y, S_{\text{CD}}^z)$ on the top surface of the TI, marked as plane 1 in Fig. 8. The in-plane component S_{CD}^y, expected from back-of-the-envelope calculations sketched in the preceding paragraph, is an

order of magnitude larger than the out-of-plane component S_{CD}^z arising due to hexagonal warping (Bansil et al. 2016) of the Dirac cone on the TI surface. The spin texture on the bottom surface of the TI, marked as plane 3 in Fig. 8, has opposite sign to that shown on the top surface because of opposite direction of spins wounding along single Fermi circle on the bottom surface. In addition, a more complicated spin texture in real space (on a grid of points with $\simeq 0.4$ Å spacing), akin to noncollinear intra-atomic magnetism (Nordström and Singh 1996) but driven here by current flow, emerges within $\simeq 2$ nm thick layer below the TI surfaces. This is due to the penetration of evanescent wave functions from the metallic surfaces into the bulk of the TI, as shown in Fig. 8 by plotting $\mathbf{S}_{\mathrm{CD}}(\mathbf{r})$ within plane 2.

The conventional unpolarized charge current injected into the HM layer in Fig. 2b along the x-axis generates transverse spin Hall currents (Vignale 2010; Sinova et al. 2015; Wang et al. 2016) due to strong SOC in such layers. In 3D geometry, spin Hall current along the y-axis carries spins polarized along the z-axis, while the spin Hall current along the z-axis carries spins polarized along the y-axis (Wang et al. 2016). Thus, the effect of the spin Hall current flowing along the z-axis and entering FM layer resembles STT that would be generated by a fictitious polarizing FM layer with fixed magnetization along the y-axis and with charge current injected along the z-axis. While this mechanism is considered to play a major role in the generation of the DL component of SOT (Liu et al. 2012), as apparently confirmed by the Kubo-formula+ncDFT modeling (Freimuth et al. 2014; Mahfouzi and Kioussis 2018), the product of signs of the FL and DL torque components is negative in virtually all experiments (Yoon et al. 2017) (except for specific thicknesses of HM = Ta (Kim et al. 2013) layer). Although HM layer in Fig. 2b certainly generates spin Hall current in its bulk, such spin current can be largely suppressed by the spin memory loss (Dolui and Nikolić 2017; Belashchenko et al. 2016) as electron traverses HM/FM interface with strong SOC. Thus, in contrast to positive sign product in widely accepted picture where SHE is most responsible for the DL component of SOT and Edelstein effect is most responsible for the FL component of SOT, negative sign product indicates that a single dominant mechanism could be responsible for *both* DL and FL torque.

In order to explore how such single mechanism could arise in the absence of spin Hall current, one can calculate torque in a number of specially crafted setups, such as the one chosen in Fig. 9a where a single ultrathin FM layer (consisting of 4 or 10 MLs) is considered with the Rashba SOC present either only in the bottom ML (marked as 0 in Fig. 9a) or in all MLs but with decreasing strength to mimic the spin-orbit proximity effect exemplified in Fig. 3g, h. The setup in Fig. 9a is motivated by SOT experiments (Sklenar et al. 2016; Shao et al. 2016; Guimarães et al. 2018; Lv et al. 2018) on FM/monolayer-TMD heterostructures where SHE is absent due to atomically thin spin-orbit-coupled material. In such bilayers, clean and atomically precise interfaces have been achieved, while back-gate voltage (Lv et al. 2018) has been employed to demonstrate control of the ratio between the FL and DL components of SOT. Note that when bulk TMD or its even-layer thin films are centrosymmetric, its monolayer will be noncentrosymmetric crystal which results in lifting of the spin degeneracy and possibly strong SOC effects (Zhu et al. 2011).
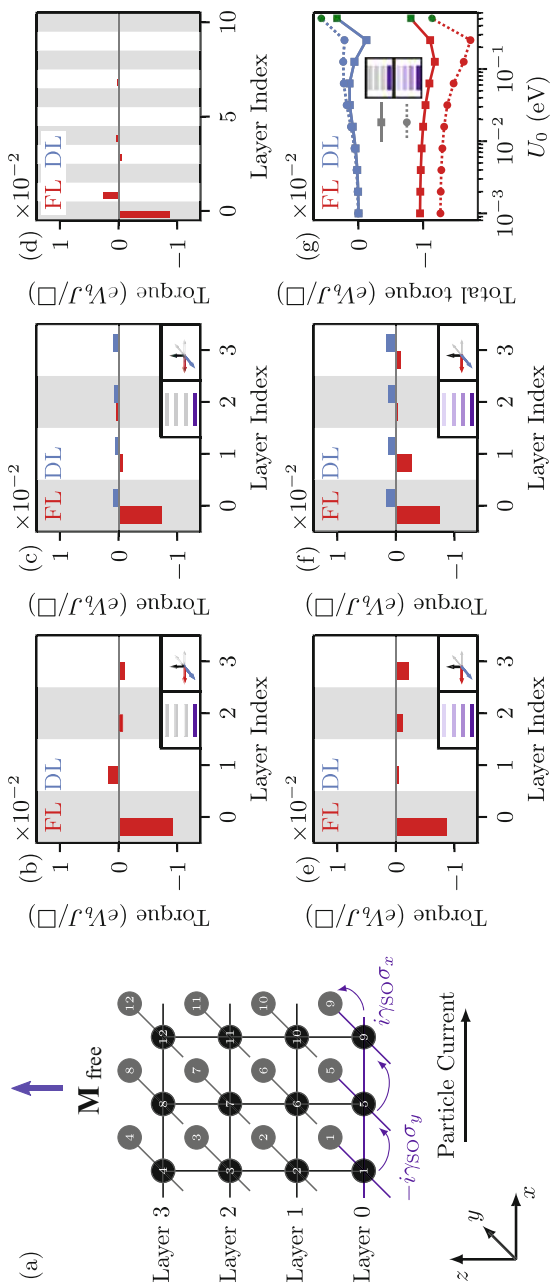
Fig. 9 (**a**) Schematic view of an ultrathin FM layer comprised of few MLs, modeled as infinite square tight-binding lattices (with lattice constant a) in the xy-plane. The bottom layer 0 is assumed to be in direct contact with a spin-orbit-coupled material like TMD. (**b**)–(**g**) The layer-resolved FL and DL components of SOT vector per unit area $\square = a^2$ for the device in (**a**) where FM film is 4 MLs thick in all panels except (**d**) where the thickness is 10 MLs. The bottom layer 0 is assumed to host the Rashba SOC of strength $\gamma_{SO} = 0.2$ eV, induced by the spin-orbit proximity effect from monolayer-TMD (which is assumed to be insulating and, therefore, not explicitly considered), as illustrated in Fig. 3h for Co/monolayer-MoS$_2$. This is twice as strong as the exchange field in the FM layer, $\gamma_{SO}/J = 2$ in Eq. (31). In panels (**e**) and (**f**) the Rashba SOC was also added in layers $l > 0$ with strength 50% smaller than in layer $l − 1$ (i.e., $\gamma_{SO} = 0.2$, 0.1, 0.05, and 0.025 eV in layers 0–3, respectively). Panel (**g**) plots FL and DL components of SOT summed over all layers as a function of the strength of a homogeneous on-site potential $U_\mathbf{n} = U_0$ in Eq. (33) added into the central region of the thickness of 5 sites along the x-axis. The green square and dot in panel (**g**) mark SOT components at $U_0 = 0.5$ eV whose layer-resolved values are shown in panels (**c**) and (**f**), respectively. The temperature is set at $T = 300$ K, Fermi energy is $E_F = 1$ eV, and a grid of $N_{k_y} = 2000$ equally spaced k_y-points was used to sample periodically repeated lattice in the y-direction. The red and blue arrows in the insets of panels (**b**), (**c**), (**e**), and (**f**) denote $\mathbf{M}_{free} \times \mathbf{e}_y$ and $\mathbf{M}_{free} \times (\mathbf{M}_{free} \times \mathbf{e}_y)$ unit vectors along the direction of FL and DL components of SOT, respectively. The quad-graphs in the insets of panels (**b**), (**c**), (**e**), (**f**), and (**g**) signify the presence or absence of the Rashba SOC within layers 0–3 using the shade of the corresponding line

In order to be able to controllably switch SOC on and off in different MLs or to add other types of single-particle interactions, a setup is described in Fig. 9a using TBH defined on an infinite cubic lattice of spacing a with a single orbital per site located at position $\mathbf{n} = (n_x a, n_z a)$

$$\hat{H} = \sum_{\mathbf{n}} \varepsilon_{\mathbf{n}} \hat{c}_{\mathbf{n}}^\dagger \hat{c}_{\mathbf{n}} + \sum_{\langle \mathbf{nn'} \rangle} \hat{c}_{\mathbf{n}}^\dagger \mathbf{t}_{\mathbf{nn'}} \hat{c}_{\mathbf{n'}} - J \sum_{\mathbf{n}} \hat{c}_{\mathbf{n}}^\dagger \mathbf{M}_{\text{free}} \cdot \boldsymbol{\sigma} \hat{c}_{\mathbf{n}}. \tag{31}$$

Here $\hat{c}_{\mathbf{n}}^\dagger = (\hat{c}_{\mathbf{n}\uparrow}^\dagger \ \ \hat{c}_{\mathbf{n}\downarrow}^\dagger)$ is the row vector of operators which create electron at site $\mathbf{n}$ in spin-$\uparrow$ or spin-$\downarrow$ state, and $\hat{c}_{\mathbf{n}} = (\hat{c}_{\mathbf{n}\uparrow} \ \ \hat{c}_{\mathbf{n}\downarrow})$ is the column vector of the corresponding annihilation operators. The spin-dependent nearest-neighbor hoppings in the xz-plane form a 2×2 matrix in the spin space

$$\mathbf{t}_{\mathbf{nn'}} = \begin{cases} -t\mathbf{I}_2, & \text{for } \mathbf{n} = \mathbf{n'} + a\mathbf{e}_z, \\ -t\mathbf{I}_2 - i\gamma_{\text{SO}}\sigma_y, & \text{for } \mathbf{n} = \mathbf{n'} + a\mathbf{e}_x, \end{cases} \tag{32}$$

where $\gamma_{\text{SO}} = \alpha_{\text{SO}}/2a$ measures the strength of the Rashba SOC on the lattice (Nikolić et al. 2006), $\mathbf{I}_2$ is the unit 2×2 matrix, and $\mathbf{e}_\alpha$ are the unit vectors along the axes of the Cartesian coordinate system. The on-site energy

$$\varepsilon_{\mathbf{n}} = \left[U_{\mathbf{n}} - 2t \cos\left(k_y a\right) \right] \mathbf{I}_2 - 2\gamma_{\text{SO}}\sigma_x \sin(k_y a), \tag{33}$$

includes the on-site potential energy $U_{\mathbf{n}}$ (due to impurities, insulating barrier, voltage drop, etc.), as well as kinetic energy effectively generated by the periodic boundary conditions along the y-axis which simulate infinite extension of the FM layer in this direction and require k_y-point sampling in all calculations. The infinite extension along the x-axis is taken into account by splitting the device in Fig. 9a into semi-infinite left lead, central region of arbitrary length along the x-axis, and semi-infinite right lead, all of which are described by the Hamiltonian in Eq. (31) with the same values for $t = 1\,\text{eV}$, $J = 0.1\,\text{eV}$, and the γ_{SO} chosen in all three regions. Thus, γ_{SO} is homogeneous within a given ML, and always present $\gamma_{\text{SO}} = 0.2\,\text{eV}$ on layer 0, but it can take different values in other MLs. The Fermi energy is set at $E_F = 1.0\,\text{eV}$ to take into account possible noncircular Fermi surface (Lee et al. 2015) effects in realistic materials.

The SOT is often studied (Manchon and Zhang 2008; Haney et al. 2013; Lee et al. 2015; Li et al. 2015; Pesin and MacDonald 2012a; Ado et al. 2017; Kalitsov et al. 2017) using the Rashba ferromagnetic model in 2D, which corresponds to just a single layer in Fig. 9a. In that case, only FL torque component is found (Manchon and Zhang 2008; Kalitsov et al. 2017) due to the Edelstein effect and in the absence of spin-dependent disorder (Pesin and MacDonald 2012a; Ado et al. 2017). This is also confirmed in the 3D transport geometry in Fig. 9b, d, e where SOT vector, $\frac{2J}{\hbar}\mathbf{S}_{\text{CD}} \times \mathbf{M}_{\text{free}}$, on layer 0 has only FL component pointing in the $\mathbf{M}_{\text{free}} \times \mathbf{e}_y$ direction, as long as the device is infinite, clean, and homogeneous. In addition, nonzero FL component of SOT in Fig. 9b, d was also found on layers above layer 0

despite the fact that only layer 0 hosts $\gamma_{SO} \neq 0$. This is due to vertical transport along the z-axis in 3D geometry of Fig. 9a, but in the absence of the Rashba SOC on other layers, such effect decays fast as we move toward the top layer, as shown in Fig. 9d for 10 MLs thick FM film. The presence of the Rashba SOC with decreasing γ_{SO} on MLs above layer 0 generates additional nonequilibrium spin density $\mathbf{S}_{CD} \propto \mathbf{e}_y$ on those layers and the corresponding enhancement of the FL component of SOT on those layers in Fig. 9e.

We note that the Kubo-Bastin formula (Bastin et al. 1971) adapted for SOT calculations (Freimuth et al. 2014) predicts actually nonzero DL component of SOT vector for the Rashba ferromagnetic model in 2D due to the change of electronic wave functions induced by an applied electric field termed the "Berry curvature mechanism" (Lee et al. 2015; Kurebayashi et al. 2014; Li et al. 2015). Despite being apparently intrinsic, i.e., insensitive to disorder, this mechanism can be completely canceled in specific models when the vertex corrections are taken into account (Ado et al. 2017). It also gives positive sign product (Lee et al. 2015; Li et al. 2015) of DL and FL components of SOT contrary to majority of experiments where such product is found to be negative (Yoon et al. 2017). It is emphasized here that no electric field can exist in the ballistic transport regime through clean devices analyzed in Fig. 9b, d, e, for which the Kubo-Bastin formula also predicts unphysical divergence (Freimuth et al. 2014; Mahfouzi and Kioussis 2018; Lee et al. 2015; Kurebayashi et al. 2014; Li et al. 2015) of the FL component of SOT. Adding finite voltage drop within the central region, which is actually unjustified in the case of infinite, clean, homogeneous device, results in nonzero DL component of SOT also in the NEGF calculations (Kalitsov et al. 2017). However, the same outcome can be obtained simply by introducing constant potential $U_{\mathbf{n}} = U_0$ on each site of the central region acting as a barrier which reflects incoming electrons, as demonstrated in Fig. 9c, f, g. In the presence of both SOC and such barrier, spin-dependent scattering (Pesin and MacDonald 2012a; Ado et al. 2017) is generated at the lead/central-region boundary which results in nonzero component of $\mathbf{S}_{CD}$ in the direction $\mathbf{M}_{\text{free}} \times \mathbf{e}_y$ and the corresponding DL component of SOT $\propto \mathbf{M}_{\text{free}} \times (\mathbf{M}_{\text{free}} \times \mathbf{e}_y)$ acting on edge magnetic moments (Mahfouzi et al. 2016). This will, therefore, induce inhomogeneous magnetization switching which starts from the edges and propagates into the bulk of FM layer, as observed in experiments (Baumgartner et al. 2017) and micromagnetic simulations (Baumgartner et al. 2017; Mikuszeit et al. 2015).

Interestingly, Fig. 9g also demonstrates that the signs of the DL and FL component are opposite to each other for almost all values of $U_{\mathbf{n}} = U_0$ (except when $U_0 \simeq J$), as observed in the majority of SOT experiments (Yoon et al. 2017). Importantly, the spin-orbit proximity effect within the MLs of FM layer close to FM/spin-orbit-coupled-material interface, as illustrated in Figs. 3g, h and mimicked by introducing the Rashba SOC of decaying strength within all MLs of FM thin film in Fig. 9a, enhances both FL and DL components of SOT. This is demonstrated by comparing solid ($\gamma_{SO} \neq 0$ only on layer 0) and dashed ($\gamma_{SO} \neq 0$ on all layers 0–3) lines in Fig. 9g. This points out at a knob that can be exploited to enhance SOT by searching for optimal combination of materials capable to generate

penetration of SOC over long distances within the FM layer (Marmolejo-Tejada et al. 2017). In fact, in the case of FM/TI and FM/monolayer-TMD heterostructures, proximity SOC coupling within the FM layer is crucial for SOT efficiency (Wang et al. 2017) where it has been considered (Mellnik et al. 2014) that applied current will be shunted through the metallic FM layer and, therefore, not contribute to nonequilibrium spin density generation at the interface where SOC and thereby induced in-plane spin textures are naively assumed to reside.

5   Conclusions

This chapter reviews a unified first-principles quantum transport approach, implemented by combining NEGF formalism with ncDFT calculations, to compute both STT in traditional magnetic multilayers with two FM layers (i.e., the polarizing and analyzing FM layers with fixed and free magnetizations, respectively) and SOT in magnetic bilayers where only one of the layers is ferromagnetic. In the latter case, the role of the fixed magnetization of the polarizing FM layer within spin valves or MTJs is taken over by the current-driven nonequilibrium spin density in the presence of strong SOC introduced by the second layer made of HM, 3D TI, WSM, or monolayer-TMD. This approach resolves recent confusion (Freimuth et al. 2014; Kalitsov et al. 2017) in the literature where apparently only the Kubo formula, operating with expressions that integrate over the Fermi sea in order to capture change of wave functions due to the applied electric field and the corresponding interband electronic transitions (Freimuth et al. 2014; Lee et al. 2015; Li et al. 2015), can properly obtain the DL component of SOT. In addition, although the Kubo formula approach can also be integrated with first-principles calculations (Freimuth et al. 2014; Mahfouzi and Kioussis 2018), it can only be applied to a single device geometry (where infinite FM layer covers infinite spin-orbit-coupled-material layer while current flows parallel to their interface) and in the linear-response transport regime. In contrast, NEGF+ncDFT approach reviewed in this chapter can handle arbitrary device geometry, such as spin valves and MTJs exhibiting STT or bilayers of the type FM/spin-orbit-coupled-material which are made inhomogeneous by attachment to NM leads, at vanishing or finite applied bias voltage. In contrast to often employed 2D transport geometry (Manchon and Zhang 2008; Haney et al. 2013; Lee et al. 2015; Li et al. 2015; Pesin and MacDonald 2012a; Ado et al. 2017; Kalitsov et al. 2017; Ndiaye et al. 2017) for SOT theoretical analyses, it is emphasized here the importance of 3D transport geometry (Kim et al. 2017; Ghosh and Manchon 2018) to capture both the effects at the FM/spin-orbit-coupled-material interface and those further into the bulk of the FM layer. Finally, ultrathin FM layers employed in SOT experiments can hybridize strongly with the adjacent spin-orbit-coupled material to acquire its SOC and the corresponding spin textures on all of the FM monolayers. Such "hybridized ferromagnetic metals" can have electronic and spin structure (Fig. 3) which is quite different from an isolated FM layer, thereby requiring usage of both 3D geometry and first-principles Hamiltonians [of either tight-binding (Freimuth et al. 2014; Mahfouzi and Kioussis

2018) or pseudopotential-LCAO-ncDFT (Theurich and Hill 2001) type] to predict the strength of SOT in realistic systems and optimal materials combinations for device applications of the SOT phenomenon.

References

Ado IA, Tretiakov OA, Titov M (2017) Microscopic theory of spin-orbit torques in two dimensions. Phys Rev B 95:094401

Areshkin DA, Nikolić BK (2010) Electron density and transport in top-gated graphene nanoribbon devices: first-principles Green function algorithms for systems containing a large number of atoms. Phys Rev B 81:155450

Aronov AG, Lyanda-Geller YB (1989) Nuclear electric resonance and orientation of carrier spins by an electric field. JETP Lett 50:431

Atomistix Toolkit (ATK) 2017.2. http://www.quantumwise.com

Bahramy M, King PDC, de la Torre A, Chang J, Shi M, Patthey L, Balakrishnan G, Hofmann P, Arita R, Nagaosa N, Baumberger F (2012) Emergent quantum confinement at topological insulator surfaces. Nat Commun 3:1159

Bansil A, Lin H, Das T (2016) *Colloquium*: topological band theory. Rev Mod Phys 88:021004

Bastin A, Lewiner C, Betbeder-Matibet O, Nozières P (1971) Quantum oscillations of the Hall effect of a fermion gas with random impurity scattering. J Phys Chem Solids 32:1811

Baumgartner M, Garello K, Mendil J, Avci CO, Grimaldi E, Murer C, Feng J, Gabureac M, Stamm C, Acremann Y, Finizio S, Wintz S, Raabe J, Gambardella P (2017) Spatially and time-resolved magnetization dynamics driven by spin-orbit torques. Nat Nanotech 12:980

Belashchenko KD, Kovalev AA, van Schilfgaarde M (2016) Theory of spin loss at metallic interfaces. Phys Rev Lett 117:207204

Berger L (1996) Emission of spin waves by a magnetic multilayer traversed by a current. Phys Rev B 54:9353

Berkov DV, Miltat J (2008) Spin-torque driven magnetization dynamics: micromagnetic modeling. J Magn Magn Mater 320:1238

Blöchl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953

Borders WA, Akima H, Fukami S, Moriya S, Kurihara S, Horio Y, Sato S, Ohno H (2017) Analogue spin-orbit torque device for artificial-neural-network-based associative memory operation. Appl Phys Expr 10:013007

Brandbyge M, Mozos JL, Ordejón P, Taylor J, Stokbro K (2002) Density-functional method for nonequilibrium electron transport. Phys Rev B 65:165401

Brataas A, Bauer GEW, Kelly PJ (2006) Non-collinear magnetoelectronics. Phys Rep 427:157

Bulik IW, Scalmani G, Frisch MJ, Scuseria GE (2013) Oncollinear density functional theory having proper invariance and local torque properties. Phys Rev B 87:035117

Capelle K, Vignale G, Györffy BL (2001) Spin currents and spin dynamics in time-dependent density-functional theory. Phys Rev Lett 87:206403

Carva K, Turek I (2009) Landauer theory of ballistic torkances in noncollinear spin-valves. Phys Rev B 80:104432

Ceperley DM, Alder BJ (1980) Ground state of the electron gas by a stochastic method. Phys Rev Lett 45:566

Chang P-H, Markussen T, Smidstrup S, Stokbro K, Nikolić BK (2015) Nonequilibrium spin texture within a thin layer below the surface of current-carrying topological insulator Bi_2Se_3: a first-principles quantum transport study. Phys Rev B 92:201406(R)

Chantis AN, Belashchenko KD, Tsymbal EY, van Schilfgaarde M (2007) Tunneling anisotropic magnetoresistance driven by resonant surface states: first-principles calculations on an Fe(001) surface. Phys Rev Lett 98:046601

Christen T, Büttiker M (1996) Gauge-invariant nonlinear electric transport in mesoscopic conductors. Europhys Lett 35:523

Dolui K, Nikolić BK (2017) Spin-memory loss due to spin-orbit coupling at ferromagnet/heavy-metal interfaces: ab initio spin-density matrix approach. Phys Rev B 96:220403(R)

Edelstein VM (1990) Spin polarization of conduction electrons induced by electric current in two-dimensional asymmetric electron systems. Solid State Commun 73:233

Eich FG, Gross EKU (2013) Transverse spin-gradient functional for noncollinear spin-density-functional theory. Phys Rev Lett 111:156401

Eich FG, Pittalis S, Vignale G (2013) Transverse and longitudinal gradients of the spin magnetization in spin-density-functional theory. Phys Rev B 88:245102

Ellis MOA, Stamenova M, Sanvito S (2017) Multiscale modeling of current-induced switching in magnetic tunnel junctions using ab initio spin-transfer torques. Phys Rev B 96:224410

Evans RFL, Fan WJ, Chureemart P, Ostler TA, Ellis MOA, Chantrell RW (2014) Atomistic spin model simulations of magnetic nanomaterials. J Phys Condens Matter 26:103202

Fan Y, Upadhyaya P, Kou X, Lang M, Takei S, Wang Z, Tang J, He L, Chang L-T, Montazeri M, Jiang GY, Nie T, Schwartz RN, Tserkovnyak Y, Wang KL (2014) Magnetization switching through giant spin-orbit torque in a magnetically doped topological insulator heterostructure. Nat Mater 13:699

Freimuth F, Blügel S, Mokrousov Y (2014) Spin-orbit torques in Co/Pt(111) and Mn/W(001) magnetic bilayers from first principles. Phys Rev B 90:174423

Garello K, Miron IM, Avci CO, Freimuth F, Mokrousov Y, Blügel S, Auffret S, Boulle O, Gaudin G, Gambardella P (2013) Symmetry and magnitude of spin-orbit torques in ferromagnetic heterostructures. Nat Nanotech 8:587

Ghosh S, Manchon A (2018) Spin-orbit torque in a three-dimensional topological insulator-ferromagnet heterostructure: crossover between bulk and surface transport. Phys Rev B 97:134402

Guimarães MHD, Stiehl GM, MacNeill D, Reynolds ND, Ralph DC (2018) Spin-orbit torques in $NbSe_2$/Permalloy bilayers. Nano Lett 18:1311

Han J, Richardella A, Siddiqui SA, Finley J, Samarth N, Liu L (2017) Room-temperature spin-orbit torque switching induced by a topological insulator. Phys Rev Lett 119:077702

Haney PM, Stiles MD (2010) Current-induced torques in the presence of spin-orbit coupling. Phys Rev Lett 105:126602

Haney PM, Waldron D, Duine RA, Núñez AS, Guo H, MacDonald AH (2007) Current-induced order parameter dynamics: microscopic theory applied to Co/Cu/Co spin-valves. Phys Rev B 76:024404

Haney PM, Lee H-W, Lee K-J, Manchon A, Stiles MD (2013) Current induced torques and interfacial spin-orbit coupling: semiclassical modeling. Phys Rev B 87:174411

Heiliger C, Stiles MD (2008) Ab Initio studies of the spin-transfer torque in magnetic tunnel junctions. Phys Rev Lett 100:186805

Heiliger C, Czerner M, Yavorsky BY, Mertig I, Stiles MD (2008) Implementation of a nonequilibrium Green's function method to calculate spin-transfer torque. J Appl Phys 103:07A709

Hernández AR, Lewenkopf CH (2013) Nonlinear electronic transport in nanoscopic devices: nonequilibrium Green's functions versus scattering approach. Eur Phys J B 86:131

Jia X, Xia K, Ke Y, Guo H (2011) Nonlinear bias dependence of spin-transfer torque from atomic first principles. Phys Rev B 84:014401

Johansson A, Henk J, Mertig I (2018) Edelstein effect in Weyl semimetals. Phys Rev B 97:085417

Junquera J, Paz O, Sánchez-Portal D, Artacho E (2001) Numerical atomic orbitals for linear-scaling calculations. Phys Rev B 64:235111

Kalitsov A, Nikolaev SA, Velev J, Chshiev M, Mryasov O (2017) Intrinsic spin-orbit torque in a single-domain nanomagnet. Phys Rev B 96:214430

Karrasch C, Meden V, Schönhammer K (2010) Finite-temperature linear conductance from the Matsubara Green's function without analytic continuation to the real axis. Phys Rev B 82:125114

Katine JA, Albert FJ, Buhrman RA, Myers EB, Ralph DC (2000) Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars. Phys Rev Lett 84:3149

Kent AD, Worledge DC (2015) A new spin on magnetic memories. Nat Nanotech 10:187

Kim J, Sinha J, Hayashi M, Yamanouchi M, Fukami S, Suzuki T, Mitani S, Ohno H (2013) Layer thickness dependence of the current-induced effective field vector in Ta|CoFeB|MgO. Nat Mater 12:240

Kim K-W, Lee K-J, Sinova J, Lee H-W, Stiles MD (2017) Spin-orbit torques from interfacial spin-orbit coupling for various interfaces. Phys Rev B 96:104438

Kresse G, Furthmüller J (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15

Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. Phys Rev B 47:558

Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. Phys Rev B 59:1758

Kubota H, Fukushima A, Yakushiji K, Nagahama T, Yuasa S, Ando K, Maehara H, Nagamine Y, Tsunekawa K, Djayaprawira DD, Watanabe N, Suzuki Y (2008) Quantitative measurement of voltage dependence of spin-transfer torque in MgO-based magnetic tunnel junctions. Nat Phys 4:37

Kurebayashi H, Fang JSD, Irvine AC, Skinner TD, Wunderlich J, Novák V, Campion RP, Gallagher BL, Vehstedt EK, Zârbo LP, Výborný K, Ferguson AJ, Jungwirth T (2014) An antidamping spin-orbit torque originating from the Berry curvature. Nat Nanotech 9:211

Lee K-S, Go D, Manchon A, Haney PM, Stiles MD, Lee H-W, Lee K-J (2015) Angular dependence of spin-orbit spin-transfer torques. Phys Rev B 91:144401

Levy PM, Fert A (2006) Spin transfer in magnetic tunnel junctions with hot electrons. Phys Rev Lett 97:097205

Li H, Gao H, Zârbo LP, Výborný K, Wang X, Garate I, Doğan F, Čejchan A, Sinova J, Jungwirth T, Manchon A (2015) Intraband and interband spin-orbit torques in noncentrosymmetric ferromagnets. Phys Rev B 91:134402

Liu L, Pai C-F, Li Y, Tseng HW, Ralph DC, Buhrman RA (2012) Spin-torque switching with the giant spin Hall effect of tantalum. Science 336:555

Locatelli N, Cros V, Grollier J (2014) Spin-torque building blocks. Nat Mater 13:11

Lv W, Jia Z, Wang B, Lu Y, Luo X, Zhang B, Zeng Z, Liu Z (2018) Electric-field control of spin-orbit torques in WS_2/permalloy bilayers. ACS Appl Mater Interfaces 10:2843

MacNeill D, Stiehl GM, Guimaraes MHD, Buhrman RA, Park J, Ralph DC (2017a) Control of spin-orbit torques through crystal symmetry in WTe_2/ferromagnet bilayers. Nat Phys 13:300

MacNeill D, Stiehl GM, Guimarães MHD, Reynolds ND, Buhrman RA, Ralph DC (2017b) Thickness dependence of spin-orbit torques generated by WTe_2. Phys Rev B 96:054450

Mahfouzi F, Kioussis N (2017) Current-induced damping of nanosized quantum moments in the presence of spin-orbit interaction. Phys Rev B 95:184417

Mahfouzi F, Kioussis N (2018) First-principles study of angular dependence of spin-orbit torque in Pt/Co and Pd/Co bilayers. Phys Rev B 97: 224426

Mahfouzi F, Nikolić BK (2013) How to construct the proper gauge-invariant density matrix in steady-state nonequilibrium: applications to spin-transfer and spin-orbit torques. SPIN 3:1330002

Mahfouzi F, Nikolić BK (2014) Signatures of electron-magnon interaction in charge and spin currents through magnetic tunnel junctions: a nonequilibrium many-body perturbation theory approach. Phys Rev B 90:045115

Mahfouzi F, Nikolić BK, Kioussis N (2016) Antidamping spin-orbit torque driven by spin-flip reflection mechanism on the surface of a topological insulator: a time-dependent nonequilibrium Green function approach. Phys Rev B 93:115419

Mahfouzi F, Kim J, Kioussis N (2017) Intrinsic damping phenomena from quantum to classical magnets: an ab initio study of Gilbert damping in a Pt/Co bilayer. Phys Rev B 96:214421

Manchon A, Zhang S (2008) Theory of nonequilibrium intrinsic spin torque in a single nanomagnet. Phys Rev B 78:212405

Manchon A, Miron IM, Jungwirth T, Sinova J, Zelezný J, Thiaville A, Garello K, Gambardella P (2019) Current-induced spin-orbit torques in ferromagnetic and antiferromagnetic systems. Rev Mod Phys 91:035004

Manchon A, Ryzhanova N, Vedyayev A, Chschiev M, Dieny B (2008) Description of current-driven torques in magnetic tunnel junctions. J Phys Condens Matter 20:145208

Manchon A, Zhang S, Lee K-J (2010) Signatures of asymmetric and inelastic tunneling on the spin torque bias dependence. Phys Rev B 82:174420

Marmolejo-Tejada JM, Chang P-H, Lazić P, Smidstrup S, Stradi D, Stokbro K, Nikolić BK (2017) Proximity band structure and spin textures on both sides of topological-insulator/ferromagnetic-metal interface and their charge transport probes. Nano Lett 17:5626

Marzari N, Mostofi AA, Yates JR, Souza I, Vanderbilt D (2012) Maximally localized Wannier functions: theory and applications. Rev Mod Phys 84:1419

Mellnik AR, Lee JS, Richardella A, Grab JL, Mintun PJ, Fischer MH, Vaezi A, Manchon A, Kim E-A, Samarth N, Ralph DC (2014) Spin-transfer torque generated by a topological insulator. Nature 511:449

Mikuszeit N, Boulle O, Miron IM, Garello K, Gambardella P, Gaudin G, Buda-Prejbeanu LD (2015) Spin-orbit torque driven chiral magnetization reversal in ultrathin nanostructures. Phys Rev B 92:144424

Miron IM, Garello K, Gaudin G, Zermatten P-J, Costache MV, Auffret S, Bandiera S, Rodmacq B, Schuhl A, Gambardella P (2011) Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. Nature 476:189

Myers EB, Ralph DC, Katine JA, Louie RN, Buhrman RA (1999) Current-induced switching of domains in magnetic multilayer devices. Science 285:867

Ndiaye PB, Akosa CA, Fischer MH, Vaezi A, Kim E-A, Manchon A (2017) Dirac spin-orbit torques and charge pumping at the surface of topological insulators. Phys Rev B 96:014408

Nikolić BK, Zârbo LP, Souma S (2006) Imaging mesoscopic spin Hall flow: spatial distribution of local spin currents and spin densities in and out of multiterminal spin-orbit coupled semiconductor nanostructures. Phys Rev B 73:075303

Nordström L, Singh DJ (1996) Noncollinear intra-atomic magnetism. Phys Rev Lett 76:4420

Oh S-C, Manchon S-YPA, Chschiev M, Han J-H, Lee H-W, Lee J-E, Nam K-T, Jo Y, Kong Y-C, Dieny B, Lee K-J (2009) Bias-voltage dependence of perpendicular spin-transfer torque in asymmetric MgO-based magnetic tunnel junctions. Nat Phys 5:898

Okuma N, Nomura K (2017) Microscopic derivation of magnon spin current in a topological insulator/ferromagnet heterostructure. Phys Rev B 95:115403

Openmx 3.8. http://www.openmx-square.org/

Ozaki T (2003) Variationally optimized atomic orbitals for large-scale electronic structures. Phys Rev B 67:155108

Ozaki T (2007) Continued fraction representation of the Fermi-Dirac function for large-scale electronic structure calculations. Phys Rev B 75:035123

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865

Perez N, Martinez E, Torres L, Woo S-H, Emori S, Beach GSD (2014) Chiral magnetization textures stabilized by the Dzyaloshinskii-Moriya interaction during spin-orbit torque switching. Appl Phys Lett 104:213909

Pesin DA, MacDonald AH (2012a) Quantum kinetic theory of current-induced torques in Rashba ferromagnets. Phys Rev B 86:014416

Pesin D, MacDonald AH (2012b) Spintronics and pseudospintronics in graphene and topological insulators. Nat Mater 11:409

Petrović M, Popescu BS, Plecháč P, Nikolić BK (2018) Spin and charge pumping by steady or pulse current driven magnetic domain wall: a self-consistent multiscale time-dependent-quantum/time-dependent-classical approach. Phys Rev Applied 10:054038 https://journals.aps.org/prapplied/abstract/10.1103/PhysRevApplied.10.054038

Ralph D, Stiles M (2008) Spin transfer torques. J Magn Magn Mater 320:1190

Rungger I, Sanvito S (2008) Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition. Phys Rev B 78:035407

Sankey JC, Cui Y-T, Sun JZ, Slonczewski JC, Buhrman RA, Ralph DC (2008) Measurement of the spin-transfer-torque vector in magnetic tunnel junctions. Nat Phys 4:67

Sanvito S (2011) Electron transport theory for large systems. In: Bichoutskaia E (ed) Computational nanoscience. RSC Publishing, Cambridge

Schlipf M, Gygi F (2015) Optimization algorithm for the generation of ONCV pseudopotentials. Comput Phys Commun 196:36

Shao Q, Yu G, Lan Y-W, Shi Y, Li M-Y, Zheng C, Zhu X, Li L-J, Khalili Amiri P, Wang KL (2016) Strong Rashba-Edelstein effect-induced spin-orbit torques in monolayer transition metal dichalcogenide/ferromagnet bilayers. Nano Lett 16:7514

Shelley M, Poilvert N, Mostofi AA, Marzari N (2011) Automated quantum conductance calculations using maximally-localised Wannier functions. Comput Phys Commun 182:2174

Sinova J, Valenzuela SO, Wunderlich J, Back CH, Jungwirth T (2015) Spin Hall effects. Rev Mod Phys 87:1260

Sklenar J, Zhang W, Jungfleisch MB, Jiang W, Saglam H, Pearson JE, Ketterson JB, Hoffmann A (2016) Perspective: Interface generation of spin-orbit torques. J Appl Phys 120:180901

Slonczewski JC (1996) Current-driven excitation of magnetic multilayers. J Magn Magn Mater 159:L1

Soumyanarayanan A, Reyren N, Fert A, Panagopoulos C (2016) Emergent phenomena induced by spin-orbit coupling at surfaces and interfaces. Nature 539:509

Stamenova M, Mohebbi R, Seyed-Yazdi J, Rungger I, Sanvito S (2017) First-principles spin-transfer torque in CuMnAs|GaP|CuMnAs junctions. Phys Rev B 95:060403

Stefanucci G, van Leeuwen R (2013) Nonequilibrium many-body theory of quantum systems: a modern introduction. Cambridge University Press, Cambridge

Stiles MD, Zangwill A (2002) Anatomy of spin-transfer torque. Phys Rev B 66:014407

Theodonis I, Kioussis N, Kalitsov A, Chshiev M, Butler WH (2006) Anomalous bias dependence of spin torque in magnetic tunnel junctions. Phys Rev Lett 97:237205

Theurich G, Hill NA (2001) Self-consistent treatment of spin-orbit coupling in solids using relativistic fully separable *ab initio* pseudopotentials. Phys Rev B 64:073106

Thygesen K, Jacobsen K (2005) Molecular transport calculations with Wannier functions. Chem Phys 319:111

Timopheev AA, Sousa R, Chshiev M, Buda-Prejbeanu LD, Dieny B (2015) Respective influence of in-plane and out-of-plane spin-transfer torques in magnetization switching of perpendicular magnetic tunnel junctions. Phys Rev B 92:104430

Tsoi M, Jansen AGM, Bass J, Chiang W-C, Seck M, Tsoi V, Wyder P (1998) Excitation of a magnetic multilayer by an electric current. Phys Rev Lett 80:4281

Velev J, Butler W (2004) On the equivalence of different techniques for evaluating the green function for a semi-infinite system using a localized basis. J Phys Condens Matter 16:R637

Vienna Ab initio Simulation Package (VASP) 5.4. http://www.vasp.at/

Vignale G (2010) Ten years of spin Hall effect. J Supercond Nov Magn 23:3

Virtual Nanolab (VNL) 2017.2. http://www.quantumwise.com

Wang C, Cui Y-T, Katine JA, Buhrman RA, Ralph DC (2011) Time-resolved measurement of spin-transfer-driven ferromagnetic resonance and spin torque in magnetic tunnel junctions. Nat Phys 7:496

Wang L, Wesselink RJH, Liu Y, Yuan Z, Xia K, Kelly PJ (2016) Giant room temperature interface spin Hall and inverse spin Hall effects. Phys Rev Lett 116:196602

Wang S, Xu Y, Xia K (2008) First-principles study of spin-transfer torques in layered systems with noncollinear magnetization. Phys Rev B 77:184430

Wang Y, Zhu D, Wu Y, Yang Y, Yu J, Ramaswamy R, Mishra R, Shi S, Elyasi M, Teo K-L, Wu Y, Yang H (2017) Room temperature magnetization switching in topological insulator-ferromagnet heterostructures by spin-orbit torques. Nat Commun 8:1364

Winkler R (2003) Spin-orbit coupling effects in two-dimensional electron and hole systems. Springer, Berlin

Xiao J, Bauer GEW, Brataas A (2008) Spin-transfer torque in magnetic tunnel junctions: scattering theory. Phys Rev B 77:224419

Xie Y, Rungger I, Munira K, Stamenova M, Sanvito S, Ghosh AW (2016) Spin transfer torque: a multiscale picture. In: Atulasimha J, Bandyopadhyay S (eds) Nanomagnetic and spintronic devices for energy-efficient memory and computing. Wiley, Hoboken

Yang HX, Chshiev M, Kalitsov A, Schuhl A, Butler WH (2010) Effect of structural relaxation and oxidation conditions on interlayer exchange coupling in Fe/MgO/Fe tunnel junctions. Appl Phys Lett 96:262509

Yasuda K, Tsukazaki A, Yoshimi R, Kondou K, Takahashi KS, Otani Y, Kawasaki M, Tokura Y (2017) Current-nonlinear Hall effect and spin-orbit torque magnetization switching in a magnetic topological insulator. Phys Rev Lett 119:137204

Yoon J, Lee S-W, Kwon JH, Lee JM, Son J, Qiu X, Lee K-J, Yang H (2017) Anomalous spin-orbit torque switching due to field-like torque-assisted domain wall reflection. Sci Adv 3:e1603099

Zhang SS-L, Vignale G, Zhang S (2015) Anisotropic magnetoresistance driven by surface spin-orbit scattering. Phys Rev B 92:024412

Zholud A, Freeman R, Cao R, Srivastava A, Urazhdin S (2017) Spin transfer due to quantum magnetization fluctuations. Phys Rev Lett 119:257201

Zhu ZY, Cheng YC, Schwingenschlögl U (2011) Giant spin-orbit-induced spin splitting in two-dimensional transition-metal dichalcogenide semiconductors. Phys Rev B 84:153402

Part V
Low-Dimensional Materials at the Nanoscale

Low-dimensional materials at the nanoscale: Overview

24

James R. Chelikowsky

Contents

Abstract

Chapters in this section center on applications of computational methods to current and emergent materials, especially physically confined systems at the nanoscale.

1   Emergence

In this section, computational applications to "current and emergent" materials are presented.

Of course, current materials are ones of contemporary interest such as nanostructures, topological materials, and restricted dimensional materials such as graphene. Active research project focuses on such materials for device applications and fundamental understanding.

The term "emergent materials" is a more problematic one. The concept of an unanticipated property emerging as the number of atoms increase is not new. A key

J. R. Chelikowsky (✉)
Center for Computational Materials, Oden Institute for Computational Engineering and
Sciences, Departments of Physics and Chemical Engineering, The University of Texas at Austin,
Austin, TX, USA
e-mail: jrc@utexas.edu

reference in the evolution of this concept is Anderson's highly cited paper on *More Is Different* (Anderson 1972), published almost a half century ago. In his seminal paper, Anderson noted: "The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe."

Consider a simple example. Atoms carry charge, but not dipoles. The idea that an isolated atom in free space might carry a dipole would seem to violate the isotropy of free space. However, when atoms are brought together to form solids, it is possible for a resulting crystal to be a "ferroelectric," i.e., the crystal possesses spontaneous electric polarization or a permanent dipole moment. The manifestation of ferroelectricity is claimed to be an emergent property as one would not have anticipated such a property by a study of the constituent atoms. The property only occurs when large number of atoms interacts. Hence, the title of Anderson's article reflects this emergence as "more is different."

Some have disparaged the use "emergence" in physics as commonly used by claiming it is ill defined, vague, and utilized solely for the purpose of obtaining research funding (Kivelson and Kivelson 2016). One aspect of such criticism is the anthropic nature of the definition, specifically, defining emergence as involving an "unanticipated" property. An unanticipated property is clearly a subjective judgment. One might argue that a particular "emergent property" should have anticipated and therefore not "emergent." Most scientists would argue that definitions of properties should not be dependent on a subjective judgment.

In this case, one might argue that a property deemed unanticipated by one scientist might not be so characterized by another scientist. This is not an uncommon situation. As the author recalls, the Nobel Laureate Linus Pauling once said in response to a question: "It's not a difficult problem if you know what you are doing." This, it could be argued, is like saying a difficult problem is no longer difficult once you know the solution. As such one would not characterize a problem as difficult in a scientific sense.

This said, scientists plainly use anthropic avenues of thought as they are human, and such loosely defined characterizations are common and widely accepted. For example, scientists often characterize materials by words such as "complex" or "advanced," and these terms are equally problematic in terms of an objective definition as is the term: emergent. Today's advanced material may not be so a decade from now, and terming a material as complex is clearly a subjective characterization. The argument about terming a property as emergent to obtain funding may centered more on the judgment of the grant officer than on science per se. Whatever the arguments about the meaning of an "emergent" property, a material possessing such a property is likely to be of more general interest than one not.

With such issues in mind, a recently proposed definition of an "emergent" property has been proposed that attempts to more objectively quantify the term. An emergent property should be defined as "a qualitative property that can only occur in the limit where the number of microscopic constituents tends to infinity" (Kivelson and Kivelson 2016). This definition is more rigorous as it removes the issue of whether the property is "unanticipated." However, this definition does not capture

the imagination often associated with emergent behavior as implied by colloquial use and Anderson's "more is different" perspective.

2   Size, Symmetry, and Restricted Dimensionality in Materials

Emergent material properties are often associated with issues of size, symmetry, and restricted dimensionality. The authors of articles in this section for the most part focus on nanoscale systems which exhibit size-dependent properties, some of them emergent and some not.

Consider the following thought experiment about size properties. Suppose we start with a bulk crystal. We can characterize its properties as either "intensive" or "extensive." Intensive properties are not dependent on its size. Density is such a property. If we considered a $1\,cm^3$ sample or a $10\,cm^3$, sample, the density would be the same. A bandgap of a semiconductor is a similar property as it does not change in such circumstance. Extensive properties include the mass or volume of the system. Now suppose we make the size of the system smaller and smaller. As we do so, we would eventually find some "intensive" properties to be dependent on the size of the system.

The bandgap of an insulating material is an obvious example. As the material approaches the nanoscale, the bandgap changes. When an inherently intensive property becomes dependent on size, one can usually argue that the materials reside in a nanoscale regime. As a crude estimate for that size regime, consider the "size" of an excitation in a semiconductor. Typically the Bohr radius of a hydrogenic model for an electron-hole pair in a semiconductor is around 5–10 nm. For systems smaller than this size, we expect the bulk bandgap to be significantly modified.

Consider the properties of a silicon nanocrystal. Very small nanocrystals of silicon are optically active, i.e., they absorb and emit light in the visible spectrum. By "very small" we mean a few hundred atoms. As the size of such a nanocrystal increases, two interesting phenomena occur. First, the optical gap becomes independent of size and saturates to a given value. Second, the optically active material becomes optically inactive.

In Fig. 1, we illustrate computations for the ionization potential, IP, and electronic affinity, EA, along with the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) levels. The IP and EA energies are defined by

$$I = E(N - 1) - E(N)$$
$$A = E(N) - E(N + 1)$$

(1)

where E is the total electronic energy of the nanocrystal.

The quasiparticle gap is the energy to create noninteracting electron-hole pairs given by $IP - EA$. This is not the gap we expect for the absorption of light as we

Fig. 1 Evolution of the
ionization potential (IP) and
electron affinity (EA) with
quantum dot size. Also shown
are the eigenvalue levels for
the highest occupied
molecular orbital (HOMO)
and the lowest unoccupied
molecular orbital (LUMO).
(Chelikowsky in MTM
Chapter 24)
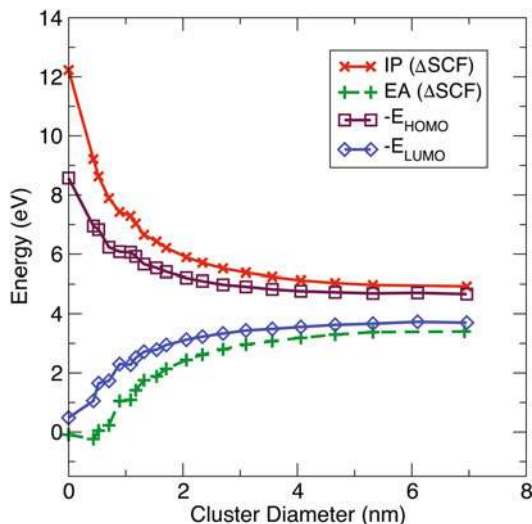
have not included interactions between the electron and hole. Also, we used density
functional theory to compute these energies. They are ground-state energies, but the
functional we used to compute these energies is approximate. The gap decreases as
the size of the nanocrystal increases and eventually approaches the gap given by the
difference in the highest occupied energy level and the lowest unoccupied energy
level.

While the quasiparticle gap is not equal to the actual optical gap, the trends are the
same. The evolution of the gap from a nanocrystal to the bulk is not what we would
call emergent with our working definition. The trend is not qualitatively different
and it is not unanticipated. As the size of nanocrystal decreases, the excitation
is physically confined to a smaller domain. We can argue using the uncertainty
principle that the energy of such an excitation should increase. This can be shown
by a simple model based on a finite well (Chan et al. 2008).

Why does an optically active Si nanocrystal become inactive as the size of the
nanocrystal increases? This effect is illustrated in Fig. 2. The oscillator strength for
optical transitions is reduced by more than an order of magnitude as the diameter of
the nanocrystal increases from 6 to 18 Å. In this sense, the optical properties of bulk
silicon are qualitatively different than for nanocrystal silicon.

The evolution of this property might again have been anticipated. In crystalline
silicon, conservation of crystal momentum prevents the absorption of light at the
bandgap threshold. The conservation of crystal momentum is a direct result of
translational symmetry, which does not exist in the limit of a small nanocrystal of
silicon.

Other nanoscale systems often reflect restricted dimensionality anisotropically
such as nanofilms and nanowires where the systems are confined in one or two
dimensions. A limiting case of a nanofilm includes two-dimensional materials

Fig. 2 Oscillator strength of
dipole-allowed transitions
near the absorption edge vs
cluster diameter. (Vasiliev
2002)
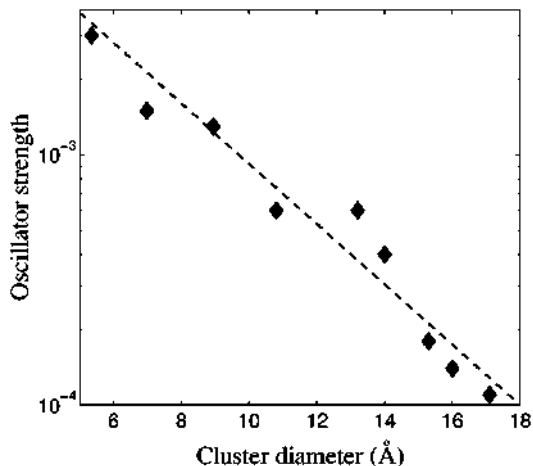
such as graphene or a single layer of MoS_2. By appropriately capitalizing on the
role of quantum confinement in these systems, workers hope to change properties
by changing the physical size of the material and not by changing its chemical
composition.

3   Computational Challenges for Emerging Materials

The study of emergent properties is computationally difficult owing to spanning
length scales and the large number of atoms involved. While the physical size
of a nanocrystal might be small, the number of atoms involved is not. Often
thousands of atoms are involved in nanoscale phenomena. Only within the last
decade have computers and algorithms evolved to the point where it is possible to
consider computations for nanostructures that are comparable in size to those made
experimentally.

A key outstanding issue remains centers on the structural properties of nanostruc-
tures, which are in general not known even for nanocrystals. While it may be safe
to assume that nanocrystal structures are similar to the bulk material, the surface
structure can be problematic. In particular, one of the issues with a confined system
is exactly how to confine it. Often nanostructures are passivated by some agent
to remove chemically active bonds on the surface or the nanocrystals themselves
embedded in inert matrixes.

While one might be able to compute the energy of systems with many atoms,
the near degeneracy of competing states often precludes a definitive prediction of
the ground state. Moreover, the number of degrees of freedoms can be too large
to optimize the nanostructure by energy minimization. Still, we can often learn
about a material by considering energetically accessible structures or an ensemble
of materials that may closely resemble experimental situations. For example, we

might consider a nanostructure created by a bulk fragment. If this system is then passivated by "capping atoms" that eliminate dangling bonds, we can obtain interesting perspectives on the role of quantum confinement.

Computational methods as outlined in the companion set of this Handbook: (*Methods: Theory and Modeling*) address many of these issues in detail.

4   Contributions

This section contains contributions that are centered on limited dimensionality at the nanoscale. The authors use a variety of computational means to examine structural, electronic, optical, and magnetic properties.

A number of the contributions focus on two-dimensional materials such as graphene sheets or nanoribbons:

- ▶ Chapter 26, "Functionalizing Two-Dimensional Materials for Energy Applications" by Bing Huang and Su-Huai Wei. This contributions is focused on the physical properties of some reprehensive 2D materials. These 2D materials include graphene, transition metal disulfides and diselenides, black phosphorus, and other metastable two-dimensional materials.
- ▶ Chapter 25, "Interaction of Hydrogen with Graphitic Surfaces, Clean and Doped with Metal Clusters" by Julio A. Alonso and María J. López. These authors review the work on the interaction of molecular and atomic hydrogen with graphene and graphene nanoribbons and address the effects due to the presence of metal clusters on the surface because of the evidence of their role in enhancing hydrogen storage.
- ▶ Chapter 29, "Interlayer Interactions in Low-Dimensional Layered Hetero-structures: Modeling and Applications" by Oded Hod. This work provides an overview of recent developments for constructing reliable, efficient, and transferable anisotropic interlayer force-fields for homogeneous and heterogeneous low-dimensional hexagonal layered materials. Applications to the study of the structural and tribological properties of layered structures are presented.
- ▶ Chapter 31, "Electronic Structure of Atomically Precise Graphene Nanoribbons" by Leopold Talirz and Carlo A. Pignedoli. This chapter guides the reader through a theoretical toolbox available for modeling the electronic structure of graphene nanoribbons, including a range of approaches from simple empirical models to state-of-the-art *ab initio* methods.

The remaining chapters center on nanostructures confined in two dimensions (nanowires) or in three dimensions (molecules, clusters, and nanocrystals).

- ▶ Chapter 27, "Spins in Semiconductor Nanocrystals" by Gustavo Martini Dalpian. Semiconductor nanocrystals can be used as hosts to trap and manipulate single spins, which can have different properties than their bulk counterparts,

owing both to quantum confinement. This chapter discusses how spins can be generated in nanocrystals by the insertion of impurities, either magnetic or not, or by magnetic polarons present at dangling bonds at their surfaces.

- ▶ Chapter 28, "Excited-State Properties of Thin Silicon Nanowires" by Li Yang. Nanowires are special in terms of anisotropy as physical confinement occurs in two directions resulting in a one-dimensional periodic system. Many-electron interactions can be computed using many-body perturbation theory methods based on the GW approximation and the Bethe-Salpeter equation. This chapter illustrates the results found with these computational tools indicating enhanced electronic interactions owing to restricted dimensionality and depressed electronic screening.

- ▶ Chapter 30, "Emergence of Functionalized Properties in Semiconductor Nanostructures" by James R. Chelikowsky. In this chapter computational methods are discussed that can be used to predict the emergence of nanoscale electronic properties in semiconductors, both intrinsic and extrinsic, and how these properties change as the size of the system decreased.

References

Anderson PW (1972) More is different. Science 177:393

Chan TL, Tiago ML, Kaxiras E, Chelikowsky JR (2008) Size limits on doping phosphorus into silicon nanocrystals. Nano Lett 8:596

Kivelson S, Kivelson SA (2016) Defining emergence in physics. Npj Quantum Mater 1:16024

Vasiliev I, Ogut S, Chelikowsky JR (2002) First-principles density functional calculations for the optical spectra of clusters and nanocrystals]. Phys Rev B 65:115416

Interaction of Hydrogen with Graphitic Surfaces, Clean and Doped with Metal Clusters

25

Julio A. Alonso and María J. López

Contents

Abstract

Hydrogen is viewed as a possible alternative to the fossil fuels in transportation. The technology of fuel-cell engines is fully developed, and the outstanding remaining problem is the storage of hydrogen in the vehicle. Porous materials, in which hydrogen is adsorbed on the pore walls, and in particular nanoporous carbons, have been investigated as potential onboard containers. Furthermore, metallic nanoparticles embedded in porous carbons catalyze the dissociation of hydrogen in the anode of the fuel cells. For these reasons the interaction of hydrogen with the surfaces of carbon materials is a topic of high technological interest. Computational modeling and the density functional formalism (DFT) are helping in the task of discovering the basic mechanisms of the interaction of hydrogen with clean and doped carbon surfaces. Planar and curved graphene provide good models for the walls of porous carbons. We first review work on the interaction

J. A. Alonso (✉) · M. J. López
Departamento de Física Teórica, Atómica y Optica, University of Valladolid, Valladolid, Spain
e-mail: jaalonso@fta.uva.es; maria.lopez@fta.uva.es

of molecular and atomic hydrogen with graphene and graphene nanoribbons, and next we address the effects due to the presence of metal clusters on the surface because of the evidence of their role in enhancing hydrogen storage.

1   Introduction

The interaction of hydrogen with the surface of carbonaceous materials is a topic of high interest in materials science. Intense work has been motivated by the concern that oil reserves in our planet may be exhausted in five decades or a bit more. Then, hydrogen is viewed as a candidate to replace gasoline as a fuel in cars (Ogden 2002). The engine technology is ready: those cars use an electric motor, based on a fuel cell in which hydrogen coming from the storage tank reacts with atmospheric oxygen producing an electric current, and the only emission from the reaction is water. More specifically, molecular hydrogen is dissociated in the anode of the fuel cell, the protons go through a proton exchange membrane and the electrons through an external circuit creating an electric current, and recombination and reaction with oxygen occurs in the cathode. These novel cars are now fabricated by the major car companies.

A crucial ingredient in this process is the storage of the hydrogen needed to feed the cell (Jena 2011). Hydrogen is a gas, and the present challenge is to find efficient ways to store enough hydrogen in a tank to be able to run the car for 500 or 600 Km. The technology currently in use consists in storing hydrogen gas in steel tanks at high pressures of 350–700 bar. However, in the opinion of many engineers, that method is not the ultimate solution. Hydrogen storage in solid materials is viewed as a safer and convenient alternative. Storage of hydrogen in light solid porous materials is a promising method because the storage mechanisms are the physical or chemical adsorption of hydrogen on the pore walls, and those porous materials show large specific surface areas up to several thousand m^2/g. An interesting class of porous materials is the family of nanoporous carbons. That family includes activated carbons synthesized by physical or chemical activation of carbon-organic precursors or coals (Jordá-Beneito et al. 2008), carbide-derived carbons synthesized by reacting metal carbides with halogens (Yushin et al. 2006), graphitic nanofibers, and carbon nanotubes. Experiments have been performed to measure the storage capacities of those materials, with the result that the amount of hydrogen stored at low temperatures (around 100 K) approaches the onboard storage targets of 6% of hydrogen in weight (wt) and 40 g of hydrogen per liter, but the storage capacity is lower at room temperature, far below the storage targets (Ansón et al. 2004; Jordá-Beneito et al. 2007; Lueking et al. 2005; Panella et al. 2005).

Experimental work has shown that the storage capacity of porous carbons can be enhanced by doping these materials with metallic atoms and nanoclusters (Contescu et al. 2009, 2011). Theoretical calculations and computer simulations have been performed to shed light on the mechanisms of interaction between hydrogen and clean or doped porous carbons. In this way, the calculations can help to improve the materials in the quest of the difficult task of fulfilling the storage targets.

2   Theoretical Methods

In most theoretical simulations of the interaction of hydrogen with porous carbons, the inner walls of the pores are modeled as single-layer graphitic-like nanostructures: graphene when simulating pores with planar surfaces, and carbon nanotubes or open fullerenes in the case of curved pore surfaces, and this model is supported by extensive computer simulations of the structure of porous carbons (López et al. 2011; De Tomás et al. 2017). The method generally used to investigate the interaction between hydrogen and the carbon substrates (clean or doped) is the density functional formalism, or DFT (Fiolhais et al. 2003), a method which combines (1) reasonable accuracy in describing the electronic structure and calculating interaction energies, on one hand, and (2) acceptable computation times, which allow simulating systems with more than 100 atoms easily, on the other. A key ingredient in the density functional calculations is the exchange-correlation energy functional and the corresponding exchange-correlation potential. The description of electronic exchange and correlation effects has advanced to a mature state, and a number of practical approximations are available. After the successful use for many years of the local density approximation (Fiolhais et al. 2003), most works today use functionals belonging to the family of generalized gradient approximations, or GGA (Perdew et al. 1996; Becke 2014). Two of the most commonly used exchange-correlation energy functionals, known as PW (Perdew et al. 1992) and PBE (Perdew et al. 1996), have been developed by Perdew and coworkers.

   The interaction between molecular hydrogen and graphitized surfaces is weak, and a part of this interaction comes from the van der Waals forces. Consequently, it is desirable to include dispersion corrections in the DFT calculations, and this field has been reviewed recently (Klimes and Michaelides 2012; Hermann et al. 2017). One family of DFT-based methods uses nonlocal correlation functionals in which the dispersion interactions are obtained directly from the electron density. In a second class of DFT methods, dispersion interactions are included empirically or semiempirically at different levels of complexity. At present, no particular method to include dispersion interactions in DFT can be generally preferred over all the others, and the best option is to restrict attention to the family of materials of interest in each particular case and perform careful tests of the dispersion methods and functionals applied to that family of materials. One example is the recent study of the physisorption of molecular hydrogen on benzene and graphene (Cabria et al. 2017), in which the results of different density functionals have been compared to ab initio quantum chemistry methods.

3   Molecular Hydrogen on Graphene

When molecular hydrogen interacts with a graphitized carbon surface without defects, the molecule binds weakly to the surface by a physisorption mechanism. The measured physisorption energy of H_2 on graphene is 0.048 eV (Mattera

Fig. 1 Relevant configurations for a H_2 molecule on top of a graphene surface. Three sites: A, on a carbon atom; B, on a carbon-carbon bond; H, on the center of a hexagon. For each site, three different orientations of the molecule axis are shown. The first subscript on the site index informs if the molecule axis is parallel of perpendicular to the graphene plane. The second subscript indicates if the axis is parallel or perpendicular to two C–C bonds of the hexagon. (Reproduced from Cabria et al. 2017, with permission of AIP Publishing)
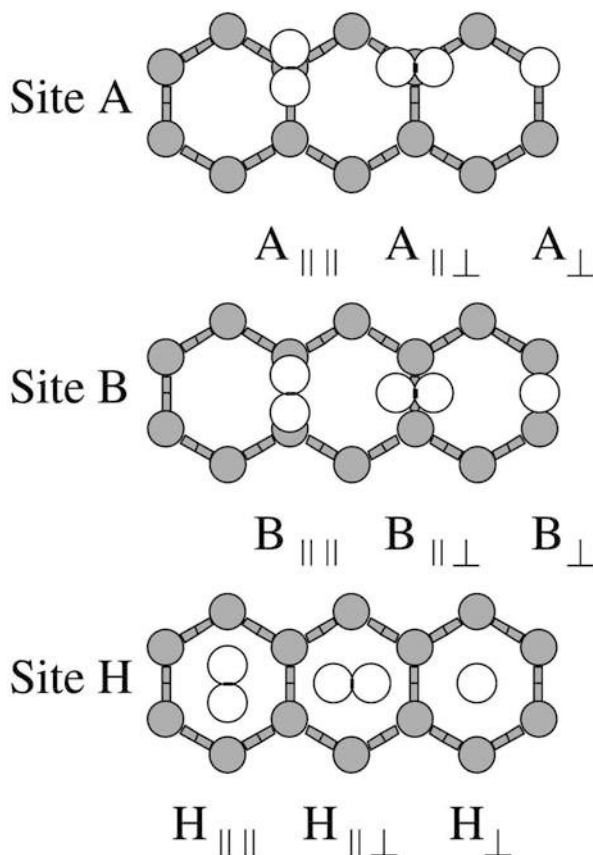
et al. 1980; Costanzo et al. 2012). DFT calculations indicate that the most stable adsorption sites are at the center of the carbon hexagons with the molecular axis parallel to the graphene plane (Cabria et al. 2017). The adsorption energy obtained by the PBE functional corrected by dispersion corrections in a form proposed by Grimme, PBE-D2 (Grimme 2006), is $E_{ads} = 0.051$ eV, in good agreement with experiment, and the equilibrium distance from the molecule to the carbon plane is $d_e = 2.85$ Å. The perpendicular orientation of the molecular axis leads to a similar adsorption energy, and adsorption energies on other sites shown in Fig. 1, bridge positions between two C atoms, or positions on top of C atoms, are 0.01 eV smaller. In summary, adsorption energies depend only weakly on the adsorption site and orientation of the molecular axis. The adsorption energy can be interpreted as a sum of several contributions, all of them small. One contribution is due to the dispersion interactions (interaction between induced fluctuating electrical dipoles). Another contribution arises from the small overlap of the electron densities of molecule and substrate. Even in the absence of any rearrangement of the overlapping densities, electronic exchange and correlation effects produce a small bonding

contribution (Arellano et al. 2000). In addition, there is a rearrangement of the electron density (or electronic relaxation) that, although small, also contributes to binding. Calculations for H_2 adsorbed on the outer wall of carbon nanotubes lead to the same conclusion (Cabria et al. 2006), and an experimental estimation (Brown et al. 2000) gave $E_{ads} = 0.062$ eV.

How much hydrogen can be adsorbed on graphene is determined by the interplay between the H_2-graphene and the H_2–H_2 interaction potentials. The H_2–H_2 potential is repulsive for distances $d < 3.4$ Å between the two molecules and has a weak attractive tail with a shallow minimum of depth 0.003–0.004 eV at $d = 3.5$ Å (Okamoto and Miyamoto 2001). But the distance between the centers of two neighbor hexagons in graphene is only 2.46 Å, and this means that full covering of graphene (or a carbon nanotube) with one H_2 molecule per hexagon is not possible. An estimate of the affordable low temperature coverage can be done by considering a model (Okamoto and Miyamoto 2001) of a two dimensional lattice of physisorbed molecules interacting through an H_2–H_2 potential $V_{HH}(r)$ and writing the energy of N molecules physisorbed on area S as

$$E = N\left(-t + \gamma V_{HH}\left(r_{nn}\right)/2\right). \tag{1}$$

E is given as a sum of two terms. The first one is the total physisorption energy of noninteracting molecules, $-Nt$, written in a way that assumes the same value, $-t$, for the physisorption of each molecule; this is supported by the observation made above that the physisorption energies depend only weakly on the adsorption site. The second term accounts for the intermolecular H_2–H_2 interactions. The H_2–H_2 interactions are restricted to the nearest-neighbor molecules, and this is accounted for by the coordination number γ in the lattice ($\gamma = 6$ for a triangular lattice). The number of physisorbed molecules per area S is 1.154 S/r_{nn}^2 for a triangular lattice. Taking t as a parameter with values between 0.04 and 0.05 eV, and minimizing E with respect to the nearest-neighbor separation r_{nn}, leads to a coverage of 4.5–4.6% (weight of adsorbed hydrogen divided by the total weight of hydrogen and substrate). Since the walls of the pores are essentially graphene nanoribbons, this estimation, which neglects the effect of temperature, already indicates the difficulties of reaching the hydrogen storage targets with clean porous carbons.More precise estimations of the amount of hydrogen stored in porous carbons have been obtained with methods which combine thermodynamics and quantum mechanics (Patchkovsky et al. 2005; Cabria et al. 2007, 2011). As a result of pumping in hydrogen by applying an external pressure P_{ext}, the amount of stored hydrogen at a temperature T is determined by the thermodynamic equilibrium between a condensed phase formed by the hydrogen adsorbed on the pore walls and a gas phase corresponding to the non-adsorbed hydrogen in the material. First, the pores are modeled as cavities formed by two parallel graphene layers separated by a distance d (other pore geometries, cylindrical and spherical, have also been studied), and the potential confining the molecules in the pore is calculated using DFT or other quantum chemical methods. When d is small, the hydrogen molecules interact with the two walls and the confinement energy is larger. The quantum states for

the motion of the H_2 molecules in the confining potential well of the pores are calculated by solving the time-independent Schrödinger's equation, and the energies ε_i of those states are used to evaluate the partition function of the adsorbed H_2 phase, $Z_{ads} = \sum_i e^{-\frac{\varepsilon_i}{kT}}$, where k is Boltzmann's constant. The ratio $Z_{ads}/Z_{free} = K_{eq}$, where Z_{free} is the partition function of the gas of compressed free molecules confined in the volume of the pores, gives the equilibrium constant K_{eq} between the adsorbed and the compressed phases. K_{eq} allows relating P_{ext} and the pressure of the adsorbed phase, or internal pressure P_{int}, through the van't Hoff equation

$$K_{eq} = e^{\frac{1}{RT}\int_{P_{ext}}^{P_{int}} V_{molar}(P,T)dP} \tag{2}$$

where R is the gas constant. A key ingredient in this equation is the molar volume $V_{molar}(P,T)$, which is known as a function of P and T from existing parameterizations of the equation of state of hydrogen. Then, P_{int} is obtained for given values of P_{ext} and T from Eq. (2). Finally, using again the equation of state, the molar volume of the adsorbed hydrogen phase is obtained in terms of T and P_{int} (or P_{ext}). The predictions at low temperature are promising (Cabria et al. 2011): at 77 K, the gravimetric and volumetric storage capacities reach values of 3 wt% and 60 g/L, respectively, and the last one fulfills the technological target. Although such temperatures are too low for automotive applications, the prediction could be useful for outer space vehicles. Figure 2 presents calculated storage capacities for three pore shapes, planar, cylindrical, and spherical, at $T = 300$ K and $P = 10$ MPa, as a
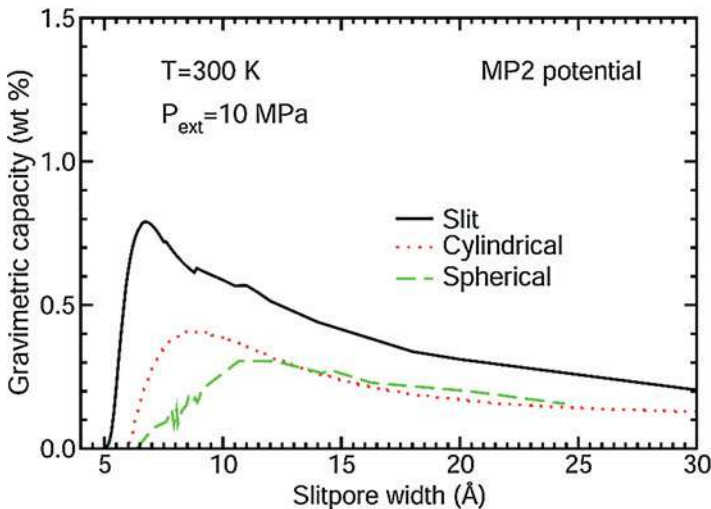
Fig. 2 Excess gravimetric hydrogen storage capacities of planar, cylindrical, and spherical pores as a function of pore size, for confining potentials obtained using the Möller-Plesset (MP2) method. The results correspond to a temperature of 300 K and an external pressure of 10 MPa. (Reproduced from Alonso et al. 2013, with permission of Cambridge Univ. Press)

function of the pore size. The gravimetric storage capacities are below 1 wt%, but this value has to be increased by a factor in between one and two, because computer simulations indicate that about 50% of the pore walls in the porous carbons expose their two sides to gas adsorption; that is, those walls are shared by two pores. Figure 2 shows a maximum in each of the curves; that is, there is an optimum pore size near 1 nm, although the precise optimal value depends on the pore shape. In summary, the calculations predict that clean porous carbons do not fulfill the targets at room temperature, and this is consistent with experiment (Yushin et al. 2006). The ultimate reason is that the adsorption energies of H_2 on pristine graphitic surfaces, planar or curved, are below 0.10 eV, too small for the purposes of storage.

4   Atomic Hydrogen on Graphene

Atomic hydrogen adsorbs on graphene on top of C atoms, and the C atoms bonded to H protrude out of the carbon plane by 0.35–0.48 Å in DFT-GGA calculations (Ferro et al. 2004; Borodin et al. 2011), changing their hybridization partially from sp^2 to sp^3. This is a chemisorption state. The chemisorption energies are 0.81–0.86 eV, although there is an adsorption barrier of 0.2 eV (Jeloaica and Sidis 1999), confirmed recently (Aréou et al. 2011). Chemisorbed single H atoms have been observed on graphene supported on SiC (0001) surfaces (Merino et al. 2015). A physisorption state has been detected in scattering experiments (Ghio et al. 1980), and the calculations indicate that the physisorbed H atoms sit on top of the hollow sites at the hexagon centers. The physisorption binding energy is 0.03 eV, a value close to the thermal energy at room temperature, $kT = 0.025$ eV, and the activation energy barrier between the physisorption and chemisorption states is only 0.13 eV (Psofogiannakis and Froudakis 2009).

Using standard techniques for depositing atomic hydrogen on graphene, the most probable clusters forming at low temperature for H coverages of 1–2 at.% are dimers. Two different configurations of dimers have been observed by STM (Hornekaer et al. 2006; Merino et al. 2015); see Fig. 3, and DFT calculations have allowed to interpret those two configurations (Hornekaer et al. 2006; Psofogiannakis and Froudakis 2009; Merino et al. 2015): a structure with the two H atoms chemisorbed on adjacent C atoms (ortho dimer) and a second structure with the two H atoms chemisorbed on opposite carbons of an hexagonal ring, that is, as third neighbors (para dimer). The two C atoms protrude out of the carbon plane. The para is slightly more stable than the ortho configuration, the chemisorption energies of the dimers (with respect to two isolated H atoms) being 2.87 and 2.60 eV, respectively. These values are larger than the sum of the chemisorption energies of two well-separated H atoms, 1.72 eV, and consequently the two chemisorbed atoms interact substantially both in the para and ortho configurations. However, the above values of the dimer chemisorption energies are smaller than the energy required to split the H_2 molecule in two H atoms (4.5 eV), and this means that chemisorption of H only occurs starting with atomic H. Recombination of the two H atoms and desorption of the molecule has been observed in temperature-programmed desorption experiments (TPD). Those experiments indicate that the ortho are more
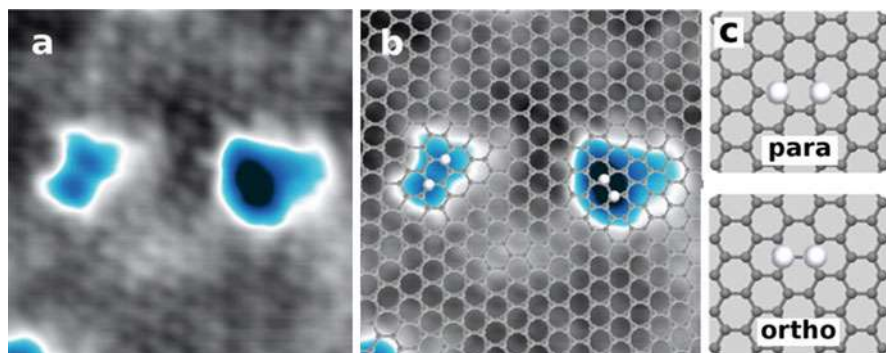
Fig. 3 (**a**) STM images of hydrogen dimers chemisorbed on a graphene layer grown on a SiC (0001) surface. In (**b**) the graphene grid has been superimposed, and the suggested position for the H atoms is indicated. (**c**) Model hydrogen dimers: para dimer (third-nearest neighbors) and ortho dimer (first neighbors). (Reproduced from Merino et al.2015, with permission of ACS)

stable than the para dimers under annealing, and this has been interpreted as the para dimers having a lower recombination energy barrier (Hornekaer et al. 2006).

The barrier for the migration of an isolated H atom from its stable position on top of a C atom to a position on top of a nearest-neighbor C atom obtained in DFT calculations is 0.99 eV, slightly less than the desorption barrier of 1.1 eV (Borodin et al. 2011), and a direct leap to a second nearest-neighbor C atom is less likely than two nearest-neighbor jumps, because the barrier is higher in the former case. Consequently, transport of atomic hydrogen on graphene is subject to a strong competition between diffusion and desorption. Molecular dynamics (MD) simulations in which the semiempirical Brenner potential (Brenner 1990) was used to model the interaction between C and H atoms have been performed to study the competition between diffusion and desorption (Borodin et al. 2011). The temperatures, 1200–2000 K, were high to allow for the observation of desorption and diffusion events within the simulation times. The simulations reveal that hydrogen atoms move by diffusion jumps and that the average number of jumps made by a hydrogen atom during its lifetime (the residence time on the graphene before it desorbs) is very small, close to one at those temperatures, slightly increasing as the temperature decreases. This is because the average lifetime increases from about 8 ps at 2000 K to about 0.5 ns at 1200 K. However, a small fraction of atoms performs more jumps, up to ten jumps for long lifetime events. Apart from jumps to first neighbor positions, correlated jump chains over two to five C–C bonds were observed. Interestingly, extrapolation to room temperature reveals that a hydrogen atom can travel two or three nanometers before detachment. This affects the nucleation and growth of hydrogen clusters on graphene: a dimer can grow by capturing H atoms deposited at distances of several Å or less from the dimer. Diffusion of H from one side of a graphene layer to the other side through a carbon hexagon is very unlikely because the calculations predict the existence of a high diffusion barrier between 3 and 4 eV (Miao et al. 2011).

Carbon atoms with coordination 2 surrounding vacancies or holes (groups of vacancies) in graphene and at the edges of graphene nanoribbons have unsaturated dangling bonds and are highly reactive, acting as catalysts for the dissociation of H_2. This molecule easily dissociates at those regions, because the dissociation barriers are small: 0.3 eV or less, depending on the particular exchange-correlation functionals used in the DFT calculations and the inclusion of dispersion interactions (Bores et al. 2012). The values of the chemisorption energies, measured with respect to the free molecule, depend on the system: 5.65 eV and 2.53 eV on zigzag and armchair edges of nanoribbons, respectively (Bores et al. 2012), and 3.06 eV on graphene vacancies (Granja-DelRío et al. 2017). The reason for the large difference between the two ribbon edges is that the separation between the two hydrogen atoms in armchair edges, 1.9 Å, is lower than in zigzag edges, 2.4 Å, and the closer proximity between the H atoms in the armchair edge weakens the C-H bonds. This becomes reflected in the changes of the density of electronic states. Adsorption and dissociation of H_2 at zigzag edges do not close the electronic gap at the Fermi level of the nanoribbons. In contrast, dissociation of H_2 at armchair edges introduces occupied states at the Fermi level and closes the gap, which is a signature of weaker C–H bonds.

Nanoribbons with zigzag edges support electronic states near the Fermi level which are localized at one edge of the ribbon (Son et al. 2006). States localized at one ribbon edge have spin orientation opposite to that of the states localized at the other edge, and this antiferromagnetic-like arrangement gives a total spin $S = 0$ to the ribbon. However, the dissociative adsorption of a hydrogen molecule at a zigzag edge changes the magnetic character of the nanoribbon, producing a configuration of non-compensated spins (Bores et al. 2012): the spin orientation of the electrons supplied by the two hydrogen atoms is opposite to that of the electrons of the edge states at the adsorbing edge, and then the net spin at the adsorbing edge is reduced, whereas the localized states at the other edge suffer no changes. As a result, the nanoribbon gets a magnetic moment of 2 μ_B. This is reflected in Fig. 4, where the difference between the electronic densities with spins up and down has been plotted. The left panel corresponds to a clean nanoribbon with excess of spin up on the upper edge and excess of spin down on the lower edge, and the right panel shows that the dissociative adsorption of H_2 at the upper edge quenches part of the spin up excess. Hydrogenation of the nanoribbon edges then provides a way to control the magnetism of carbon nanoribbons (Kusakabe and Maruyama 2003).

Graphitized glassy carbon grains present in interstellar clouds are believed to catalyze the recombination of H atoms to produce H_2 gas. Molecular dynamics simulations reveal that graphitized low density glassy carbons (densities lower than that of graphite) can be viewed as a disordered assembly of planar and curved graphene-like fragments leaving substantial space (pores) in between (López et al. 2011) and that the number of pores and their size decrease as the density of the material approaches that of graphite (Alonso et al. 2017). Those planar and curved fragments bear a substantial amount of defects, and also many open edges, mostly of zigzag type. Defects play an important role in carbon chemistry, including the above process of H–H recombination. A common family of defects is formed by holes,
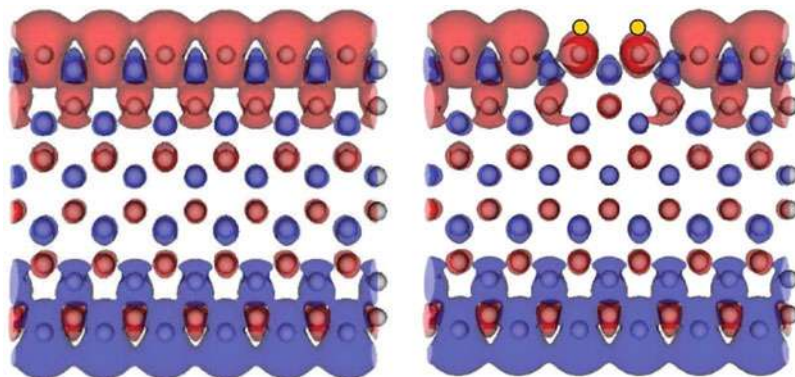
Fig. 4 Difference between spin up and spin down electron densities, $\rho_\uparrow\left(\overrightarrow{r}\right) - \rho_\downarrow\left(\overrightarrow{r}\right)$, of a clean graphene ribbon (left) and a ribbon with a dissociated hydrogen molecule at the upper edge (right). Blue and red colors represent positive and negative isosurfaces, respectively, of value 0.01 e/au^3

or multi-vacancies, having reactive unsaturated dangling bonds. Experiments have confirmed the presence of those defects in samples of chemically produced defective graphene and in samples obtained by ball milling of graphite powder (Cavallari et al. 2016). The evidence arises from neutron vibrational spectroscopy of the hydrogenated samples and, more specifically, from the comparison of the measured inelastic neutron scattering (INS) spectra with the modeled INS spectra obtained from DFT-based molecular dynamics simulations, which provide the vibrational spectrum for hydrogen on specific graphene sites. The experimental spectra are well reproduced by considering in the model a proper proportion of H atoms chemisorbed at the edges of the flakes and inside small holes. Interestingly, the INS spectra change significantly with temperature in the temperature range between 5 and 300 K, and this is due to some H atoms moving, that is, changing their local environment. Because the binding of the chemisorbed H atoms to dangling bonds on edges or void sites is strong, the only way to understand the local motion of the H atoms at those temperatures is by quantum tunneling.

Graphene can be saturated with chemisorbed atomic hydrogen and an interesting new phase results with formula CH, known as graphane. Its existence was predicted by calculations based on a cluster expansion method (Sluiter and Kawazoe 2003), and later using DFT (Sofo et al. 2007), and it has been produced by different experimental methods (Sahin et al. 2015). In graphane, every C atom of the layer is covalently bonded to an H atom, and as a consequence the hybridization in the carbon layer changes from sp^2 to sp^3. The H atoms are bonded to C atoms on both sides of the carbon plane in an alternating manner, as shown in Fig. 5; that is, every C–C bond connects C atoms with H attached at opposite sides of the layer. All the C–H bond lengths are close to 1.10 Å, which is the typical value for C–H bonds in organic chemistry, and the bond lengths of the C–C bonds are close to the single C–C bonds in diamond. Graphene is a zero-gap semiconductor; however, graphane shows an electronic gap, and the best theoretical calculations performed,

Fig. 5 Structure of graphene in chair configuration. Dark and light spheres represent C and H atoms, respectively. Notice that each C atom bonded to an H atom above the carbon layer is surrounded by three C atoms bonded to H atoms below the carbon layer. (Reproduced from Sofo et al. 2007, with permission of APS)
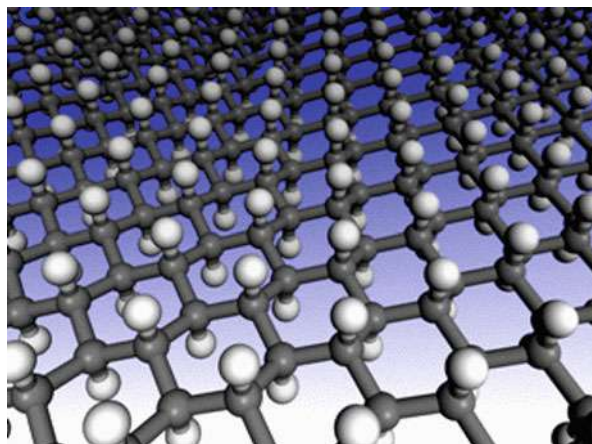
using the many-body GW approximation, predict a value of 5.2 eV (Lebègue et al. 2009). The binding energy in a GGA calculation is quite large, 6.6 eV per atom, which means that graphane is more stable than any other hydrocarbon with the 1/1 carbon/hydrogen atomic composition. The conformation described above is known as the *chair* conformation. There are other possible arrangements, but those alternative isomeric conformations have lower binding energies.

5   Adsorption of Hydrogen on Graphene Doped with Non-transition Elements

Although the technological hydrogen storage targets are not achieved with clean graphitized porous carbons, experimental evidence indicates that doping the material with metallic atoms, clusters, or nanoparticles enhances the amount of stored hydrogen (Yang 2000; Bath et al. 2010). On the other hand, when the stored hydrogen feeds the fuel cell, the H_2 molecules are dissociated in the anode, which is formed by metallic nanoparticles embedded in a porous carbon. For these reasons, the interest in understanding the details of the interaction of hydrogen with metallic atoms and clusters supported on graphene (a good model for the carbon pore walls) has become a priority.

A family of dopants enhancing hydrogen storage is formed by metals with *s* and *p* electrons, like Li or K (Chen et al. 1999; Yang 2000). In this case the beneficial effect of the dopant atoms consists in increasing the binding energy of the hydrogen molecule to the pore walls. The most stable adsorption positions for a Li atom on graphene are the hollow sites above the center of carbon hexagons, and Li rests at a distance 1.7–1.8 Å above the carbon plane. The bonding, with strength close to 1 eV, is an admixture of covalent and ionic contributions, and the Li atom is partially ionized, transferring about half an electron to graphene (Khantha et al. 2004). The adsorption energies of H_2 increase substantially on Li-doped graphene compared to

Fig. 6 Polarization of the charge of an adsorbed H_2 molecule due to the presence of a Li atom nearby. Black and white regions indicate enhancement and depletion of electronic charge, respectively, compared to a free molecule. The position of the Li atom is indicated by the small red sphere. Upper panel, the molecule is placed above a C–C bond with the molecular axis perpendicular to the C layer. Lower panel, the molecule is placed above the center of a hexagon with the molecular axis parallel to the layer plane
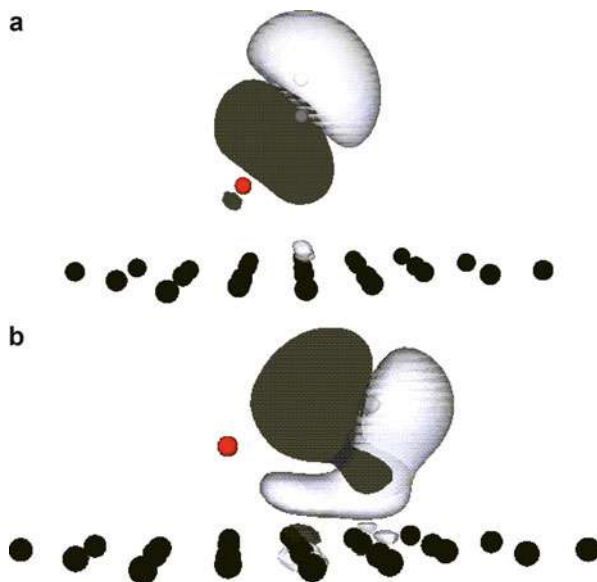
clean graphene, reaching values of 0.17–0.20 eV per molecule (Cabria et al. 2005), and the same occurs for K-doping (Froudakis 2001). The increase of the adsorption energies is due to the polarization of the charge of H_2 by the electric field of the neighbor charged alkali atoms. The polarization of the charge of the H_2 molecule is shown in Fig. 6. Due to the increased adsorption energy of H_2 in the case of Li doping, the storage capacity of Li-doped porous carbons becomes enhanced. A rough estimate along the lines of the method presented in Sect. 3 indicates that the gravimetric hydrogen storage capacity grows to about 1.5 wt% in the case of slit pores, twice the value in Fig. 2 (Alonso et al. 2013). A proposed advantage of Li is that Li atoms do not cluster on a graphitized surface (Sun et al. 2006), because the experimental (Bréchignac et al. 1994) and calculated (Brito et al. 2014) binding energies of small Li clusters are smaller than the binding energy of a Li atom to graphene.

The alkaline-earth dopants Ca and Sr show some similarities with the simpler alkali dopants. The most stable positions on graphene obtained in DFT calculations are the hollow sites also, and the adsorption binding energy of Ca atoms amounts to 0.63 eV using the PBE approximation (Chan et al. 2008) and 0.99 eV using LDA (Ataca et al. 2009). DFT-PW calculations for Ca adsorbed on C_{60} predict a larger binding energy, 1.3 eV, due to the curvature of the fullerene (Yoon et al. 2008). Those values are higher than the binding energies per atom (E_b) of small free Ca clusters (Dai and Blaisten-Barojas 2008): $E_b = 0.07$ eV (Ca_2), 0.22 eV (Ca_3), 0.44 eV (Ca_4), 0.48 eV (Ca_5), 0.51 eV (Ca_6), and 0.62 eV (Ca_7). Consequently, dispersion of Ca on graphene, instead of three-dimensional clustering, is expected. In contrast, the binding strength of Be and Mg to graphene is too weak. The stronger bonding of Ca (and Sr) to graphene arises from the transfer of electronic charge from

the valence s electrons of these atoms to the π^* bands of graphene and back donation to the previously unoccupied d orbitals of the alkaline-earth atoms through their hybridization with the π^* states. The electrostatic repulsion between the charged dopant atoms prevents clustering, and the most favorable coverage of Ca is 12.5%, that is, one Ca atom for every four C hexagons, for single-side adsorption (or 25% for double-side adsorption). The charged dopants generate an electric field favoring the adsorption of molecular hydrogen around those atoms. DFT calculations predict that a single Ca or Sr atom attached to C_{60} can adsorb up to five H_2 molecules with quite constant binding energies, 0.2 eV per molecule in GGA calculations and 0.4 eV in LDA calculations (Yoon et al. 2008). C_{60} coated with 32 Ca atoms (one on each face of the fullerene) was explored, and it was found that the system can adsorb 92 H_2 molecules, nearly three per Ca atom, with the same binding energies given above. In the case of graphene, each Ca atom binds up to four or five H_2 molecules, depending on the Ca coverage (Ataca et al. 2009).

Interesting features appear when Ca is adsorbed on narrow graphene nanoribbons because of the influence of the ribbon edges and for graphene with holes (multivacancy defects). The Ca atoms prefer to occupy the hollow sites on carbon hexagons on the ribbon edge, instead of hexagons in the central part of the ribbon (Lee et al. 2010). The charge transferred to the carbon ribbon is more localized in the first case and leads to a stronger electrostatic interaction between Ca and the substrate and to larger binding energies. On the other hand, the binding energies on those sites are sensitive to the type of edges. For nanoribbons terminated in zigzag edges, the binding energies obtained from DFT-PW calculations are 2.0 eV on the edges and 1.1 eV on the central part of the ribbon. The electron transfer leaves the Ca atoms charged, and the electrostatic repulsion prevents the charged atoms from coming too close together. The calculated optimal separation on zigzag edges is 10 Å. The binding is weaker in the case of nanoribbons with armchair edges: 0.73 and 0.68 eV on the edges and ribbon center, respectively, close to the binding energy on extended graphene. Consistent with the above results, the binding energy of Ca on C hexagons surrounding multi-vacancy holes in which the arrangement of the C atoms is of zigzag type is relatively large, 1.3 eV. The average adsorption energies of molecular hydrogen around a Ca atom doping the edge of a zigzag nanoribbon are 0.18, 0.21, 0.20, 0.17, 0.16, and 0.13 eV per molecule for 1, 2, 3, 4, 5, and 6 H_2 molecules, respectively.

A few atoms such as nitrogen or boron, which have nearly the same size as carbon and can easily form three chemical bonds, can replace C atoms in graphene forming genuine substitutional impurities. Most other atoms are larger than carbon and do not fit in a graphene vacancy, protruding outward from the carbon plane. The presence of substitutional boron affects the adsorption of hydrogen on graphene. Interestingly, the B atoms do not provide stable adsorption sites for atomic hydrogen (Ferro et al. 2003). However, the presence of B enhances the binding strength of H on C atoms neighbors to B atoms. Substituted graphene models with compositions C_7B (12.5% B) and C_3B (25% B) have been studied (Ferro et al. 2004) using DFT. The calculated chemisorption energies of atomic H on C_7B and C_3B are 2.37 and 1.70 eV, respectively, and the protrusion out of the carbon plane of the

C atom involved in the C-H bond is smaller than in pure graphene. At a first sight, it may appear that the binding energy of H does not increase with rising B concentration. However, in those two models the substitutional B atoms are distributed in such a way that there are no B-B first neighbors. Allowing for the occurrence of B-B first neighbors in modified models with C_3B composition, the adsorption energy of atomic H is 3.49 eV, indicating that the location of the B atoms has a strong influence on the layer reactivity. Obviously, the probability of finding B-B nearest-neighbors in materials prepared in the laboratory increases with B concentration, so it will be easier to find those pairs in C_3B as compared to C_7B composition. Taking this into account, a correlation arises between the B concentration and the strength of the H bonding which is consistent with the experimental observation that retention of H in graphitic materials exposed to hydrogen plasma increases with B concentration (Schenk et al. 1995). Graphitic materials are exposed to hydrogen plasma, for instance, in experiments of nuclear fusion in tokamaks because carbon materials coat the walls in contact with the plasma. B also influences the diffusion of H atoms through a doped graphene layer, from one side to the other: the activation energy barrier drops from about 4 eV in pure graphene to 0.91 eV in C_7B (through a hexagon having two B atoms) and to 0.54 eV in C_3B (through a hexagon with a B atom). The lowering of the barrier is also consistent with the increased retention of hydrogen in doped carbon materials exposed to hydrogen plasma, because the H atoms find the walls more permeable and penetrate deeply into the solid. For H diffusion on the layer plane, the activation energy, near 1 eV on pure graphene, drops to 0.5 eV for B-doped graphene with composition C_3B. In the latter case, one possible diffusion path for H is from a C atom to a first neighbor C. An alternative path is from C to a second neighbor C atom through an intermediate B atom; in this case the maximum of the activation barrier occurs when H is on top of the B atom. As discussed above, substitutional B induces a better H retention capability in graphite. At the same time, experiments indicate that recombination into molecular hydrogen and desorption of H_2 under heating becomes easier in the B-doped carbons (Schenk et al. 1995). This result is not in conflict with the larger H chemisorption energies in the doped material (Ferro et al. 2004): recombination of two H atoms and desorption of H_2 on clean and doped graphene are complex processes involving several steps with activation barriers associated, and the presence of substitutional B lowers those barriers.

6 Adsorption of Hydrogen on Graphene Doped with Transition Metal Clusters

Experimental work provides evidence for an enhancement of the storage capacity of porous carbons doped with transition metal nanoparticles: vanadium (Zacharia et al. 2005), nickel (Wang and Yang 2008), ruthenium (Wang and Yang 2008), palladium (Contescu et al. 2009; Contescu et al. 2011), and platinum (Kim et al. 2008), among others. Since the adsorption of hydrogen is a surface process, clustering of the metal

Table 1 Calculated adsorption energies of metal atoms (M) on graphene, E_{ads}, binding energy per atom of free trimers, $E_b(M_3)/at$, and hexamers, $E_b(M_6)/at$, and experimental cohesive energy or the metal, E_c. All energies are given in eV

M	E_{ads} [1]	$E_b(M_3)/at$	$E_b(M_6)/at$	E_c
Ti	1.869	2.701 [2]	3.678 [2]	4.85
Fe	0.748	2.1 [3]	3.0 [3]	4.28
Pd	1.081	1.25 [4]	1.96 [4]	3.89
Au	0.096	1.66 [5]	2.56 [5]	3.81

[1] Chan et al. (2008), [2] Kumar et al. (2007), [3] Kima et al. (2014), [4] Cabria et al. (2012), and [5] Fernández et al. (2004)

atoms to form nanoparticles is usually viewed as a disadvantage. Actually, clustering of transition metal atoms on graphene or other carbon nanostructures is common; examples can be found in work on Ti (Sun et al. 2005) and Pd (Cabria et al. 2010). The reason for clustering is that the binding strength between the dopant atoms is larger than the adsorption energy of the dopant atom on graphene. This point is illustrated in Table 1 for a few selected examples: Ti, Fe, Pd, and Au. For all of them, the adsorption energy of the atom on graphene is not only smaller than the cohesive energy of the metal but also smaller than the binding energies per atom of small clusters (the Table reports results for trimers and hexamers). The fact that the adsorption energies of metal atoms on graphene are lower than the cohesive energies of the corresponding bulk metals is a general feature, but for a few non-transition metals, the adsorption energies are larger than the binding energies per atom of small clusters, as discussed above in the specific case of Ca. The enhancement of the hydrogen storage in the doped carbons has been generally ascribed to the mechanism of spillover from the nanoparticle to the substrate (Conner and Falconer 2006; Prins 2012), which is discussed below.

Theoretical work on the interaction of hydrogen with transition metal (TM) clusters supported on graphene is abundant, and the case of palladium, a dopant of interest in experimental studies (Contescu et al. 2009, 2011; Zacharia et al. 2005), is selected here to illustrate the main features of the interaction. DFT-GGA calculations predict that the lowest energy structure of free Pd_3 is an equilateral triangle, Pd_4 is a tetrahedron, Pd_5 is a trigonal bipyramid (a second structure, a square pyramid, is very close in energy, only 0.05 eV above), and Pd_6 is an octahedron (Cabria et al. 2012). The magnetic state of all these clusters is a spin triplet (magnetic moment $M = 2 \mu_B$). In general, the most stable structure of the free Pd cluster is also the most stable structure of the cluster adsorbed on pristine graphene. Supported Pd_3 has a triangular structure vertically oriented, Pd_4 is a tetrahedron lying on a triangular face, and Pd_6 is an octahedron lying on one of its faces. The rule also works for selected larger clusters studied, like Pd_{13} (Blanco-Rey et al. 2016). The only exception in the small-size range covered above occurs for Pd_5: in this case, the lowest energy structure of the adsorbed cluster is the square pyramid lying on its base.

Table 2 Calculated binding energies of Pd_n clusters to graphene. Adsorption energies for molecular adsorption of H_2 and dissociated chemisorption (H + H) on Pd_n clusters supported on graphene and dissociation barriers. Energies are given in eV. Adsorption energies are measured with respect to separated Pd_n@graphene and free H_2. Dissociation barriers near zero are lower than 0.03 eV

	Pd_2	Pd_3	Pd_4	Pd_5	Pd_6	Pd_{13}
$E_b(Pd_n@graph)$	1.26	0.76	0.84	0.49[a]; 0.68[b]	1.14	1.41
$E_{ads}(H_2)$	1.06	0.53	0.42	0.69	0.56	0.50
$E_{ads}(H+H)$	1.66	1.95	1.47	1.24	1.19	1.42
Barrier	Near 0	Near 0	0.70	0.26	0.30	Near 0

[a]Trigonal bipyramid
[b]Square pyramid

Pd_n clusters supported on graphitized surfaces easily adsorb hydrogen. Molecular H_2 binds to the cluster usually on top of carbon atoms, and the adsorption energies vary between 0.5 and 1 eV depending on the cluster size (Cabria et al. 2012). Those binding energies are substantially higher compared to those for H_2 adsorbed on clean graphene, but it should be noticed that the molecule is bound to the cluster. The binding can be interpreted as a mixture of a covalent contribution (electron density accumulates in the region between the molecule and the Pd atom) and polarization of the charge of the molecule. The H–H bond length of the adsorbed molecules becomes slightly elongated compared to free H_2, but the molecule is only activated and the H–H bond is not broken. On the other hand, a dissociated state also exists with the two atoms separately chemisorbed on the surface of the Pd cluster, and it is, in fact, more stable than the molecular state, at least for low covering. Table 2 shows the adsorption energies for molecular and dissociative adsorption. In a cluster there are usually different adsorption sites with slightly different adsorption energies, and the adsorption energies of this Table correspond to the highest in each case. The path to dissociation starting with the adsorbed molecule needs to surmount an activation barrier whose height depends sensitively on the cluster size, being very small in some cases.

Table 2 shows that the binding strength between small Pd_n clusters and graphene is not large. Furthermore, the binding becomes weakened after the adsorption of a hydrogen molecule, and adsorption of additional molecules enhances that effect (López et al. 2014). For instance, the energies required to desorb H_2-Pd_n metal-hydrogen complexes (0.97 eV for H_2-Pd_2, 0.77 eV for H_2-Pd_3, 0.63 eV for H_2-Pd_4, 0.60 eV for H_2-Pd_5, and 1.0 eV for H_2-Pd_6) are competitive with the energies required for desorbing the H_2 molecule (the values of $E_{ads}(H_2)$ in the Table 2). This will produce an undesirable contamination effect on the fuel cell in the step of releasing the adsorbed gas from the container. The problem disappears when the clusters are anchored to defects on the graphitized pore walls. This is because graphene defects are very reactive and the binding of the clusters to the defects is strong: between 5 and 6 eV for Pd_n clusters ($n = 1$–6) anchored to graphene vacancies (López et al. 2014). In spite of the strong binding, the deformation of the

structure of the anchored Pd clusters is minor compared with free clusters or with the clusters adsorbed on perfect graphene. Fortunately, the pore walls of graphitized porous carbons contain abundant defects which play a double role. First, by tightly anchoring small clusters, the formation of large clusters is hindered. Second, the contaminating effect disappears because metal-hydrogen complexes are not freed.

H$_2$ shows two adsorption modes on Pd clusters anchored to graphene vacancies: molecular adsorption (H$_2$) and dissociative adsorption (H+H), and there is competition between those two modes. Adsorption of H$_2$ on anchored Pd$_6$ has been studied in detail (Granja et al. 2015). The most stable structure of Pd$_6$ anchored to a vacancy is the octahedron resting on one of its triangular faces. That structure shows only tiny distortions with respect to free Pd$_6$, and the adsorption energy is 5.62 eV. The graphene layer becomes distorted around the anchored cluster: the C atoms underneath the Pd cluster protrude out of the carbon layer by 0.5–0.7 Å toward the Pd cluster, and C atoms of the next shell protrude out by about 0.3–0.4 Å. Up to seven H$_2$ molecules can be adsorbed on the Pd cluster with sizable adsorption energies: the adsorption energy of the first adsorbed molecule is 0.74 eV, and the individual adsorption energies of new added molecules steadily decrease by increasing coverage, reaching the value of 0.26 eV for the seventh added molecule. Beyond the seventh, new added molecules have very low adsorption energies.

For a single molecule, however, the dissociative (H+H) adsorption configuration is more stable than molecular adsorption, although dissociation requires surmounting an activation barrier. Dissociation drives a change of the structure of the anchored Pd$_6$ cluster: the distance between two neighboring Pd atoms increases, and the cluster structure changes from the octahedron to an incomplete pentagonal bipyramid. The dissociative chemisorption energy, measured with respect to the free H$_2$ molecule and the anchored octahedron, is 1.25 eV. This energy can be viewed as the difference of two terms: the dissociative chemisorption energy on the pentagonal bipyramid, 1.36 eV, minus the energy of 0.11 eV required to transform the anchored octahedron into an incomplete pentagonal bipyramid. The anchored incomplete bipyramid supports up to three dissociated molecules, plus other six non-dissociated.

Even if the isolated Pd atom and the Pd metal are nonmagnetic, small Pd clusters show nonzero magnetic moments (Moseler et al. 2001); see Table 3 for Pd$_n$, $n = 3$–6, and Pd$_{13}$ (Cabria et al. 2012). The magnetic moments can be manipulated by depositing the clusters on graphitized substrates and by adsorption of hydrogen (López et al. 2017). The spin magnetic moment of Pd$_{13}$, $\mu = 6 \ \mu_B$, is partially quenched to $\mu = 4 \ \mu_B$ upon depositing the cluster on pristine graphene, but the magnetic moments of the smaller clusters in the table remain unchanged. Adsorption of hydrogen on the supported Pd clusters leads to further quenching of the spin magnetic moment in all cases, although with subtle differences between clusters. The magnetic moments of supported Pd$_3$ and Pd$_{13}$ decrease due to the adsorption of molecular hydrogen (from $\mu = 2 \ \mu_B$ to 0 in Pd$_3$ and from $\mu = 4$ to $2 \ \mu_B$ in Pd$_{13}$). The subsequent dissociation of H$_2$ does not change the magnetic moment anymore.

Table 3 Calculated magnetic moments (in units of μ_B) of Pd_n clusters in different situations: free Pd clusters; Pd clusters deposited on pristine graphene; with molecular hydrogen adsorbed; with dissociated hydrogen

	Pd_3	Pd_4	Pd_5	Pd_6	Pd_{13}
Free	2	2	2	2	6
Pd_n on graphene	2	2	2	2	4
With H_2	0	2	2	2	2
With H+H	0	0	0	0	2

On the other hand, the magnetic moments of Pd_4, Pd_5, and Pd_6 do no change upon adsorption of molecular H_2, and the quenching occurs when the molecule dissociates.

The mechanism responsible for the enhancement of hydrogen storage on metal-doped porous carbons is under study. Evidently, the hydrogen adsorbed on the metal nanoparticles adds up to the hydrogen directly adsorbed on the carbon pore walls. A mechanism usually assumed in experimental works is spillover (Lueking and Yang 2004). In this mechanism the incoming H_2 molecule first interacts with the adsorbed nanoparticle, dissociates there, and the two H atoms independently diffuse through the nanoparticle. After reaching the nanoparticle border, the atoms spill to the substrate, diffuse further through the substrate surface until coming at rest, and end up adsorbed on the substrate. This mechanism has been investigated by performing DFT molecular dynamics simulations for the case of Pd clusters on a graphene substrate (Blanco-Rey et al. 2016). For incoming translational energies of 0.125 eV, a fraction of the incident H_2 molecules rebound on the cluster and return to the gas, a second group of molecules end up adsorbed on the cluster, and a third group dissociates after adsorption on the cluster. However, a sufficiently large number of simulations were performed, and not a single case of hydrogen spilling toward the carbon substrate was observed. The ultimate reason is that the activation barriers for spillover are large (larger than 2 eV in the two clusters studied, Pd_6 and Pd_{13}). This result suggests taking the spillover hypothesis with some care, at least in the case of Pd nanoparticles.

7   Conclusions

Computational modeling and the density functional formalism (DFT) are useful tools helping to discover the mechanisms of the interaction of hydrogen with clean and doped carbon surfaces, modeled as a graphene layer. Understanding those mechanisms is crucial in the improvement of hydrogen fuel cells, in the development of efficient hydrogen containers for transportation and in many catalytic processes.

References

Alonso JA, Cabria I, López MJ (2013) Simulation of hydrogen storage in porous carbons. J Mater Res 28:589–604

Alonso L, López MJ, Alonso JA (2017) Computer simulations of the structure of nanoporous carbons and higher density phases of carbon. In: Angilella GGN, Amovilli C (eds) Many-body approaches at different scales: a tribute to N.H. March on the occasion of his 90th birthday. Springer, New York, pp 21–34

Ansón A, Benham L, Jagiello J, Callejas MA, Benito AM, Maser WK, Züttel A, Sudan P, Martínez MT (2004) Hydrogen adsorption on a single carbon nanotube material. Nanotechnology 15:1503–1508

Arellano JS, Molina LM, Rubio A, Alonso JA (2000) Density functional study of adsorption of molecular hydrogen on graphene layers. J Chem Phys 112:8114–8119

Aréou E, Cartry G, Layet JM, Angot T (2011) Hydrogen-graphite interaction: experimental evidences of an adsorption barrier. J Chem Phys 134:014701

Ataca C, Aktürk E, Ciraci S (2009) Hydrogen storage of calcium atoms adsorbed on graphene: first-principles plane wave calculations. Phys Rev B 79:041406

Bath VV, Contescu GNC, Baker FS (2010) Atypical hydrogen uptake on chemically-activated, ultramicroporous carbon. Carbon 48:1331–1340

Becke AD (2014) Perspective: fifty years of density-functional theory in chemical physics. J Chem Phys 140:18A301

Blanco-Rey M, Juaristi JI, Alducin M, López MJ, Alonso JA (2016) Is spillover relevant for hydrogen adsorption and storage in porous carbons doped with palladium nanoparticles? J Phys Chem C 120:17357–17364

Bores C, Cabria I, Alonso JA, López MJ (2012) Adsorption and dissociation of molecular hydrogen on the edges of graphene nanoribbons. J Nanopart Res 14:1263

Borodin VA, Vehviläinen TT, Ganchenkova MG, Nieminen MN (2011) Hydrogen transport on graphene: competition of mobility and desorption. Phys Rev B 84:075486

Bréchignac C, Busch H, Cahuzac P, Leygnier J (1994) Dissociation pathways and binding energies of lithium clusters from evaporation experiments. J Chem Phys 101:6992–7002

Brenner DW (1990) Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. Phys Rev B 42:9458–9471

Brito BGA, Candido L, Teixeira Rabelob JN, Hai GQ (2014) Binding energies of small lithium clusters: a comparison of different theoretical calculations. Chem Phys Lett 616–617: 212–216

Brown CM, Yildirim T, Newmann DA, Heben MJ, Gennett T, Dillon AC, Alleman JL, Fischer JR (2000) Quantum rotation of hydrogen in single-wall carbon nanotubes. Chem Phys Lett 329:311–316

Cabria I, López MJ, Alonso JA (2005) Enhancement of hydrogen physisorption on graphene and carbon nanotubes by Li doping. J Chem Phys 123:204721

Cabria I, López MJ, Alonso JA (2006) Density functional study of molecular hydrogen coverage on carbon nanotubes. Comput Mater Sci 35:238–242

Cabria I, López MJ, Alonso JA (2007) The optimum average nanopore size for hydrogen storage in carbon nanoporous materials. Carbon 45:2649–2658

Cabria I, López MJ, Alonso JA (2010) Theoretical study of the transition from planar to three-dimensional structures of palladium clusters supported on graphene. Phys Rev B 81:035403

Cabria I, López MJ, Alonso JA (2011) Simulation of the hydrogen storage in nanoporous carbons with different pore shapes. Int J Hydrog Energy 36:10748–10759

Cabria I, López MJ, Fraile S, Alonso JA (2012) Adsorption and dissociation of molecular hydrogen on palladium clusters supported on graphene. J Phys Chem C 116:21179–21189

Cabria I, López MJ, Alonso JA (2017) Searching for DFT-based methods that include dispersion interactions to calculate the physisorption of H_2 on benzene and graphene. J Chem Phys 146:214104

Cavallari C, Pontiroli J-RM, Johnson M, Aramini M, Gaboardi M, Parker SF, Ricco M, Rols S (2016) Hydrogen motions in defective graphene: the role of surface defects. Phys Chem Chem Phys 18:24820–24824

Chan KT, Neaton JB, Cohen ML (2008) First-principles study of metal adatom adsorption on graphene. Phys Rev B 77:235430

Chen P, Wu X, Lin J, Tan KL (1999) High H_2 uptake by alkali-doped carbon nanotubes under ambient pressure and moderate temperatures. Science 285:91–93

Conner WC, Falconer JL (2006) Spillover in heterogeneous catalysis. Chem Rev 95:759–788

Contescu CI, Brown CM, Liu Y, Bhat VV (2009) Detection of hydrogen spillover in palladium modified activated carbon fibers during hydrogen adsorption. J Phys Chem C 113:5886–5890

Contescu CI, van Benthem K, Li S, Bonifacio CS, Pennycook SJ, Jena P, Gallego NC (2011) Single Pd atoms in activated carbon fibers and their contribution to hydrogen storage. Carbon 49:4050–4058

Costanzo F, Silvestrelli PL, Ancilotto F (2012) Physisorption, diffusion, and chemisorption pathways of H_2 molecule on graphene and on (2,2) carbon nanotube by first principles calculations. J Chem Theory Comput 8:1288–1294

Dai Y, Blaisten-Barojas E (2008) Energetics, structure, and electron detachment spectra of calcium and zinc neutral and anion clusters: a density functional theory study. J Phys Chem A 112:11052–11060

De Tomás C, Suarez-Martinez I, Vallejos-Burgos F, López MJ, Kaneko K, Marks NA (2017) Structural prediction of graphitization and porosity in carbide-derived carbons. Carbon 119:1–9

Fernández EM, Soler JM, Garzón IL, Balbás LC (2004) Trends in the structure and bonding of noble metal clusters. Phys Rev B 70:165403

Ferro Y, Marinelli F, Allouche A, Brosset C (2003) Density functional theory investigation of H adsorption on the basal plane of boron-doped graphite. J Chem Phys 118:5650–5657

Ferro Y, Marinelli F, Jelea A, Allouche A (2004) Adsorption, diffusion, and recombination of hydrogen on pure and boron-doped graphite surfaces. J Chem Phys 120:11882–11888

Fiolhais C, Nogueira F, Marques MAL (eds) (2003) A primer in density functional theory. Springer, Berlin

Froudakis GE (2001) Why alkali-metal-doped carbon nanotubes possess high hydrogen uptake. Nano Lett 1:531–533

Ghio E, Mattera L, Salvo C, Tommasini F, Valbusa U (1980) Vibrational spectrum of H and D on the (0001) graphite surface from scattering experiments. J Chem Phys 73:556–561

Granja A, Alonso JA, Cabria I, López MJ (2015) Competition between molecular and dissociative adsorption of hydrogen on palladium clusters deposited on defective graphene. RSC Adv 5:47945–47953

Granja-DelRío A, Alonso JA, López MJ (2017) Competition between palladium clusters and hydrogen to saturate graphene vacancies. J Phys Chem C 121:10843–10850

Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem 27:1787–1799

Hermann J, DiStasio RA, Tkatchenko A (2017) First-principles models for van der Waals interactions in molecules and materials: concepts, theory, and applications. Chem Rev 117:4714–4758

Hornekaer L, Sljivancanin Z, Xu W, Otero R, Rauls E, Stensgaard I, Laegsgaard E, Hammer B, Besenbacher F (2006) Metastable structures and recombination pathways for atomic hydrogen on the graphite (0001) surface. Phys Rev Lett 96:156104

Jeloaica L, Sidis V (1999) DFT investigation of the adsorption of atomic hydrogen on a cluster-model graphite surface. Chem Phys Lett 300:157–162

Jena P (2011) Materials for hydrogen storage: past, present and future. J Phys Chem Lett 2:206–211

Jordá-Beneito M, Suárez-García F, Lozano-Castelló D, Cazorla-Amorós D, Linares-Solano A (2007) Hydrogen storage on chemically activated carbons and carbon nanomaterials at high pressures. Carbon 45:293–302

Jordá-Beneito M, Lozano-Castelló D, Suárez-García F, Cazorla-Amorós D, Linares-Solano A (2008) Advanced activated carbon monoliths and activated carbons for hydrogen storage. Microporous Mesopor Mat 112:235–242

Khantha M, Cordero NA, Molina LM, Alonso JA, Girifalco LA (2004) Interaction of lithium with graphene: an ab initio study. Phys Rev B 70:125422

Kim BJ, Lee YS, Park SJ (2008) Preparation of platinum-decorated porous graphite nanofibers, and their hydrogen storage behaviors. J Colloid Interface Sci 318:530–533

Kima E, Mohrland A, Weck PF, Pang T, Czerwinski KR, Tománek D (2014) Magic numbers in small iron clusters: a first-principles study. Chem Phys Lett 613:59–63

Klimes J, Michaelides A (2012) Perspective: advances and challenges in treating van der Waals dispersion forces in density functional theory. J Chem Phys 137:12090

Kumar TJD, Weck PF, Balakrishnan N (2007) Evolution of small Ti clusters and the dissociative chemisorption of H_2 on Ti. J Phys Chem C 111:7494–7500

Kusakabe K, Maruyama M (2003) Magnetic nanographite. Phys Rev B 67:092406

Lebègue S, Klintenberg M, Eriksson O, Katsnelson MI (2009) Accurate electronic band gap of pure and functionalized graphane from GW calculations. Phys Rev B 79:245117

Lee H, Ihm J, Cohen ML, Louie SG (2010) Calcium-decorated graphene-based nanostructures for hydrogen storage. Nanoletters 10:793–798

López MJ, Cabria I, Alonso JA (2011) Simulated porosity and electronic structure of nanoporous carbons. J Chem Phys 135:104706

López MJ, Cabria I, Alonso JA (2014) Palladium clusters anchored on graphene vacancies and their effect on the reversible adsorption of hydrogen. J Phys Chem C 118:5081–5090

López MJ, Blanco-Rey M, Juaristi JI, Alducin M, Alonso JA (2017) J Phys Chem C 121:20756–20762

Lueking AD, Yang RT (2004) Hydrogen spillover to enhance hydrogen storage-study of the effect of carbon physicochemical properties. Appl Catal A 265:259–268

Lueking AD, Pan L, Narayanan DL, Clifford CEB (2005) Effect of expanded graphite lattice in exfoliated graphite nanofibers on hydrogen storage. J Phys Chem B 100:12710–12717

Mattera L, Rosatelli R, Salvo C, Tommasini F, Valbusa U, Vidali G (1980) Selective adsorption of $^1H_2$ and $^2H_2$ on the (0001) graphite surface. Surf Sci 93:515–525

Merino P, Svec M, Martínez JI, Mutombo P, Gonzalez C, Martín-Gago JA, de Andres PL, Jelinek P (2015) Ortho and para hydrogen dimers on G/SiC(0001): combined STM and DFT study. Langmuir 31:233–239

Miao M, Liu Y, Wu T, Wang Q, Gubbins KE (2011) Does a hydrogen atom/proton diffuse through graphene? Diff Fundamentals.org 16:66

Moseler M, Häkkinen H, Barnett RN, Landman U (2001) Structure and magnetism of neutral and anionic palladium clusters. Phys Rev Lett 86:2545–2548

Ogden JM (2002) Hydrogen: the fuel of the future? Phys Today 55(4):69–75

Okamoto Y, Miyamoto Y (2001) Ab initio investigation of physisorption of molecular hydrogen on planar and curved graphenes. J Phys Chem B 105:3470–3474

Panella B, Hirscher M, Roth S (2005) Hydrogen adsorption in different carbon nanostructures. Carbon 43:2209–22114

Patchkovsky S, Tse JS, Yurchenko SN, Zhechkov L, Heine T, Seifert G (2005) Graphene nanostructures as tunable storage media for molecular hydrogen. Proc Nat Acad Sci USA 102:10439–10444

Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais (1992) Atoms, molecules, solids and surfaces: applications of the generalized gradient approximation for exchange and correlation. Phys Rev B 46:6671–6687

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Prins R (2012) Hydrogen spillover. Facts and fiction. Chem Rev 112:2714–2738

Psofogiannakis GM, Froudakis GE (2009) DFT study of the hydrogen spillover mechanism on Pt-doped graphite. J Phys Chem C 113:14908–14915

Sahin H, Leenaerts O, Singh SK, Peeters FM (2015) Graphane. WIREs Comput Mol Sci 5:255–272

Schenk A, Winter B, Lutterloh C, Biener J, Schubert UA, Küppers (1995) The origin of reduced chemical erosion of graphite based materials induced by boron doping. J Nucl Mater 220:767–770

Sluiter MHF, Kawazoe Y (2003) Cluster expansion method for adsorption: application to hydrogen chemisorption on graphene. Phys Rev B 68:085410

Sofo JO, Chaudhari AS, Barber GD (2007) Graphane: a two-dimensional hydrocarbon. Phys Rev B 75:153401

Son YW, Cohen ML, Louie SG (2006) Energy gaps in graphene nanoribbons. Phys Rev Lett 97:216803

Sun Q, Wang Q, Jena P, Kawazoe Y (2005) Clustering of Ti on a C_{60} surface and its effect on hydrogen storage. J Am Chem Soc 127:14582–14583

Sun Q, Jena P, Wang Q, Marquez M (2006) First-principles study of hydrogen storage on $Li_{12}C_{60}$. J Am Chem Soc 128:9741–9745

Wang L, Yang RT (2008) Hydrogen storage properties of carbons doped with ruthenium, platinum, and nickel nanoparticles. J Phys Chem C 112:12486–12494

Yang RT (2000) Hydrogen storage by alkali-doped carbon nanotubes–revisited. Carbon 38:623–641

Yoon M, Yang SD, Hicke CH, Wang E, Geohegan D, Zhang Z (2008) Calcium as the superior coating metal in functionalization of carbon fullerenes for high-capacity hydrogen storage. Phys Rev Lett 100:206806

Yushin G, Dash R, Jagiello J, Fischer JE, Gogotsi Y (2006) Carbide-derived carbons: effect of pore size on hydrogen uptake and heat of adsorption. Adv Funct Mater 16:2288–2293

Zacharia R, Kim KY, Fazle Kibria AKM, Nahm KS (2005) Enhancement of hydrogen storage capacity of carbon nanotubes via spill-over from vanadium and palladium nanoparticles. Chem Phys Lett 412:369–375

Functionalizing Two-Dimensional Materials for Energy Applications

26

Bing Huang and Su-Huai Wei

Contents

Abstract

In this chapter, the authors have systemically reviewed the electronic properties of chemically functionalized two-dimensional (2D) materials for various applications. Hydrogenation and oxidization of graphene and silicene have been predicted to be efficient approaches to open sizable band gaps in these 2D materials, and the values of the band gaps can be tuned by varying the concentration of chemical absorbers. In this way these materials become very promising for both low-gap and large-gap energy-related applications. Modulation doping of

B. Huang · S.-H. Wei (✉)
Materials and Energy Division, Beijing Computational Science Research Center, Beijing, China
e-mail: suhuaiwei@csrc.ac.cn

epitaxial 2D materials via substrates provides an alternative way to enhance the dopant and carrier densities in 2D materials. Meanwhile, the carrier mobility of the host 2D materials can be largely maintained after doping. The electronic and magnetic properties of transition-metal (TM)-doped graphene and single-layer boron nitride (BN) can be effectively controlled by the choice of the TM atoms and/or the intrinsic surface defects, which make TM-doped 2D materials holding great potential for spintronic applications. Finally, the authors show that alloying could be an important way to extend the employment of 2D materials in specific energy applications.

1   Introduction

2D materials are regarded as a new generation of nanomaterials that can exhibit unique electronic and optical properties compared to the traditional semiconductors. As in the case of traditional semiconductors, the use of two-dimensional (2D) materials for various energy-related applications is strongly dependent on the possibility of effectively controlling their electronic and optoelectronic properties via various "external" approaches. In the past years, various concepts and ideas have been proposed, including chemical functionalization, doping, and alloying, to tune the electronic properties of 2D materials and especially of graphene-based materials.

In the following, the authors will describe these ideas in several parts: in Sect. 2, the authors focus on the computational methods and models mentioned in this review chapter; in Sect. 3 the authors focus on hydrogenation and oxidization of graphene and single-layer silicon (silicene); in Sect. 4, the authors focus on the substrate-enhanced *n*- and *p*-type doping in graphene and magnetic doping properties in graphene and boron nitride; in Sect. 5, the authors focus on the discussion of structural and electronic properties of 2D C-N, Si-P, and Si-S alloys; and in Sect. 6, the authors will give a conclusion of this chapter.

2   Methods and Models

All the density functional theory (DFT) calculations the authors discuss were performed with the VASP code (Kresse and Furthmüller 1996). Projector-augmented-wave (PAW) potentials are used to describe the core electrons, and the generalized gradient approximation (GGA) for the exchange-correlation functional with the Perdew-Burke-Ernzerhof (PBE) functional (Perdew et al. 1996) is selected in our structural relaxation calculations. The Γ-centered *k*-point mesh with sufficient *k*-point density is used for the Brillouin-zone integration for all the structures. All the structures are fully relaxed until the force on each atom is less than 0.01 eV/Å. It is well known that DFT calculations both in the local-density approximations (LDA) and in the general gradient approximations (GGA) for the exchange-correlation functional underestimate the band gaps of semiconductors. Because calculations of the band gaps using hybrid exchange-correlation functionals

improve the agreement with experiment for both conventional semiconductors and low-dimensional materials (Garza and Scuseria 2016), the authors adopt the Heyd-Scuseria-Ernzerhof (HSE06) hybrid functional (Heyd et al. 2003) to calculate the electronic structures of all the systems the authors study. In particular, it has been shown that the HSE06 functional can describe the band gaps of new silicon phases very well, close to GW results (Xiang et al. 2013; Botti et al. 2012). To simulate the 2D material, a 15-Å vacuum layer is adopted to simulate isolated graphene, BN, silicene, etc.

To search for the ground-state structures at different alloy concentrations, the authors treat some of chemical functionalized 2D materials as alloy systems, and the authors use the cluster expansion (CE) method established in the alloy theory to describe the system. In the CE approach, the alloy Hamiltonian is mapped onto a generalized Ising Hamiltonian (Ferreira et al. 1989; Huang et al. 2013a). Generally, CE is able to represent any alloy energy by an appropriate selection of effective cluster interactions (J), and then it can predict the absolute lowest-energy structures within certain supercells. The J values can be determined by fitting them to the energies of selected configurations computed from first principles (van de Walle et al. 2002). The cross-validation (CV) score, obtained by a least-square fit, is used to measure the CE quality, and the authors have made sure that all the CE fittings in our study can reach a reasonably low CV score. The phase diagram of alloys is calculated by the Monte Carlo methods (van de Walle et al. 2002).

Some of the alloy structures discussed in this chapter were determined by the genetic algorithm (GA) in combination with DFT calculations to search for the global minimum structures with different alloy concentrations (Liu et al. 2007). For a given supercell with the fixed alloy concentration, the authors first generate tens of initial structures, and then the authors fully optimize both the cell parameters and the internal coordinates. To generate a new population, the authors perform the cut-and-splice crossover operator on parent structures chosen through the tournament selection. The next generation is chosen from the lowest-energy clusters among the parents and offspring. The process is repeated until the ground-state structure is located.

3 Chemical Functionalization of 2D Materials

Graphene, a single layer of carbon atoms arranged in a honeycomb lattice, has been a focus of recent research efforts (Novoselov et al. 2005, 2006; Zhang et al. 2005) due to its unique semimetallic zero-gap electronic structure and the massless Dirac fermion behavior. The unusual electronic and structural properties make graphene a promising material for the next generation of faster and smaller electronic devices. However, graphene lacks an essential feature for controlled and reliable transistor operation and energy applications, namely, a band gap. Several schemes have been proposed to open a band gap in graphene. Among them, chemical functionalization, via hydrogen- and oxygen-related chemical groups (-O, -OH, -COOH, etc.), is found to be a promising way to manipulate the electronic properties of graphene. Similar

to the studies in graphene system, chemical functionalization of other 2D systems by hydrogen and oxygen groups has also been extensively studied in the past years.

3.1   Hydrogenation of Graphene

The hydrogenation of graphene, as a prototype of covalent chemical functionalization, has been studied extensively (Stojkovic et al. 2003; Tsetseris and Pantelides 2012; Boukhvalov et al. 2008; Slijivancanin et al. 2011; Sofo et al. 2007). However, the fully hydrogenated graphene, known as graphane C_1H_1, is an insulator with a very large band gap (>5 eV) (Sofo et al. 2007), which is unsuitable for most energy-related applications. Therefore, some ideas were proposed to tune the band gap of 2D graphene by varying the hydrogen concentration.

Firstly, the hydrogenation of infinite 2D graphene is examined. Full hydrogenation of graphene has been extensively studied (Sofo et al. 2007), but the mechanism of hydrogenation is not well understood. It is well known that hydrogen atoms tend to absorb on the top of carbon atoms. To find the lowest-energy structure of graphene with different coverages of hydrogen atoms, Xiang et al. (2009) used the "cluster expansion" (CE) method (van de Walle et al. 2002). First, they consider the C_2T_4 alloy (see Fig. 1a) with the top sites T occupied by H atoms or vacancies. It is found that there are four important pair interactions as shown in Fig. 1a. The interaction between two H atoms adsorbed on the same C atom is repulsive (the J value for this configuration is 0.34 eV). However, the interaction parameter between two H atoms adsorbed on different sides of the two adjacent C atoms is attractive. Using the cluster expansion Hamiltonian, they can obtain the energy of a given alloy configuration and thus the ground-state structure of the partially hydrogenated graphene. As shown in Fig. 1b–e, the strong H clustering or phase separations always occur at different H concentrations on graphene surface; therefore, the global ground state of a partially hydrogenated graphene will divide into graphene and graphane parts (Xiang et al. 2009).

The strong phase separation in graphene-graphane system can be applied to modulate the electronic properties of finite graphene, e.g., with 1D edges (Xiang et al. 2009). First, one can consider the adsorption of a single H atom on a zigzag-edge-terminated graphene nanoribbon, as shown in Fig. 2a. The calculated adsorption energies are shown in Fig. 2b, and the positive adsorption energy is a consequence of the formation of C-H bond. Therefore, the added H atom prefers to locate at the edge carbon atom (site 1) than other sites by at least 1.1 eV, because the number of carbon-carbon p_z-p_z π bonds is the largest in this configuration. Experimentally, it was found that the atomic layer deposition of metal oxide on graphene grows actively on edges, indicating that the chemical reactivity at the edges of graphene is high (Wang et al. 2008, 2009). Interestingly, the dependence of the adsorption energy on the distance between the adsorbed site and the edge is not monotonic. It displays an odd-even oscillation with a smaller adsorption energy at even sites, and the adsorption energy of even (odd) sites increases (decreases) with the distance; eventually the energy difference between even and odd site adsorptions becomes

Fig. 1 (**a**) The C2T4
structure used in the cluster
expansion calculations. T
refers to the top site of
graphene. The important pair
interactions are indicated by
arrows. The numbers (in eV)
give the pair interaction
parameters of the cluster
expansion expression. (**b**) The
ground-state structure
(graphane) of hydrogenated
graphene with $n[H]/n[C] = 1$.
(**c**) The lowest-energy
structure of hydrogenated
graphene with
$n[H]/n[C] = 0.5$ among all
possible structures with no
more than four carbon atoms
per cell. (**d**) The
lowest-energy structure of
hydrogenated graphene with
$n[H]/n[C] = 0.5$ among all
possible structures with no
more than eight carbon atoms
per cell. (**e**) The ground-state
structure of a very large
graphene cell with 24
adsorbed H atoms. The
enclosed region by dashed
lines denotes the cell.
(Adopted from Xiang et al.
2009)
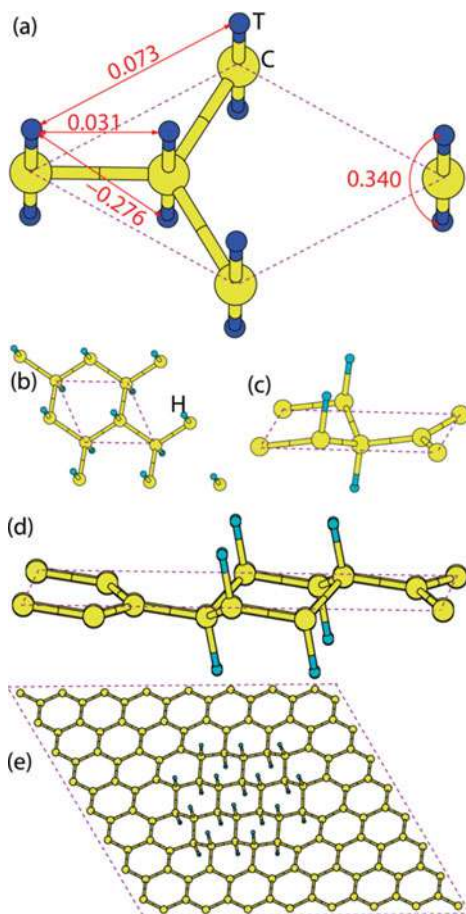
very small. The smallest adsorption energy at site 2 might be due to the presence of
two rather unstable edge carbon atoms near site 2 that participate in the formation
of only one π-π bond.

A second H atom will adsorb on the opposite site of the carbon atom (site 2)
adjacent to the edge carbon atom (site 1) to which the first H atom is bound, so as
to saturate the broken bond. This configuration is more stable than that with two
H atoms adsorbed on two outermost edge carbon atoms, for which the adsorption
energy is about twice the adsorption energy of a single H atom on an edge carbon
atom. The third H atom is expected to adsorb on the edge carbon atom adjacent to
the carbon atom to which the second H atom is bound. If the number of H atoms is
equal to the total number of edge carbon atoms (on both edges), all the H atoms will
adsorb on the outermost carbon atoms of one edge of graphene, as shown in Fig. 2c.
When the number of H atoms is twice the total number of edge carbon atoms, the
excess H atoms will bind with the outermost carbon atoms of another edge, resulting
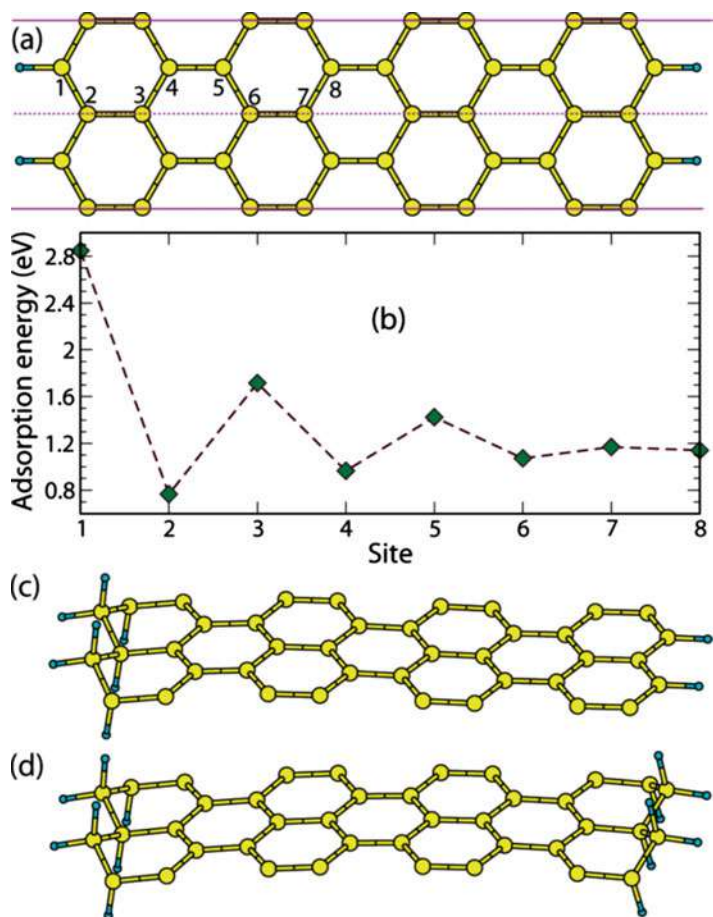
Fig. 2 (**a**) The structure of zigzag-edge graphene with numbers indicating different nonequivalent carbon sites. The supercell adopted in the DFT calculations is denoted by solid horizontal lines. (**b**) The adsorption energy of one H atom on different carbon sites as labeled in (**a**). (**c**, **d**) Ground-state structures of 8-ZGNR with one and two adsorbed H atoms per edge carbon atom, respectively. (Adopted from Xiang et al. 2009)

in a symmetric configuration (see Fig. 2d). The H atoms adsorb on the outermost zigzag bare carbon chain of one edge and then on the outermost zigzag bare carbon chain of the other edge. The process of alternating hydrogenation continues until no more free H atom is available. It should be noticed that this conclusion is very general and independent on the specific edge configurations.

To understand the electronic properties of partially hydrogen-functionalized graphene, Xiang et al. compare partially hydrogenated zigzag-edge graphene nanoribbon (ZGNR) with four bare zigzag carbon rows and hereafter referred to as 8-ZGNR-4, with 4-ZGNR without adsorbed H atoms as a representative example (Xiang et al. 2009). In both cases, they find that the electronic ground

state is the antiferromagnetic (AFM) state in which each of the two electronic edge states is ferromagnetically ordered but the two edge states are antiferromagnetically coupled to each other (Son et al. 2006). Moreover, they have nearly the same local magnetic moment ($\sim$0.10 μ_B). Because the density functional theory (DFT) calculations based on local-density approximations (LDA) or generalized gradient approximations (GGA) are well known to seriously underestimate the band gap of semiconductors, they calculate the band structure of partially hydrogenated graphene by employing the HSE06 functional (Heyd et al. 2003), which was shown to give a good band gap for many semiconductors including ZGNRs (Hod et al. 2008). The HSE06 band structures calculated for 4-ZGNR and 8-ZGNR-4 in the AFM state are shown in Fig. 3a, b, respectively. In the energy region of the band gap, the band structure of 8-ZGNR-4 is similar to that of 4-ZGNR, and the band gap (1.44 eV) of 8-ZGNR-4 is close to that (1.53 eV) of 4-ZGNR. Therefore, the electronic properties of partially hydrogenated graphene are determined by those of its graphene part, i.e., the bare zigzag carbon rows. Moreover, Xiang et al. also predict that ZGNRs with a small number of bare zigzag carbon rows, e.g., N = 4, are promising solar cell absorption materials; 8-ZGNR-4 (and N-ZGNR-4 with N > 4) has a direct band gap that is close to the optimal value ($\sim$1.5 eV) (Thompson and Frechet 2008) for the solar energy harvesting and high carrier mobility.

Besides double-sided functionalized graphene, which can be realized in a chemical functionalized freestanding graphene, Xiang et al. have studied the single-sided hydrogen-functionalized graphene (Xiang et al. 2010a), which might be realized in a functionalized epitaxial graphene. The thermodynamically stable single-side hydrogenated graphene structures are nontrivial and unexpected, as shown in Fig. 4 for several different H concentrations. This is significantly different from the case of

Fig. 3 Electronic band structures of (**a**) 4-ZGNR and (**b**) partially hydrogenated 8-ZGNR with four bare zigzag carbon rows (i.e., 8-ZGNR-4) in the AFM state from HSE06 calculations. The horizontal dashed lines denote the top of the valence bands. The insets show the structures of 4-ZGNR and 8-ZGNR-4. (Adopted from Xiang et al. 2009)
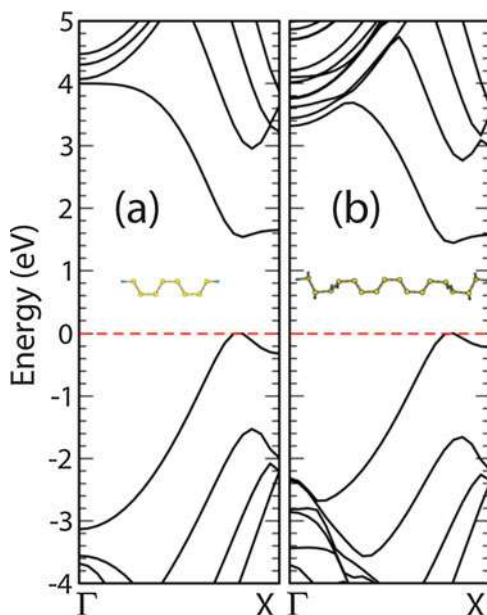
Fig. 4 (**a** and **b**): The low-energy structures of two single-side hydrogenated graphene at a H concentration of 1/8. (**c**) and (**d**) are two proposed functionalized structures in Boukhvalov et al. (2008) and Slijivancanin et al. (2011), respectively. The numbers (in eV/H) in rectangles denote the average adsorption energies. The unit cells are marked. The hydrogen coverages (Hc) are shown. (Adopted from Xiang et al. 2010a)
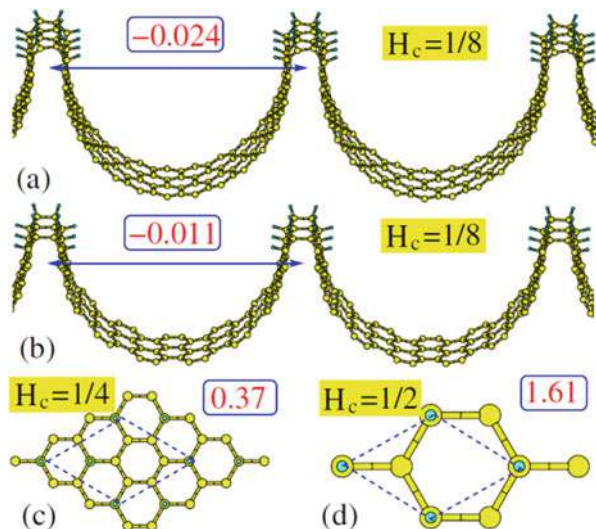
double-sided functionalization of graphene. The peculiar adsorption configuration is a consequence of the competition between electronic kinetic energy and elastic bending strain energy: the hydrogen atoms tend to be close to each other to lower the kinetic energy of carbon π electrons, while the phase separation between sp^3 carbon part and sp^2 graphene is prohibited by the large strain caused by the single-sided hydrogen adsorption. Previous studies based on the same computational methods as those of Xiang et al. (2010a) suggested two possible single-side hydrogenated graphene structures (Boukhvalov and Katsnelson 2008a; Zhou et al. 2009), as shown in Fig. 4c, d. However, the formation energies are much higher (by at least 0.3 eV) than the one proposed by Xiang et al. (2010a), as shown in Fig. 4. This suggests that the macroscopic single-sided hydrogen adsorption on graphene is a cooperative phenomenon, which can hardly be described by the adsorption preference of a small number of hydrogen atoms. It should be noted that even with the same H coverage (0.25 or 0.5), the single-side hydrogenated graphene structure based on the new bent models has a much lower energy although the single-side hydrogenated graphene is thermodynamically unstable at this H coverage.

3.2   Oxidization of Graphene

Graphene oxide (GO) is of particular interest since chemical reduction in GO has been demonstrated as a promising route for mass production of graphene (Eda et al. 2008; Tung et al. 2009; Li et al. 2008; Gilje et al. 2007; Stankovich et al. 2006; Brodie 1859). In addition, GO shows promise for use in several technological applications such as polymer composite (Li et al. 2008), dielectric layers in nanoscale electronic devices, and the active region of chemical sensors.

GO, which was first prepared by Brodie in 1859 (Stankovich et al. 2006), consists of oxidized graphene layers with the hexagonal graphene topology (Wilson et al. 2009). Although it is now widely accepted that GO bears hydroxyl and epoxide functional groups on its basal plane (Gao et al. 2009), as confirmed by a recent high-resolution solid-state 13C-NMR measurement (Cai et al. 2008), the complete structure of GO has remained elusive because of the pseudorandom chemical functionalization of each layer.

Several first-principles studies (Kudin et al. 2008; Boukhvalov and Katsnelson 2008b; Yan et al. 2009) have been performed to investigate the structure and energetics of epoxide and hydroxyl groups on single-layer graphene. And some possible structures were manually selected and examined to obtain a good structural model for GO. However, the selection of these structures may not be the real ground-state structure of GO, which may give incorrect conclusions on the electronic properties of GO in the real samples. Xiang et al. adopted the global structural search method in combination with density functional theory (DFT) to search the global minimum structures of oxidized graphene with different oxidation levels (Xiang et al. 2010b). As a first step, they considered only the arrangement of epoxide groups on single-layer graphene to simplify the model.

The GA simulations (Xiang et al. 2010b) reveal two low-energy structures of C_1O, as shown in Fig. 5. In both structures, each of the carbon atoms is bonded to two neighbor carbon atoms and two O atoms; thus there is no sp^2 carbon. The structure shown in Fig. 5b is rather simple. There are epoxy pair chains along the zigzag direction which are parallel to each other. The authors will refer to this model as the epoxy pair model (D_{2h} symmetry) hereafter. The formation of an isolated epoxy pair was suggested in a theoretical study on the graphene oxidative process (Li et al. 2009; Zhang et al. 2009a). It was also shown that an isolated epoxy pair is less stable than a carbonyl pair (Li et al. 2009). However, in the fully oxidized graphene case, it is found in Xiang et al. (2010b) that that the epoxy pair chain is more stable than the carbonyl pair chain. The lowest-energy structure (C_{2v} symmetry) (Fig. 5a) of C_1O has a lower energy than the epoxy pair model. The small energy difference suggests that both phases might coexist at certain finite temperature. In the lowest-energy structure, there are isolated six-membered carbon rings, which are connected to the neighboring six-membered carbon rings through two epoxy pairs and four unzipped epoxy groups. For each six-membered carbon ring, there are two normal epoxy groups. Therefore, the complex structure is termed as the mix model. One can obtain the epoxy pair model from the mix model by moving the unzipped epoxy groups on top of the normal epoxy groups.

When the oxygen concentration is reduced, i.e., the O/C ratio is less than 1, it is found that the strong phase separation occurs, which will separate the CO_x ($x < 1$) structure into CO_1 part and graphene part (Huang et al. 2013a; Xiang et al. 2010b). The origin of this phase separation can be understood by the additional cluster expansion (CE) calculations (Huang et al. 2013a).

The formation energy (E_f) of selected 120 CO_x different structures is calculated using ab initio methods with the maximum O/C ratio of 1, as shown in Fig. 6a. The J values defining the CE are obtained by fitting to the DFT-GGA energy
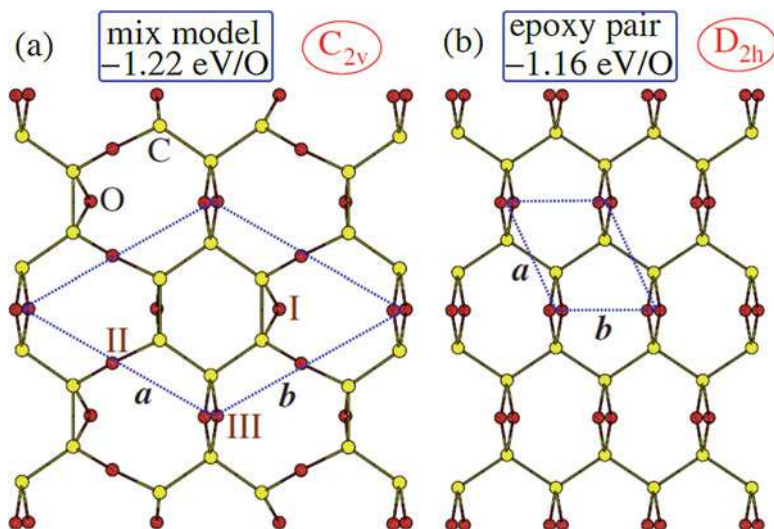
Fig. 5 (**a**) The structure of the mix model (C_{2v} symmetry) which is predicted to be the lowest-energy C_1O phase. (**b**) The low-energy C_1O phase with the epoxy pair model (D_{2h} symmetry), which has a higher energy by only 0.06 eV/O than the mix model. "I," "II," and "III" in (**a**) indicate normal epoxy, unzipped epoxy, and epoxy pair, respectively. The unit cells are enclosed by dashed lines. The numbers give the formation energy of the oxidized graphene structures. (Adopted from Xiang et al. 2010b)

values of selected GO structures, and the CE contains 26 J up to four-point clusters which are shown in Fig. 6b. Apparently, J are dominated by negative values (attractive interactions for oxygen atoms), and the strengths of negative J are significantly larger than those of positive ones (repulsive interactions for oxygen atoms). Interestingly, the authors find that these negative J are largely (>60%) contributed by the interactions of oxygen atoms at the different graphene sides. For example, the two largest negative J values for pairs (marked as A and B in Fig. 6b) are contributed by the interactions between the nearest and third-nearest oxygen atom pairs at different sides, while the largest negative J for triple cluster oxygen interaction (marked as C in Fig. 6b) is a combination of A and B, as shown in Fig. 6c. The constructed CE are then used to calculate the E_f of all enumerated symmetry-inequivalent structures ($\sim$25,000), and the results are shown in Fig. 6a. Overall, the authors find that there are no intermediate ground states for $0 < x < 1$, which is consistent with the dominant negative values of J for short-range pairs and strongly indicates that the ground states of CO_x ($x < 1$) will separate into a graphene phase and a CO_1 phase. As the typical values of x in CO_x in the current experiments are usually smaller than 0.5, the authors can understand that inhomogeneous phase generally exists in almost all the GO samples. Our calculations explain quite well the experimental discovery that thermally reduced GO is separated into unoxidized
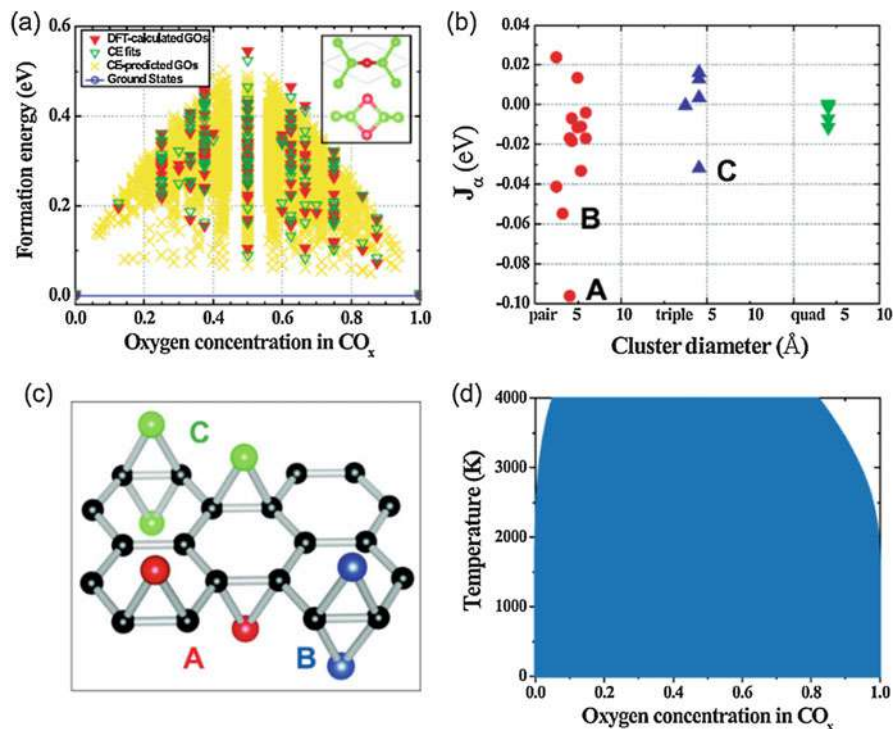
Fig. 6 (**a**) The calculated formation energies of CO_x (with respect to graphene and CO_1 shown in Fig. 5b) along with the corresponding CE fits as a function of oxygen concentration x in the GO system. The formation energies of 25,000 symmetry-inequivalent structures calculated from CE are also plotted here. Inset: the structure of CO_1. (**b**) Effective cluster interactions J as a function of cluster diameter, fitted with CE (the empty and point clusters are excluded). (**c**) The cluster figures associated with the two largest negative J values for pairs (marked as A and B in (**b**)) and the largest negative J value for triple cluster (marked as C in (**b**)) are shown as different colors. (**d**) The MC-calculated phase diagram in CO_x alloys. The area of phase separation in this phase diagram is also shaded blue. (Adopted from Huang et al. 2013a)

graphene regions and high-oxidized GO regions that have a quasi-hexagonal unit cell with an unusually high 1:1 O:C ratio (Mattson et al. 2011).

One may expect to enhance the solubility of GO alloys at high growth temperature, which is proved to be a very successful idea in the growth of many conventional alloys like III nitrides (Ho and Stringfellow 1996; Bergman and McHale 2011) and Cu(In, Ga) Se_2 solar cell materials (Wei et al. 1990, 1999). In order to investigate this possibility, the phase diagram for graphene CO_x (0 < x < 1) is calculated by Monte Carlo simulations, as shown in Fig. 6d. The authors find that the miscibility temperature of GO even at quite low oxygen concentration (x < 0.05) is around 4000 K, which is rather high and close to the melting point of graphite (Bundy 1989). The calculated phase diagram clearly demonstrates that it is impossible to get homogeneous GOs by the conventional Hummers methods under the typical

temperature for preparing thermally reduced GOs (Brodie 1859; Wilson et al. 2009; Gao et al. 2009; Cai et al. 2008).

3.3   Chemical Functionalization of Silicene

Silicon is the leading material in semiconductor industries because of its wide usage in transistors, solar cells, and other electronic devices. Silicene, the silicon equivalent of graphene, has recently attracted significant attention because it is predicted to be stable and has graphene-like electronic structures including the presence of the Dirac cone at the Fermi level (Takeda and Shiraishi 1994; Fagan et al. 2000; Guzman-Verri and Lew Yan Voon 2007; Cahangirov et al. 2009; Zhang et al. 2017) and because of its compatibility with mature Si-based electronics. Recent experimental data show that epitaxial silicene can be successfully grown on some metal substrates, such as Ag (Lalmi et al. 2010; Vogt et al. 2012; Feng et al. 2012; Chen et al. 2012), Ir (Meng et al. 2013), and ZrB_2 (Fleurence et al. 2012). Similar to graphene, the feasibility of employing silicene for various electronic applications depends largely on the possibility of controlling its electronic properties by chemical functionalization. In contrast to graphene, it is well known that Si favors sp^3 hybridization, whereas sp^2 hybridization is dominated in silicene. This special feature is expected to give rise to very unique properties during chemical functionalization of silicene. For example, the external adsorbate or dopant may largely perturb the unfavorable sp^2 Si bonds and induce some unexpected structural transitions.

For silicene, F, O, and H are selected as typical adsorbates for chemical functionalization because they are successful for modifying the electronic properties of graphene (Huang et al. 2013b). Our calculations in Vogt et al. (2014) show that F (H) prefers to stay on the top of the Si atom, while O prefers to stay at the bridge site of a Si-Si bond, as in graphene. However, because silicene is less stable than bulk Si, currently, it can only be epitaxially grown on some special substrates to maintain the 2D structure and avoid the formation of bulk Si. For epitaxial silicene, the adsorbates are expected to attach only on the top basal plane of silicene. Therefore, the authors first focus on (single-sided) F- or O- functionalized silicene.

For pristine silicene (inset of Fig. 7a), the calculations in Huang et al. (2013b) show that the Si-Si bond length is 2.25 Å (shorter than 2.36 Å in bulk Si) and the buckling height is 0.44 Å, in agreement with previous calculations (Chen et al. 2012). For fluorinated silicene (SiF_x), the highest F concentration is $x = 1$. In the structure of SiF_1 (inset of Fig. 7a), the Si-Si bond length is stretched to 2.41 Å, and then silicene becomes flat. Although the Si atoms in SiF_1 are fully fourfold coordinated, the bond lengths and bond angles are far from those pertinent to sp^3 hybridization in bulk Si. The E_f values of selected 115 SiF_x structures are calculated using first-principles methods, and they are in the range of -0.4 eV$\sim$0.4 eV, as shown in Fig. 7a. The CE Hamiltonian of SiF_x is obtained by fitting to the 115 first-principles energies, and then all the energies of SiF_x alloys are calculated using the CE approach, and the results are shown in Fig. 7a. Interestingly, three intermediate ground states are discovered, as plotted in Fig. 7c–e for $Si_{12}F_4$, Si_6F_3,

and $Si_{12}F_{11}$, respectively. The intermediate ground states are determined by the fact that all the E_f of SiF_x with the concentrations between two neighboring intermediate ground states are above the ground-state line. For $Si_{12}F_4$ and Si_6F_3 (Cs symmetry), spontaneous structural warping appears after F adsorption; i.e., the arrangement of the nearly planar honeycomb lattices in silicene is now transformed to some ordered rippled structures. In the structure of $Si_{12}F_4$ (Si_6F_3), 8 (5) out of 12 (6) Si atoms are fourfold coordinated with the neighboring Si and F atoms to maximize the sp^3 hybridization and lower the total energies, and the average bond length between fourfold-coordinated Si is the same as that in bulk Si. For the structure of $Si_{12}F_{11}$, the basic honeycomb structure is maintained after fluorination, as shown in Fig. 7e. The only Si atom without F adsorption in $Si_{12}F_{11}$ is pushed down and forms a bond length of 2.35 Å with the neighboring Si atoms to maximize the ideal sp^3-like hybridization.

For the epitaxial SiOx, the highest O concentration is x = 0.5 and its structure is shown in Fig. 7b. The E_f with a total of 97 alloy structures are calculated using first-principles methods and fitted by the CE Hamiltonian. Using the CE Hamiltonian, our search identified a single intermediate ground-state $Si_{14}O_4$, as shown in Fig. 7b. In the structure of $Si_{14}O_4$ (C_{2V} symmetry), the original hexagonal sublattice shrinks to a rippled rectangular sublattice, as shown in Fig. 7f. Interestingly, in this structure all the Si atoms form four bonds with the neighboring Si or O atoms to realize the sp^3 hybridization, while each O atom forms two bonds with the neighboring Si atoms. This kind of spontaneous structural transition results in a new ground state. Furthermore, the authors find that chemical functionalization of epitaxial silicene is irreversible because of the induced structural transitions. This scenario is significantly different from that of graphene.

It is interesting to understand the chemical functionalization effects on the electronic properties of silicene. The electronic band structures for these discovered intermediate ground states of SiF_x and SiO_x are shown in Fig. 8. The authors find that $Si_{12}F_4$ is an intrinsic semiconductor with a slightly indirect band gap of 1.11 eV, similar to bulk Si. However, the energy difference between the direct and indirect band gaps (0.07 eV) in $Si_{12}F_4$ is extremely small based on the HSE06 calculations, compared to that of bulk Si (2.3 eV in experiments or HSE06 calculations), indicating that it may replace bulk Si for optoelectronic applications. The CBM and VBM of $Si_{12}F_4$ are mainly contributed by the electrons of the threefold-coordinated Si atoms. Interestingly, Si_6F_3 behaves as an *n*-type semiconductor with an indirect band gap of 0.51 eV. In Si_6F_3, there is only one Si atom that is threefold coordinated, and the structure could be considered a silicon fluoride system with an F (anion) vacancy. The anion vacancy state has one unpaired electron close to the conduction band and is strongly hybridized with the conduction band. Similar to Si_6F_3, $Si_{12}F_{11}$ can also be considered a silicon fluoride with an F vacancy, but the defect state in this 2D structure is quite localized and is resonant inside the valence band. As a result, $Si_{12}F_{11}$ behaves as a *p*-type semiconductor with a direct band gap of 2.17 eV. $Si_{14}O_4$ shows unexpected metallic behavior (negative band gap), partially because of the large sublattice distortion. The band crossing the Fermi level is mostly contributed by the hybridized *3s* and $3P_z$ states from the Si atoms (those that are not bonded with oxygen atoms).
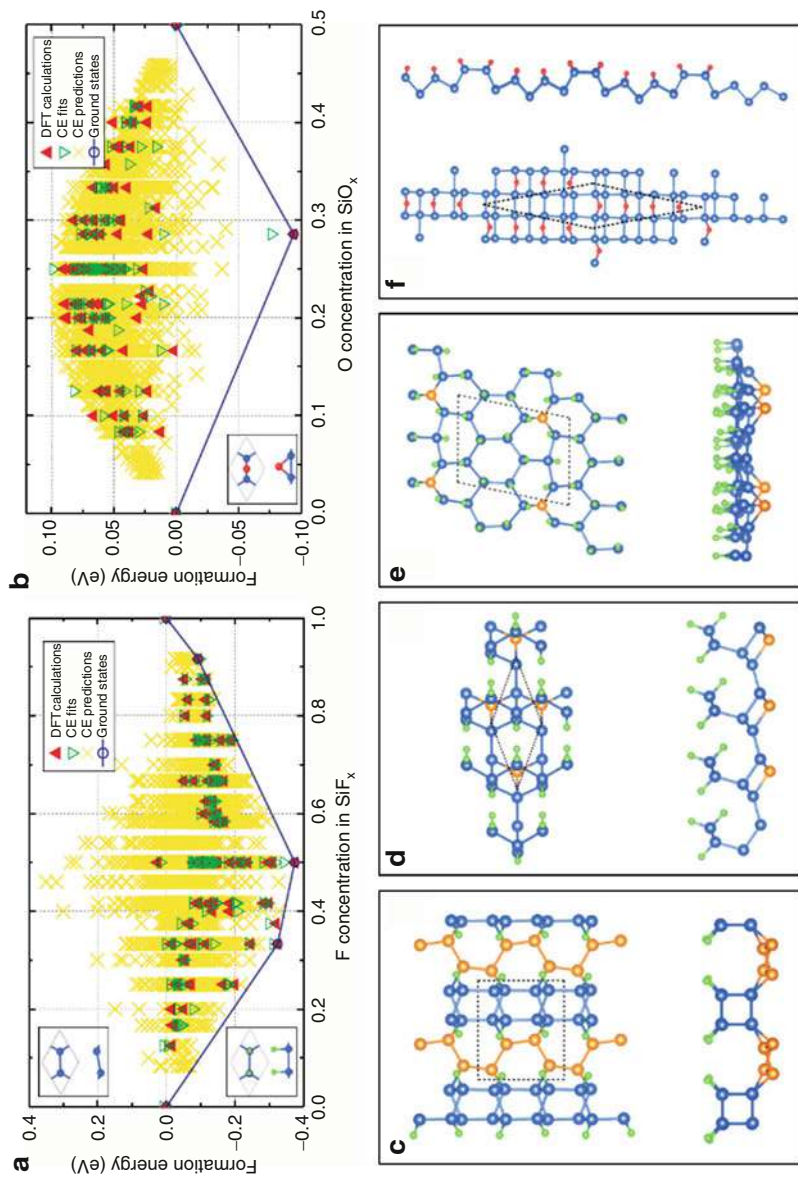
Fig. 7 (continued)

Besides single-layer silicene, bilayer silicene has also been synthesized in the experiments (Vogt et al. 2014). Especially, bilayer silicene is more stable than single-layer one. Therefore, the authors have also studied the chemical functionalized bilayer silicene. By using a similar process for the study of chemical functionalized monolayer silicene, the authors have studied hydrogenated bilayer silicene. Interestingly, for single-sided hydrogenation, six intermediate ground states are discovered, as plotted in Fig. 9 for $Si_{16}H_2$ (C_{2v} symmetry), $Si_{28}H_6$ (C_1 symmetry), $Si_{24}H_6$ (C_2 symmetry), $Si_{28}H_8$ (C_2 symmetry), $Si_{12}H_4$ (C_{2v} symmetry), and $Si_{16}H_6$ (C_{2v} symmetry). One common characteristic of these ground states is that the adsorbed H atoms on the top of some specific Si atoms break the underneath Si-Si bonds and induce some local structural distortions. The average Si-Si bond length (2.34 Å) and bond angle (106°) at the structural distortion sites are close to those in bulk Si, which indicates that the induced distortion is to generate the sp^3 hybridization and lower the total energies. Generally, the hexagonal sublattice of silicene is kept during hydrogenation, which is significantly different from that of hydrogenated single-layer silicene (Huang et al. 2013b) and consistent with the fact that bilayer silicene is more stable than single-layer silicene.

For double-sided SiH_x, six intermediate ground states are discovered at specific H concentrations, which are quite different from chemically functionalized double-sided graphene (Huang et al. 2013a). Usually, there are no intermediate ground states in chemically functionalized freestanding 2D materials with single-atom thickness because the existence of strong double-sided adsorbate-adsorbate interactions can give rise to the clustering of adsorbates (Huang et al. 2013a). For double-sided hydrogenated bilayer silicene, the formation of the sp^3 bond on one side and the subsequent increase in the layer-layer distance significantly reduce the strength of the attraction of H on opposite sides. This explains why some intermediate ground states can form in double-sided hydrogenated bilayer silicene. As shown in Fig. 4, those structures are $Si_{16}H_4$ (D_{6h} symmetry), $Si_{12}H_4$ (D_{2h} symmetry), Si_8H_4 (D_{2h} symmetry), $Si_{12}H_8$ (D_{2h} symmetry), $Si_{16}H_{12}$ (D_{2h} symmetry), and $Si_{12}H_{10}$ (D_{3h} symmetry). In these six ground states, it is found that H atoms prefer to symmetrically adsorb on the top and bottom sides of silicene, giving rise to relatively higher structural symmetry compared to single-sided hydrogenation. Similar to single-sided SiH_x, H adsorption breaks the underneath Si-Si bonds, and the local structural distortion around Si atoms can generate the ideal sp^3 hybridization. The calculated average binding energy between H atoms and bilayer silicene in these ground states is around 820 meV/H. Generally, the ground

Fig. 7 (**a**) The calculated formation energies (E_f) of epitaxial SiF_x along with the corresponding CE fits as a function of x. The E_f of 6000 symmetry-inequivalent structures calculated from CE are also plotted here. Inset: the top and side views of the structures of silicene and SiF_1. (**b**) is the same as (**a**), but for epitaxial SiO_x, the 10,000 symmetry-inequivalent structures calculated from CE are shown. The top and side views of the structure of $SiO_{0.5}$ are shown as an inset in (**b**). The top and side views of the four discovered intermediate ground-state structures for (**c**) $Si_{12}F_4$, (**d**) Si_6F_3, (**e**) $Si_{12}F_{11}$, and (**f**) $Si_{14}O_4$. In these structures, the blue, green, and red atoms represent the Si, F, and O, respectively. The three-bonded Si atoms are highlighted by orange color, and the unit cells are marked by a dashed line. (Adopted from Huang et al. 2013b)
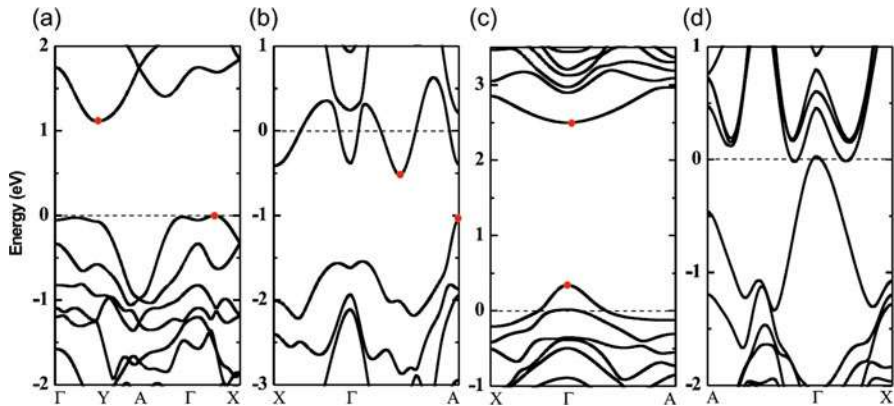
Fig. 8 The electronic band structures for these discovered ground states of epitaxial SiF_x and SiO_x: (**a**) $Si_{12}F_4$, (**b**) Si_6F_3, (**c**) $Si_{12}F_{11}$, and (**d**) $Si_{14}O_4$. The positions of CBM and VBM are highlighted by red dots. The Fermi level is set to zero. (Adopted from Huang et al. 2013b)

state of a single- or double-sided SiH_x structure with an arbitrary concentration could be phase separated into its neighboring ground states. For example, the ground state of any double-sided $SiH_{0.4}$ structure can be the combination of $Si_{12}H_4$ (x = 1/3, Fig. 10b) and Si_8H_4 (x = 1/2, Fig. 10c) structures. This kind of phase separation is also experimentally observed in chemically functionalized graphene.

In Fig. 11, the authors have plotted the calculated band gaps of these single-sided and double-sided SiH_x ground states as a function of H concentration. For some special indirect band-gap semiconductors, if the gap differences between the indirect band gap and minimum direct band gap are less than 0.1 eV, the authors call them quasi-direct gap semiconductors. Quasi-direct band-gap systems may behave as direct band-gap systems in most optoelectronic applications (Watanabe et al. 2004). As shown in Fig. 11, the authors find that there are large oscillations for the band gaps of single-sided SiH_x as a function of H concentration, which is significantly different from that of double-sided SiH_x. Single-sided $Si_{28}H_6$ and $Si_{24}H_6$ have direct (or quasi-direct) band gaps of 1.01 and 1.30 eV, respectively, and double-sided $Si_{16}H_4$ and $Si_{12}H_4$ structures have direct (or quasi-direct) band gaps of 1.52 and 1.46 eV, respectively. Clearly, these four SiH_x compounds have suitable band gaps for solar-absorbing applications. Interestingly, it is also found that double-sided Si_8H_4, $Si_{12}H_8$, $Si_{16}H_{12}$, and $Si_{12}H_{10}$ have direct (quasi-direct) band gaps of 1.85, 1.99, 2.43, and 2.85 eV, respectively, which lie within red, orange, green, and blue colors, as shown in Fig. 11b. Because Si_8H_4, $Si_{16}H_{12}$, and $Si_{12}H_{10}$ can realize the red, green, and blue colors, in a LED application, it may be possible to mix them in order to realize white color. Today, III-nitride materials, e.g., InGaN alloys, have become the front-runner for achieving phosphor-free white LEDs via the combination of RGB emitters because the alloys with different compositions can, in principle, produce blue, green, and red LEDs. However, it is known that it is exceedingly difficult to obtain red InGaN LEDs, and the efficiency of green
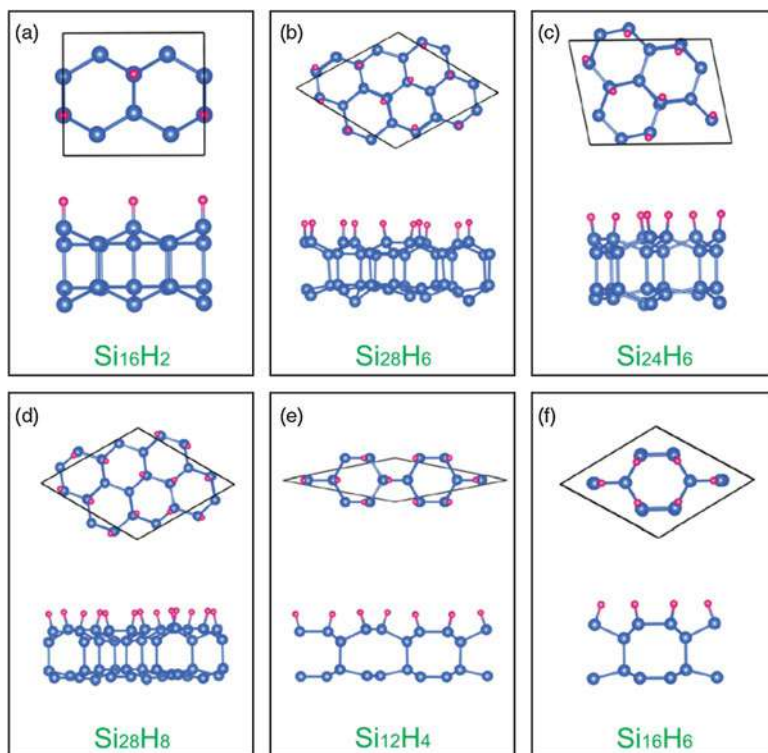
Fig. 9 The top and side views of six intermediate ground-state structures for single-sided (**a**) $Si_{16}H_2$, (**b**) $Si_{28}H_6$, (**c**) $Si_{24}H_6$, (**d**) $Si_{28}H_8$, (**e**) $Si_{12}H_4$, and (**f**) $Si_{16}H_6$. (Adopted from Huang et al. 2014)

LEDs using InGaN is also relatively low. This is partially because there is a large lattice mismatch (>10%) between GaN and InN, which can induce significant phase separations and dislocations during the formation of InGaN alloys. Obviously, the well-ordered SiH_x-layered structures the authors discovered here could overcome the large lattice-mismatch issue in InGaN alloys for high-efficiency white lighting. Moreover, it is found that layered materials have unique advantages over the conventional bulk materials for lighting because of the very strong optical transition rates at their band edges (Watanabe et al. 2004; Huang et al. 2012a).

4 Doping of 2D Materials

4.1 Enhanced Dopant and Carrier Densities in Graphene via Substrate

Controlling the type and density of charge carriers by doping is at the heart of graphene-like 2D electronics. However, the development of reliable chemical
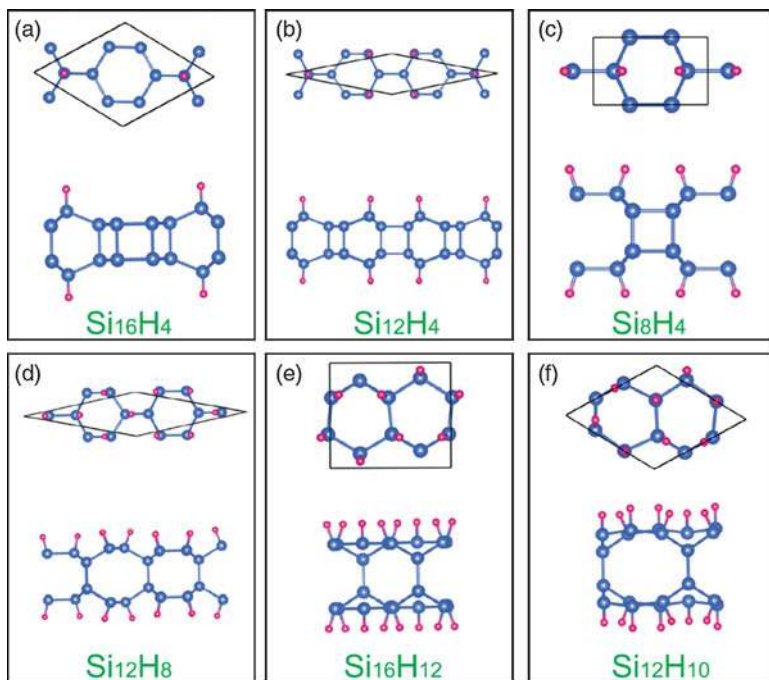
Fig. 10 The top and side views of six intermediate ground-state structures for double-sided (**a**) $Si_{16}H_4$, (**b**) $Si_{12}H_4$, (**c**) Si_8H_4, (**d**) $Si_{12}H_8$, (**e**) $Si_{16}H_{12}$, and (**f**) $Si_{12}H_{10}$. (Adopted from Huang et al. 2014)
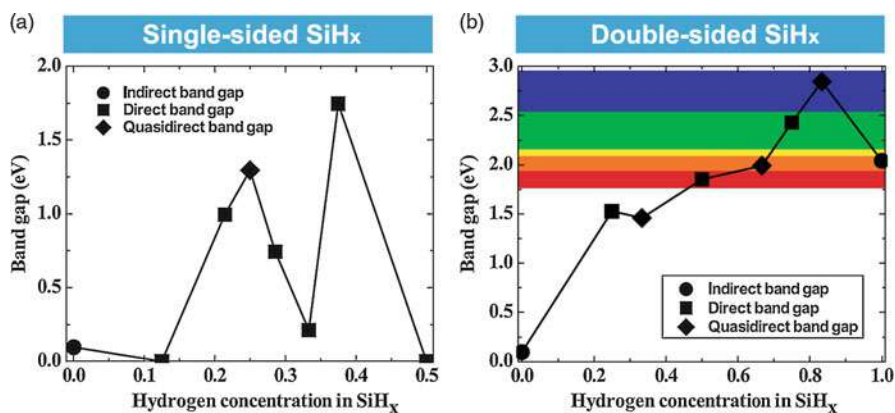
Fig. 11 The calculated band gaps of (**a**) single-sided SiH_x and (**b**) double-sided SiH_x ground states as a function of hydrogen concentration. In panel (**b**), a sketch of light spectrum is overlaid to aid visualization of the size of the band gap. (Adopted from Huang et al. 2014)

doping methods in graphene is still a real challenge. Direct doping of graphene with substitutional dopants is rather difficult, because the strong covalent C-C bonds make graphene one of the strongest materials in the world (Geim and Novoselov 2007). Until recently, doping of graphene has mainly been achieved through reactive molecular or atomic adsorbates in experiments, which is difficult to implement under current device technology. Moreover, direct doping usually suppresses the high carrier mobility of graphene evidently (Peres 2010). Comparing to mechanical exfoliation of graphene from graphite, large area epitaxial growth of graphene layers on a silicon carbide (SiC) surface shows huge application potentials and has apparent technological advantages over the mechanical exfoliation method. Interestingly, the authors find that substitutional doping could be strongly enhanced in epitaxial graphene grown on SiC substrate compared to that of freestanding graphene (Huang et al. 2011a). The formation energies of dopants can reduce by as much as ∼8 eV. The type and density of carriers in epitaxial graphene layer could be effectively manipulated by suitable dopants and surface passivation. Contrasting to direct doping of freestanding graphene, the carriers of epitaxial graphene are weakly scattered by dopants due to the spatial separation between scattering centers (dopants) and the conducting channel (epitaxial graphene).

The carriers in graphene could be manipulated by p-/n-type dopants, like in conventional semiconductors. Al, B, P, and N are considered as substitutional dopants, as they are widely used in C- and Si-based materials. The calculated formation energy differences for graphene-SiC system are shown in Fig. 12. Interestingly, Al prefers to substitute the Si atom of the SiC substrate or substitute the C atom of the interface buffer layer, and the formation energy of Al decreases dramatically by ∼8.26 eV compared to freestanding graphene case. B and P atoms prefer to substitute C atoms of the buffer layer, and the formation energies of dopants decrease largely by 2.36 and 5.27 eV, respectively, compared to freestanding graphene cases. Besides doping in the buffer layer, the formation energies of B and P at SiC surface are also reduced by 1.46 and 4.87 eV, respectively, compared to freestanding graphene cases. Contrary to the cases of Al, B, and P doping, N atom prefers to substitute the C atom of top SiC substrate with a reduction of formation energy by 3.52 eV, as shown in Fig. 12d. The similar phenomenon of substrate-enhanced doping is also found on the SiC (0001) surface, with the E_f of dopants decreased by as much as 10.8 eV.

These quite different doping behaviors between different dopants and different layers in the graphene-SiC system could be understood by considering the structural strain in the various epitaxial graphene-SiC layers, the different atomic covalent radiuses, and the electronegativity of dopants. Since the interface buffer layer is under large tensile strain, Al, B, and P atoms could be easier to substitute C atoms in this layer due to the larger covalent radiuses of Al, B, and P than sp^2 C. The rippled (deformed) structure reduces the doping difficulty in the buffer layer to some extent. Furthermore, because the electronegativity of C is larger than that of Si in SiC, Si behaves as positively charged cation, and C acts as negatively charged anion. Al atom has lower electronegativity and a little larger covalent radius than Si, so it would prefer to substitute Si atom in the SiC matrix to increase the negative
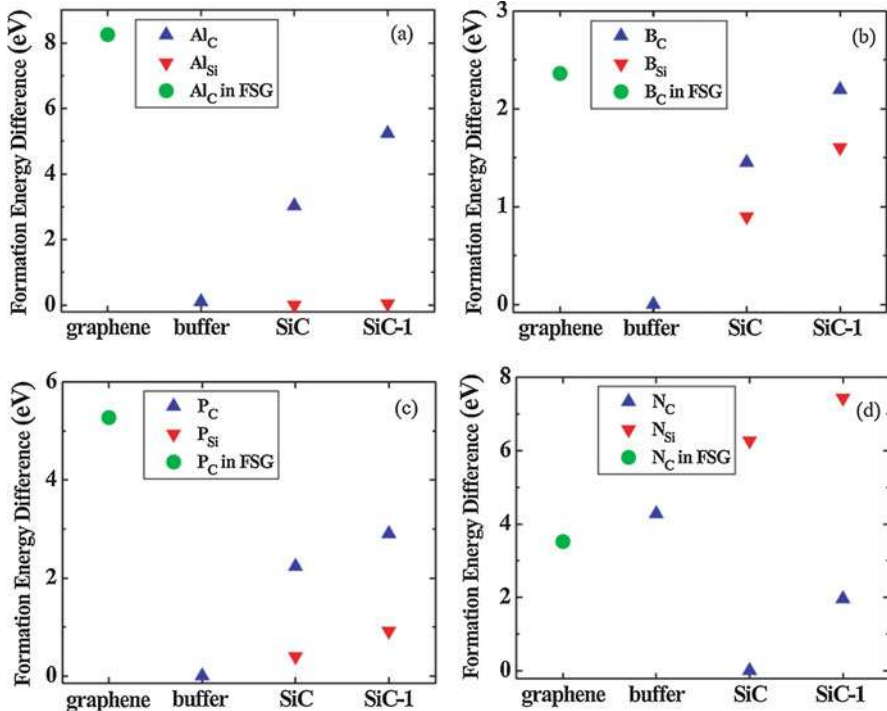
Fig. 12 The calculated formation energy differences of dopants (**a**) Al, (**b**) B, (**c**) P, and (**d**) N in a graphene-SiC system as a function of different layers. The formation energies of dopants in freestanding graphene (FSG) are also plotted in these figures for comparison. (Adopted from Huang et al. 2011a)

Coulomb interaction between Al and C. Differing from Al, B, and P, the structural tensile strain increases the difficulty of N doping in the graphene layer because N has a smaller covalent radius than C. The N atom has even higher electronegativity than C, so it would prefer to substitute the C atom in the SiC matrix to increase the negative Coulomb interaction between Si and N, which is in agreement with the calculated trend of formation energy in Fig. 12d. The above results and analysis show that overcoming the doping difficulty with significant enhanced dopant and carrier densities in graphene could be generally achieved through substrates, and it can also be applied to other layered or thin-film systems.

4.2   Magnetic Properties of Transition-Metal Doping in Graphene

Control of magnetism is the key issue of spintronics. 2D materials could have some advantages relative to conventional 3D materials used for controllable magnetism. In this section the authors demonstrate that strain and electric field can be applied to

control the magnetic moments of 2D materials effectively, which might open a new door for spintronics.

Transition-metal (TM) doping in graphene can induce magnetism, and strain can be further applied to modulate the magnetic properties of TM-graphene system (Huang et al. 2011b). The authors take the case of a single Mn atom adsorption on graphene as a primary example to demonstrate the strain effect on the TM-graphene system. The calculated strain-dependent magnetic moments are shown in Fig. 13a. A biaxial strain η is applied to the graphene system. A positive (negative) η means the TM-graphene system is under tension (compression). For pristine graphene (PG) without vacancies, the Mn atom prefers to stay at the hollow site (above the carbon hexagon center) of PG, and the total magnetic moment of the Mn atom is 5.7 μ_B, indicating that the Mn atom is in a high-spin state. The height and average bond length between Mn and its nearest C atoms are 2.09 Å and 2.54 Å, respectively, as shown in Fig. 13b, c. The binding energy of Mn on PG is about 0.45 eV, as shown in Fig. 13d. When a biaxial tensile strain η is applied for Mn-PG, interestingly, the magnetic ground state changes from high spin to low spin ($\sim$3.1 μ_B) abruptly at the critical strain of $\eta = 0.05$, as shown in Fig. 13a. Accompanying the magnetic transition, the height and average bond length between the Mn atom and its nearest C atoms decrease sharply to 1.50 and 2.16 Å, respectively, as shown in Fig. 13b, c. The binding energy of the Mn atom increases to 1.15 eV when $\eta = 0.05$. In contrast to the sharp transitions of spin states and local adsorption structures, the binding energy increases continuously as a function of η, as shown in Fig. 13d.

TM atoms on PG have low migration barriers based on DFT-GGA calculations so that it may be mobile at room temperature (Krasheninnikov et al. 2009). However, our climbing-image nudged-elastic band calculations (Cai et al. 2008) based on DFT-GGA calculations show that the migration barrier for the Mn atom on PG increases from 0.36 eV to 0.89 eV under a tension of $\eta = 0.10$, giving rise to a decrease of more than eight orders of magnitude in the diffusion coefficient at room temperature, which implies that the migration can be suppressed significantly. It also indicates that an adequate variation of strain can be used to control the patterning of TM atoms on graphene. Furthermore, it is known that the mobility of the TM atom becomes markedly reduced in the presence of vacancies (Cretu et al. 2010). For instance, the migration barriers for TM atoms on vacancies are larger than 3 eV (Krasheninnikov et al. 2009).

Similar to the case of Mn on PG, a similar magnetic transition between high-spin and low-spin states is found for Mn atoms adsorbed near the single-vacancy (SV) sites in graphene (Mn-SV). Unlike the Mn-PG case, the unstrained Mn-SV is in a low-spin state with the magnetic moment of $\sim$3 μ_B, as shown in Fig. 13a. The local adsorption structures of the Mn atoms are shown in Fig. 13e. The binding energy of the Mn atom is 6.4 eV, which is much larger than the Mn-PG case, indicating that the Mn-SV system is quite stable. Interestingly, a compression of $\eta = -0.04$ can convert the magnetic ground state from low spin ($\sim$3 μ_B) to high spin ($\sim$5 μ_B) for Mn-SV. Under compression, the graphene layer produces a spontaneous rippling with the formation of a C-C dimer bond near SV, as shown in Fig. 13e. As results, the height and the average bond length between Mn and its nearest C atoms increase signifi-
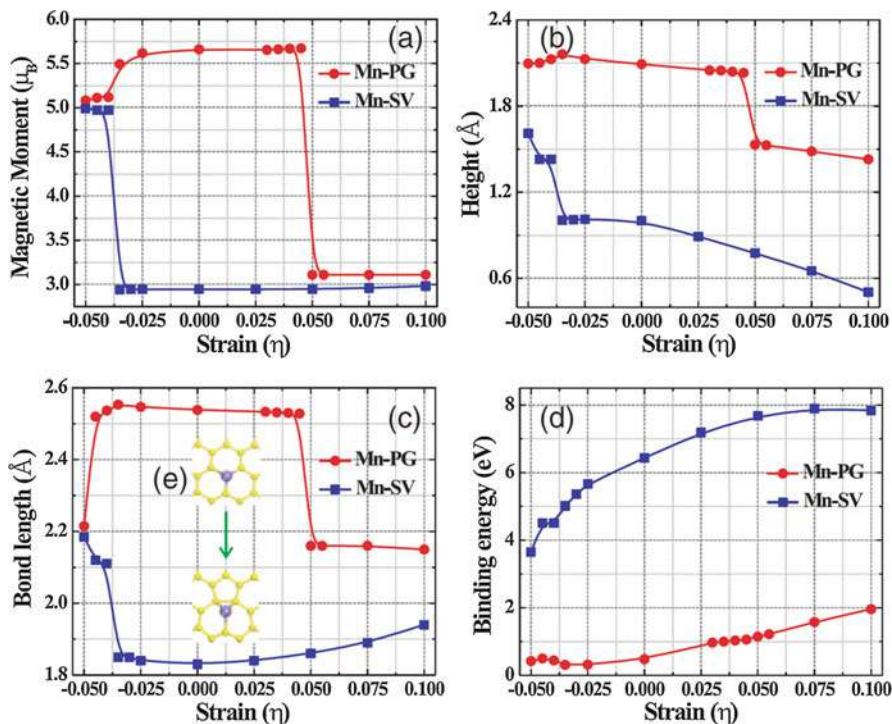
Fig. 13 (**a**) Magnetic moments of Mn atom adsorption on pristine graphene (PG) and defective graphene with a single vacancy (SV) as a function of strain. (**b**) Heights between Mn atoms and the nearest C atoms as a function of strain for Mn-PG and Mn-SV systems. (**c**) Average bond lengths between Mn atoms and the nearest C atoms as a function of strain for Mn-PG and Mn-SV systems. (**d**) Binding energies of Mn atoms on graphene as a function of strain for Mn-PG and Mn-SV systems. (**e**) Local structural transition for Mn-SV system from $\eta = 0$ (up structure) to $\eta = -0.04$ (down structure). (Adopted from Huang et al. 2011b)

cantly under compression, which weakens the binding. In general, the authors find that tensile (compressive) strain can enhance (reduce) the binding strength between TM atom and graphene, which is consistent with experimental observation (Geim and Novoselov 2007). It is interesting to observe that the electronic and magnetic structures of the Mn atom depend strongly on the strain of graphene layer, thereby leading to the strain-induced sudden change of the magnetic ground state. The strain affects the local geometry of carbon atoms around the Mn atom in two ways. It changes the height of Mn atom above the graphene layer and generates a ripple structure under the compressive strain. The change of the local environment, which determines the distance and symmetry of C p orbitals with respect to that of Mn d orbitals, can lead to a dramatic change in the magnetic d orbitals due to the strong spin-dependent hybridization between Mn d and C π orbital states.

The strain control of the magnetism generally exists for almost all the TM atoms from Sc to Zn on graphene. The range of magnetic moments of TM-graphene
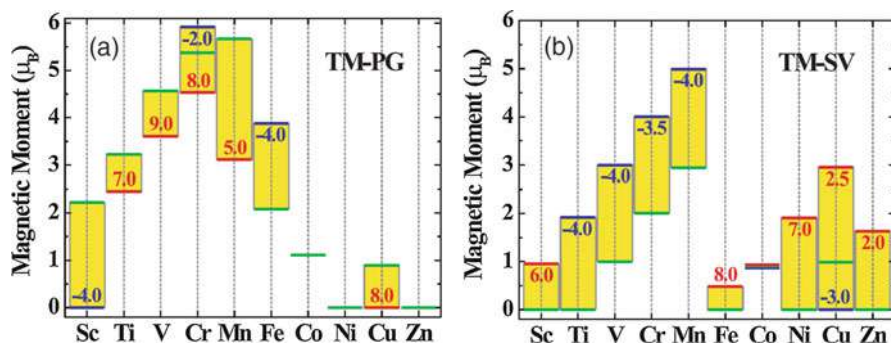
Fig. 14 The range of magnetic moments of TM atoms adsorption on graphene (**a**) without vacancy and (**b**) with SV under strain. The intrinsic magnetic moments of TM graphene without strain are shown as green lines, while the maximum and minimum magnetic moments of TM graphene caused by compressive (or tensile) strain (in unit of %) are shown by blue (or red) line. (Adopted from Huang et al. 2011b)

systems under realistic strain $-0.05 < \eta < 0.10$ is shown in Fig. 14. The transition of magnetic states for TM-PG systems under strain could be either continuous (e.g., V, Ti, and Cr) or abrupt (e.g., Sc, Mn, Fe, and Cu), as shown in Fig. 13a, which is mainly due to different strain-dependent ligand splitting and hybridization. Interestingly, strain can also control the adsorption sites of TM atom, for example, the adsorption site of Cu (Sc and Fe) on graphene converts from C-C bridge (hollow) site to hollow (C-C bridge) site at $\eta > 0.08$ ($\eta < -0.04$). The spin transition is even more remarkable in the TM-SV cases, as shown in Fig. 3b. For example, compressive strains of -0.04, -0.04, and -0.035 could transfer Ti-, V-, and Cr-SVs from low-spin to high-spin states, respectively, similar to Mn-SV. Tensile strains of 0.025, 0.06, and 0.07 could transfer Cu-, Sc-, and Ni-SVs from low-spin to high-spin states, respectively. It should be noticed that at least 66% (50%) of total magnetic moments of the Cu-SV (Ni-SV) system originates from the nearest C atoms around vacancy.

4.3 Magnetic Properties of Transition-Metal Doping in Boron Nitride

Because of the lack of band gap in graphene, as demonstrated in Fig. 13, the TM d orbitals strongly couple with the conduction or valence bands of graphene, and further control of the charge and magnetic states of TM/graphene system is quite difficult, but the reliable control of the charge and magnetic states of a magnetic material is a key step for spintronics and quantum information devices. Naturally, one may expect to solve this difficulty by replacing metallic graphene layer with some semiconducting layers, e.g., BN. Owing to the large chemical difference between cation B and anion N atoms, these ionic layer structures display large

heteropolar band gaps. Because the energies of TM *3d* orbitals can exist inside the band gap, i.e., not in the conduction or valence band as in the case of TM/graphene, the authors expect that these in-gap TM *3d* orbitals may be easier to be controlled by various approaches and the BN layer systems might be considered as preferential magnetic materials for low-dimensional spintronics. As for the case of graphene, the authors now focus on BN with intrinsic defects.

Depending on the position and symmetry of in-gap levels induced by vacancies, it is expected that they can interact with the adsorbed TM atom to achieve the controllable charge and magnetic states of TM/BN. For instance, since the energies of TM *3d* orbitals are roughly located at the middle of the gap of BN, they could behave as TM^{3+} ion after coupling to triple-acceptor V_B or TM^{2+} ion after interacting with double-acceptor V_{BN}. Since the in-gap levels induced by V_N are comparable to that of *3d* orbitals of TM in energy, instead of charge transfer, hybridization effect can play a critical role in determining the magnetic states of TM-VN.

When a TM atom is adsorbed at these vacancies (Fig. 15a), the binding energy of a TM atom on V_B is significantly larger than that of a TM atom on V_{BN} and V_N, which could be understood by the larger charge transfer effect in TM-V_B. Among these 3d TM atoms studied in Huang et al. (2012b), Sc is the one with the largest binding energy on V_B, as shown in Fig. 15b, consisting with the fact that the 3d orbitals of Sc are highest in energy than other 3d TM atoms and the large energy gain corresponds to charge transfer from Sc to BN. For TM atoms adsorption on V_{BN}, two electrons of TM atom transfer to V_{BN}, giving rise to TM^{2+} charge state. The alignment of the remaining *3d* levels exactly follows the Hund's rule, as shown in Fig. 15d. For TM atoms adsorption on V_B, TM atoms exactly exhibit TM^{3+} state after three electrons transfer to V_B. The trend of magnetic moment of TM-VN can be understood by the competing between crystal field splitting and exchange energy splitting (Huang et al. 2012b).

The understanding above suggests that the magnetic properties of TM/BN can be tuned if the authors can effectively adjust the relative energy positions of TM and/or defective BN. Especially, applying an external electric field E_{ext} can shift the relative energy-level positions because TM and BN are not in the same plane and because TM and BN can have opposite charge states when BN is defective. Taking Mn-V_B as an example (Fig. 16), Mn-V_B exhibits a high-spin configuration without an E_{ext} (Fig. 15d). A positive E_{ext} exerts a force on the system, pushing the Mn closer to the BN because Mn is positively charged and BN is negatively charged. This increases the crystal field splitting. Remarkably, when E_{ext} is increased to 1.0 V/nm, the height of Mn atom on V_B is decreased suddenly from 1.6 Å to 1.5 Å. The large increase of the crystal field splitting causes a spin crossover, i.e., charges are transferred from the spin-up channel to the spin-down channel (Fig. 16c), and the spin configuration converts from high spin to low spin, as shown in Fig. 4b. This low-spin state is found to be a metastable one with a total energy of 360 meV higher than that of the high-spin state in the absence of E_{ext}. Therefore, a negative critical E_{ext} is required to convert the low-spin state to high-spin one to achieve the spin crossover. The similar physical mechanism results in the spin configuration of Co-V_{BN} transfer
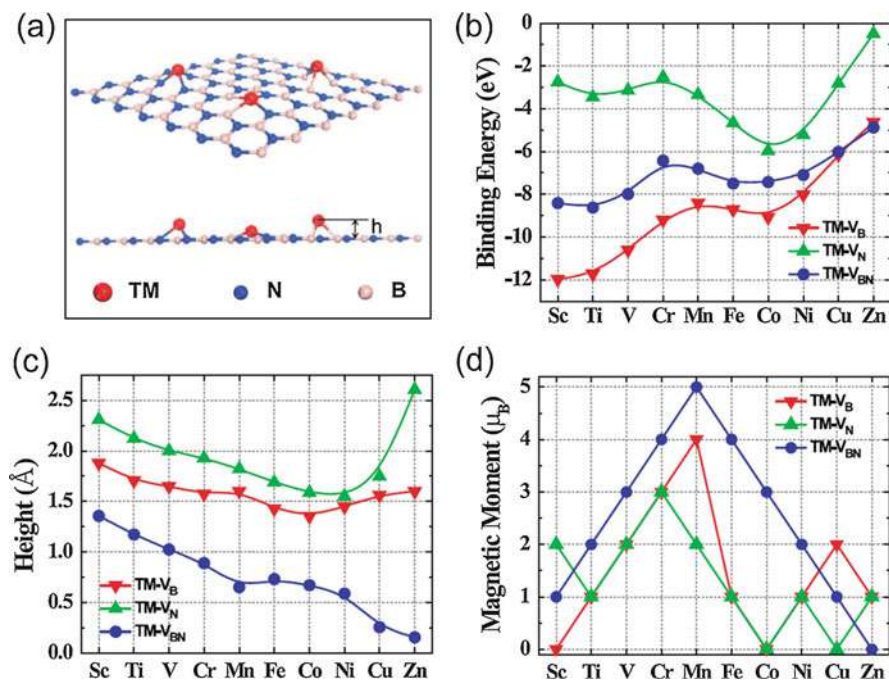
Fig. 15 (**a**) The schematic structure of TM atoms adsorbed on V_B, V_N, and V_{BN} in BN. (**b**) The binding energies of TM atoms adsorption on V_B, V_N, and V_{BN} in BN. (**c**) The height of the TM atoms above the BN surface when TM atoms adsorb on V_B, V_N, and V_{BN}. (**d**) The magnetic moments of BN with TM atoms adsorption on V_B, V_N, and V_{BN}. (Adopted from Huang et al. 2012b)

from high spin to low spin when E_{ext} is about 3.0 V/nm (Fig. 16b). The fact that larger E_{ext} is required for the Co-V_{BN} system than the Mn-V_B system is consistent with the fact that the Co-V_{BN} is less ionized, i.e., less charge transfer between TM and BN, than Mn-VB. It is worth to notice that the small critical E_{ext} in these systems indicates that it is possible to achieve this spin-crossover effect in experiments (Zhang et al. 2009b). The calculations show that this spin-crossover effect exists generally in TM/III–V and TM/II–VI layer structures, which strongly suggests that these TM/ionic layer systems could be used for electromagnetic applications such as E_{ext}-driven nonvolatile memory or energy devices.

5 Alloying of 2D Materials

Besides chemical functionalization and doping, alloying could be another choice to modulate the electronic properties of 2D materials for various energy applications (Xiang et al. 2012; Huang et al. 2015a, b; Zhu et al. 2015; Yang et al. 2017a, b; Luo et al. 2011).
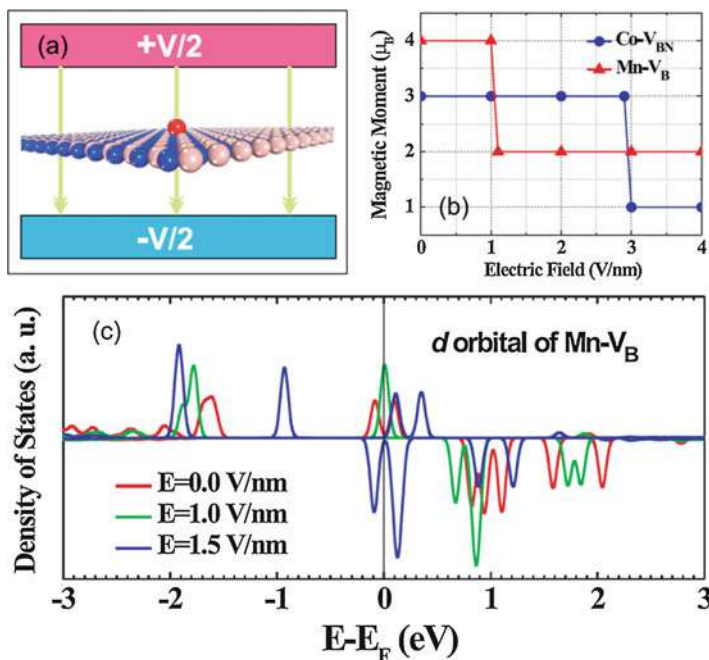
Fig. 16 (**a**) The model of a TM-BN system under an E_{ext}. An E_{ext} is oriented normally to BN surface and is assumed to be positive when it is directed downward. (**b**) The magnetic moments of Co-V_{BN} and Mn-V_B as a function of E_{ext}. (**c**) The energy shift of d orbitals for Mn-V_B as a function of E_{ext}. The Fermi level is set to zero. (Adopted from Huang et al. 2012b)

5.1   2D C-N Alloy

First, the authors focus on the 2D C-N alloy systems. Based on the CE method, Xiang et al. identified two intermediate ground states, i.e., C_3N and $C_{12}N$, as shown in Fig. 17. To further confirm the CE results, a global minimum structural optimization method based on particle swarm optimization (PSO) is adopted to confirm the CE-proposed structures. In 2D PSO simulations (Yang et al. 2017a), the graphene lattice structure is not assumed, which is different from the CE method. Instead, random structures (both atomic positions and cell parameters) are generated to initialize the simulations. The initial structures are subsequently relaxed before the PSO operations are performed. The PSO simulations show that the most stable 2D nitrogen-carbon alloy structure with 25% N concentration is indeed the structure (a p6mm plane group of 2D systems) shown in Fig. 17a. The PSO simulations for the 2D nitrogen-carbon alloy structure with 7.7% nitrogen concentration do not give structures more stable than the $C_{12}N$ shown in Fig. 17b. This result supports the stability of the proposed $C_{12}N$ structure. The authors note that the formation energies of C_3N and $C_{12}N$ are positive if graphene and the N_2 molecule are taken as the references. However, many experiments have shown that N atoms can be
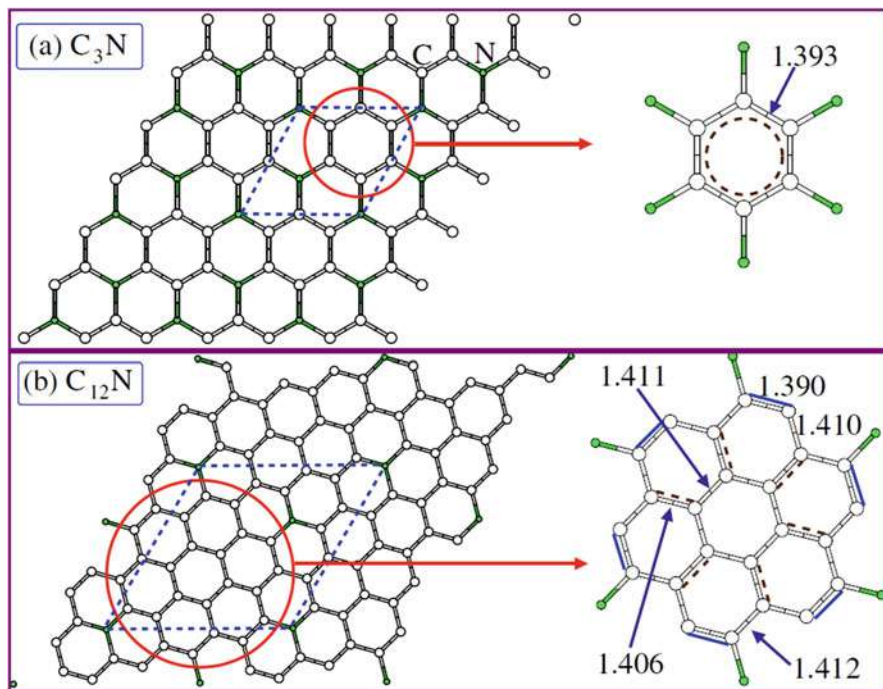
Fig. 17 (**a**) The ordered N-doped graphene superstructure for C_3N. (**b**) The ordered N-doped graphene superstructure for $C_{12}N$. Unit cells of the superstructures are enclosed by dashed lines. Each right panel displays the enlarged view of the structure that is enclosed by a red circle in the left panels. The numbers denote the bond lengths (in Å). The Clar formulas are specified by the circle, solid, and dashed lines. (Adopted from Xiang et al. 2012)

incorporated into graphene. What the authors have demonstrated is that C_3N and $C_{12}N$ are the most stable structures once N atoms are incorporated into graphene.

The band structures for the C_3N and $C_{12}N$ superstructures from the LDA calculations are shown in Fig. 18. Interestingly, both superstructures are found to be semiconductors: C_3N has an indirect band gap (about 0.26 eV from the LDA calculation). $C_{12}N$ has instead a direct band gap (about 0.62 eV from the LDA calculation) at the M point. Because the p_z orbital of a nitrogen atom will be almost fully occupied by the two π electrons, the mechanism for the band-gap opening in graphene induced by nitrogen doping can be understood by examining only the bonding situation of carbon π electrons.

The Clar's theory of the aromatic sextet (Wassmann et al. 2010) has been shown to be a simple and powerful tool to predict the stability and the electronic/magnetic structure of graphene-related systems. According to Clar's rule, for a given system, the representation with a maximum number of Clar sextets, called the Clar formula, is the most representative one. Here, the authors show the Clar formulas for C_3N and $C_{12}N$ in the right panels of Figs. 17a

Fig. 18 (**a**) The LDA band
structures for (**a**) C_3N and (**b**)
$C_{12}N$. The horizontal green
dashed lines denote the VBM
levels. The contributions from
N p_z orbitals to the wave
functions are indicated by the
radii of the solid spheres.
(Adopted from Xiang et al.
2012)
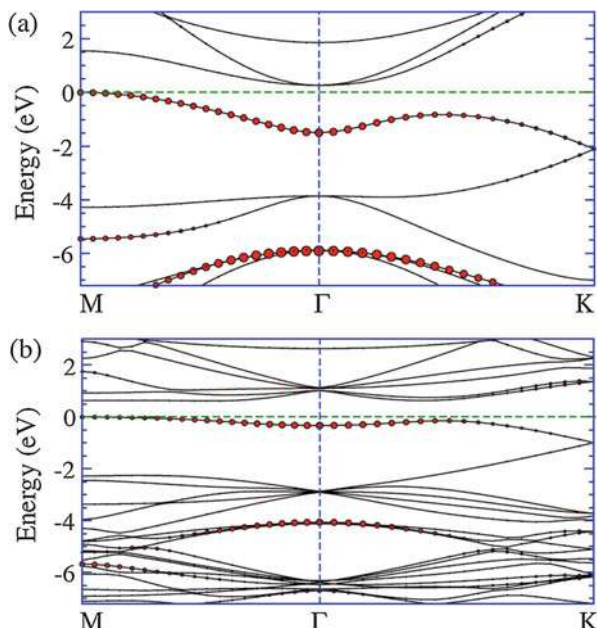
and 18b, respectively. In C_3N, each six-membered ring forms a Clar sextet.
C_3N has one unique Clar formula and thus belongs to the class 1CF of the
(pseudo)-all-benzenoid structure. In contrast, $C_{12}N$ has no Clar sextets and thus
is non-benzenoid. Because each nitrogen atom forms three single bonds with the
neighboring carbon atoms, the nitrogen atoms are excluded when the authors
construct the Clar formula. The Clar formulas shown in Fig. 17 are in good
agreement with the C-C bond lengths of relaxed structures from the DFT cal-
culation. All the C-C bond lengths in C_3N are the same: 1.393 Å. For $C_{12}N$,
some C-C bonds (marked by an additional solid line) have shorter bond lengths,
1.388 Å, which can be considered a normal C-C double bond; and there are weak
C-C double bonds (marked by an additional dashed line) with bond lengths of
1.406 Å; all the other C-C bonds are single bonds. Therefore, the authors can
understand that there is a large band gap between the occupied carbon π states
and the unoccupied carbon π states in both C_3N and $C_{12}N$. Our analysis for the
wave-function characters indicates that the highest occupied band near π is mainly
contributed by nitrogen p_z orbitals, and there are some contributions from carbon p_z
orbitals when the Brillouin-zone boundary is approached. The lowest unoccupied
band is mostly contributed by the carbon p_z orbitals. The gap between the highest
occupied carbon π state and the CBM is larger than 1.5 eV (in LDA) in both C_3N
and $C_{12}N$, which is in accord with Clar's theory. Therefore, the authors predict that
collective nitrogen doping can induce exotic semiconducting behavior in graphene
instead of *n*-type doping behavior. The real band gap should be larger because the
LDA functional underestimates the band gap. It is well known that LDA seriously

underestimates the band gap of semiconductors; thus, the authors calculate the electronic structures of C_3N and $C_{12}N$ by employing the screened HSE06 hybrid functional, which has been shown to give a good band gap for many semiconductors. Our HSE06 calculations show that C_3N and $C_{12}N$ have the band gaps of 0.96 eV and 0.98 eV, respectively. The authors note that the band gaps of the nitrogen-doped graphene are close to the band gap (about 1.12 eV) in silicon.

5.2   2D Si-P Alloy

As 2D Si and P have been achieved in the recent experiments, an interesting question is whether stable 2D silicon phosphide (Si_xP_y) monolayers with different stoichiometries can exist in the Si-P phase diagram. If so, what kinds of stable stoichiometries can they have, and do they have more attractive electronic properties than that of 2D Si or P? The answers to these questions are not only scientifically important to extend our current knowledge of 2D materials, but also can provide an approach for the discovery of new 2D functional materials beyond the existing materials. By combining global structural search (PSO) and first-principles calculations, the authors have found several stable or metastable semiconducting Si_xP_y monolayers, which can only be formed at the stoichiometries of $y/x \geq 1$, as shown in Fig. 19.

When $y/x \geq 1$, a large number of Si_xP_y monolayers with negative ΔH emerge, and most of these satisfy the classical electron counting rule, i.e., Si atoms are fourfold coordinated, while P atoms are threefold coordinated to realize full sp^3 hybridization. For Si_1P_1, the lowest-ΔH phase has a GaSe-type (P-6m2) structure with $\Delta H = -0.159$ eV/atom, as shown in Fig. 20a. The P-3ml and P21/m phases (Fig. 20b, c) have a slightly higher ΔH value than that of the P-6m2 phase. When the P composition is further increased to the stoichiometry of Si_1P_2, seven Si_1P_2 with negative H values appear in the phase diagram. The lowest-ΔH phase has Pm symmetry. As shown in Fig. 20d, the Pm Si_1P_2 structure can be considered as the stacking of one zigzag P chain on the h-BN-like SiP monolayer in each unit cell. The second lowest-ΔH phase is P-42$_1$m Si_1P_2, which has a slightly higher ΔH than that of the Pm phase. Interestingly, the Si_2P_3 pentagon is the basic building block to form the P-42$_1$m Si_1P_2 structure, as shown in Fig. 20e. For the stoichiometry of Si_1P_3, the authors find two structures with negative ΔH, and the lowest-ΔH one has C2/m symmetry, as shown in Fig. 20f.

In terms of the calculated ΔH values of all the 2D Si_xP_y, the authors can obtain the ground-state convex hull of 2D Si_xP_y, as shown in Fig. 19. Obviously, P-6m2 Si_1P_1 and Pm Si_1P_2 are located on the convex hull, indicating that they are stable against disproportioning into other neighboring compounds. It is important to compare the thermal stabilities of the predicted monolayer Si_xP_y to the existing bulk Si_xP_y phases. For the stoichiometry of Si_1P_1, the bulk Cmc21 phase has a negligibly lower ΔH value than the predicted monolayer P-6m2 phase, while the bulk F-43m phase has a significantly positive ΔH value. For the stoichiometry of Si_1P_2, the predicted monolayer Pm phase also has a quite similar ΔH value as the bulk Pbam and Pa-3 phases.
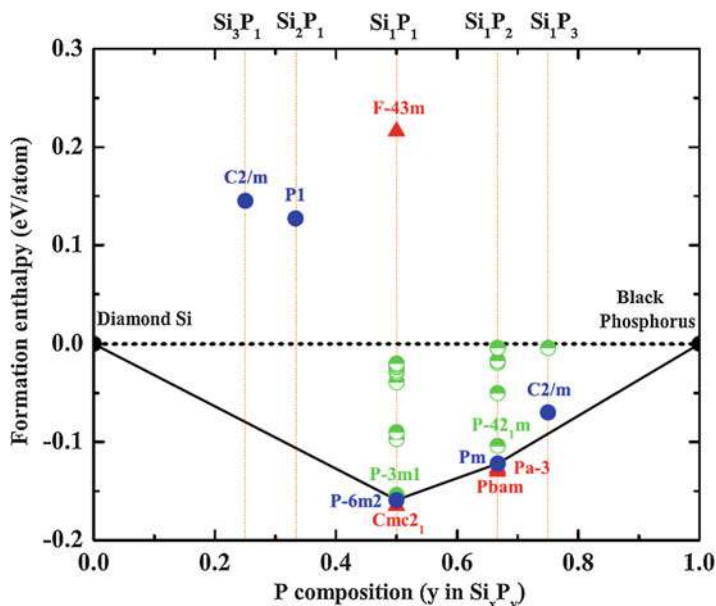
Fig. 19 Thermodynamic stability of 2D Si_xP_y indicated by the calculated formation enthalpies (ΔH) of various 2D Si_xP_y compounds with respect to elemental decomposition into diamond silicon and black phosphorus. The solid blue circles represent the predicted lowest-ΔH Si_xP_y at each P composition. The half-solid green circles represent other predicted Si_xP_y compounds with negative ΔH values. The ground-state convex hull of 2D Si-P compounds is denoted by the solid lines. For comparison, the calculated ΔH values of four previously known bulk Si_xP_y compounds, i.e., F-43m Si_1P_1, Cmc2$_1$ Si_1P_1, Pbam Si_1P_2, and Pa-3 Si_1P_2, are also plotted as red triangles. (Adopted from Huang et al. 2015a)

5.3   2D Si-S Alloy

One promising approach for the discovery of new materials is "atomic transmute" in which one type of elements is transmutated into its neighboring elements in the periodic table but the total number of valence electrons is kept unchanged. In 2D materials, it is well known that by transmutating two C atoms in group IV graphene into one B atom and one N atom, monolayer groups III–V material h-BN can be obtained with wide band gap and new functionalities. Similarly, it is possible to replace two P atoms in phosphorene or in monolayer blue phosphorus by one Si atom and one S atom and predicted new 2D materials α-SiS and β-SiS with interesting properties (Zhu et al. 2015). However, it is not clear whether these are the ground-state structures for these materials.

To identify the ground-state structures of SiS, Yang et al. have performed the global structural search method, which can effectively catch the global ground-state structure of a compound (Yang et al. 2017a). The lowest-energy structure of Si_2S_2 (Pmma symmetry) obtained from this search is shown in Fig. 21a. In Pmma-SiS, Si atoms form zigzag chains along b directions (see bc-plane projections in Fig. 21a),
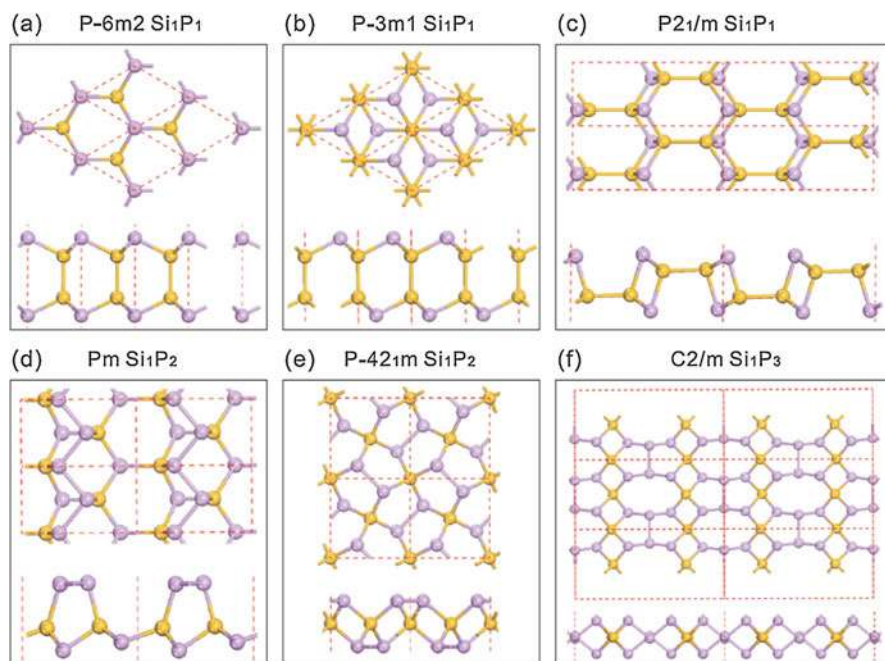
Fig. 20 Structures of 2D Si_xP_y. The top view (upper) and side view (lower) of (**a**) P-6m2 Si_1P_1, (**b**) P-3m1 Si_1P_1, (**c**) P2$_1$/m Si_1P_1, (**d**) Pm Si_1P_2, (**e**) P-42$_1$m Si_1P_2, and (**f**) C2/m Si_1P_3. Violet and yellow spheres are P and Si atoms, respectively. The unit cells are marked by the red dashed lines. (Adopted from Huang et al. 2015a)

which are sandwiched and also connected by two S layers. Each Si atom is bonded to two other Si atoms and two S atoms, while each S atom is bonded to two Si atoms. As a result, the system obeys the "octet rule," which means that both S and Si atoms have eight valence electrons to fully occupy the outmost *sp* shells and each bond (including S dangling bonds) contains two electrons; thus this structure should be chemically stable. However, the calculated phonon spectrum for this Si_2S_2 cell shows that there is one small imaginary frequency around X point, which implies that periodicity of atomic arrangements along a direction should be doubled. Indeed, the authors found by structural search that Si_4S_4 (Pma2 symmetry) is dynamically more stable, as shown in Fig. 21c. In Pma2-SiS, both top and bottom S atom layers are alternatively slightly rotated around the a axis or titled toward positive and negative b direction. This long-range periodicity in S layers lowers the total energy slightly and removes the imaginary frequency from the phonon spectrum. Consequently, the structure in Fig. 21c is both chemically and dynamically stable. Besides Pmma-SiS and Pma2-SiS, it is also found that another interesting structure in Si_2S_2 unit cell, as shown in Fig. 21b, can exist. In this structure, all the Si atoms are in ab plane and form stretched honeycomb lattices with top and bottom S layers inserted into Si-Si bonds. Because of the similarity between this structure and unbuckled silicene and graphene oxide, the authors name it as silicene sulfide,
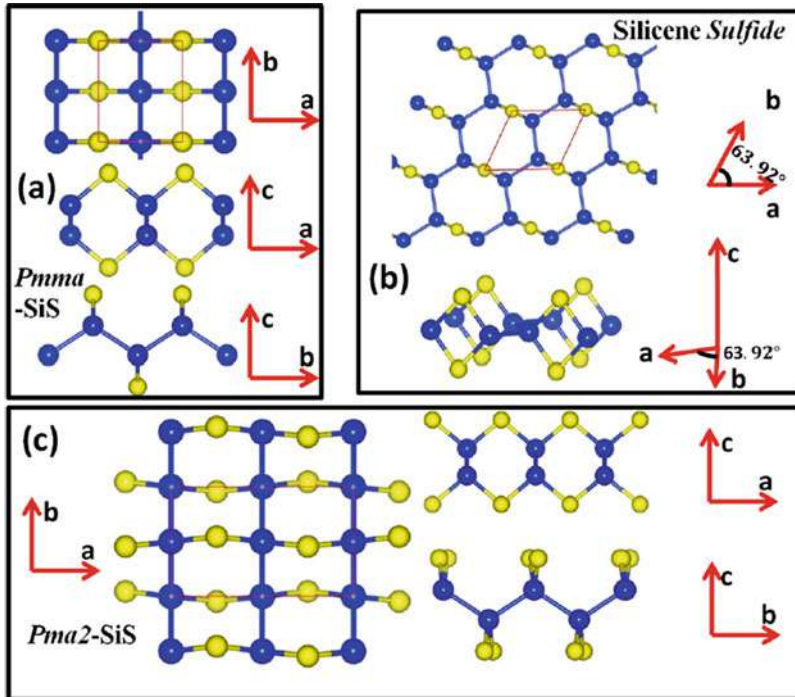
Fig. 21 Top and side views of structures for (**a**) Pmma-SiS, (**b**) silicene sulfide of Cmmm symmetry, and (**c**) Pma2-SiS. Blue is for Si and yellow is for S atoms. (Adopted from Yang et al. 2017a)

which has Cmmm symmetry. In this structure, both Si and S atoms also obey "octet rule," and thus it should be chemically stable.

The band structures and density of states of Pma2-SiS and silicene sulfide are shown in Fig. 22. Clearly, both of them are direct band-gap semiconductors with band gaps of 1.22 and 1.01 eV for Pma2-SiS and silicene sulfide, respectively, from the HSE06 calculations. The authors also performed G_0W_0 calculations for our structures and then obtained the optical band gaps of Pma2-SiS and silicene sulfide are 1.35 and 1.14 eV, respectively. Because their optical band gaps are close to the optimal values required for solar cell absorbers, they can be good candidate semiconductors for photovoltaic applications.

6   Conclusions

In conclusion, the authors have described the electronic properties of functionalized 2D materials that are useful for various applications, including solar applications, spintronic applications, and solid-state lighting. Our review clearly shows
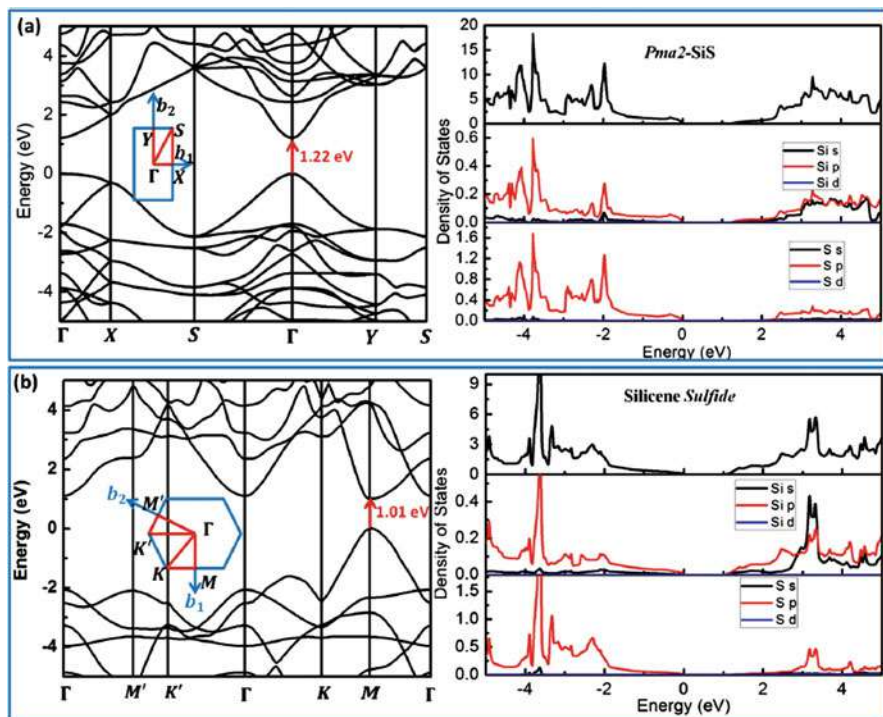
Fig. 22 Band structures and (partial) density of states for (**a**) Pma2-SiS and (**b**) silicene sulfide. (Adopted from Yang et al. 2017a)

that first-principles (density functional theory based) calculations are powerful approaches to understand the fundamental physical properties of 2D materials. For examples, our calculations and analysis have allowed us to propose new approaches to modify the optoelectronic properties of several 2D materials via chemical functionalization, to suggest a new way to enhance the doping efficiency in graphene via substrate doping and carrier injection, to show the possibility of using 2D materials in spintronic applications via magnetic atom doping, as well as to reveal that alloying can dramatically expand the exploitation of 2D materials in energy applications. Our review is intended not only to present current research activities on chemical functionalized 2D materials but also to generate new ideas in the field. In the future study of 2D materials, the authors expect that more attention will be paid to design new functionalized 2D materials by heterojunction engineering and discover novel 2D topological materials, such as quantum spin hall insulators. The authors believe that investigation with more advanced computational methods will further extend the applications of 2D systems in energy and spintronic domains.

References

Bergman L, McHale JL (eds) (2011) Handbook of luminescent semiconductor materials. Taylor Francis Group, Boca Raton

Botti S, Flores-Livas JA, Amsler M, Goedecker S, Marques MAL (2012) Low-energy silicon allotropes with strong absorption in the visible for photovoltaic applications. Phys Rev B 86:121204

Boukhvalov DW, Katsnelson MI (2008a) Chemical functionalization of graphene with defects. Nano Lett 8:4373

Boukhvalov DW, Katsnelson MI (2008b) Modeling of graphite oxide. J Am Chem Soc 130:10697

Boukhvalov DW, Katsnelson MI, Lichtenstein AI (2008) Hydrogen on graphene: Electronic structure, total energy, structural distortions and magnetism from first-principles calculations. Phys Rev B 77:035427

Brodie B (1859) On the atomic weight of graphite. Philos Trans R Soc Lond 149:249

Bundy FP (1989) Pressure-temperature phase diagram of elemental carbon. Physica A 156:169

Cahangirov S, Topsakal M, Akturk E, Sahin H, Ciraci S (2009) Two- and one-dimensional honeycomb structures of silicon and germanium. Phys Rev Lett 102:236804

Cai WW, Piner RD, Stadermann FJ, Park S, Shaibat MA, Ishii Y, Yang DX, Velamakanni A, An SJ, Stoller M, An JH, Chen DM, Ruoff RS (2008) Synthesis and solid-state NMR structural characterization of ^{13}C-labeled graphite oxide. Science 321:1815

Chen L, Liu CC, Feng B, He X, Cheng P, Ding Z, Meng S, Yao Y, Wu K (2012) Evidence for dirac fermions in a honeycomb lattice based on silicon. Phys Rev Lett 109:056804

Cretu O, Krasheninnikov AV, Rodriguez-Manzo JA, Sun L, Nieminen RM, Banhart F (2010) Migration and localization of metal atoms on strained graphene. Phys Rev Lett 105: 196102

Eda G, Fanchini G, Chhowalla M (2008) Large-area ultrathin films of reduced graphene oxide as a transparent and flexible electronic material. Nat Nanotechnol 3:270

Fagan SB, Baierle RJ, Mota R, da Silva AJR, Fazzio A (2000) Ab initio calculations for a hypothetical material: Silicon nanotubes. Phys Rev B 61:9994

Feng B, Ding Z, Meng S, Yao Y, He X, Cheng P, Chen L, Wu K (2012) Evidence of silicene in honeycomb structures of silicon on Ag(111). Nano Lett 12:3507

Ferreira LG, Wei SH, Zunger A (1989) First-principles calculation of alloy phase diagrams: The renormalized-interaction approach. Phys Rev B 40:3197

Fleurence A, Friedlein R, Ozaki T, Kawai H, Wang Y, Yamada-Takamura Y (2012) Experimental evidence for epitaxial silicene on diboride thin films. Phys Rev Lett 108:245501

Gao W, Alemany LB, Ci L, Ajayan PM (2009) New insights into the structure and reduction of graphite oxide. Nat Chem 1:403

Garza AJ, Scuseria GE (2016) Predicting band gaps with hybrid density functionals. J Phys Chem Lett 7:4165

Geim AK, Novoselov KS (2007) The rise of graphene. Nat Mater 6:183

Gilje S, Han S, Wang M, Wang KL, Kaner RB (2007) A chemical route to graphene for device applications. Nano Lett 7:3394

Guzman-Verri GG, Lew Yan Voon LC (2007) Electronic structure of silicon-based nanostructures. Phys Rev B 76:075131

Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened Coulomb potential. J Chem Phys 118:8207

Ho I, Stringfellow GB (1996) Solid phase immiscibility in GaInN. Appl Phys Lett 69:2701

Hod O, Barone V, Scuseria GE (2008) Half-metallic graphene nanodots: A comprehensive first-principles theoretical study. Phys Rev B 77:035411

Huang B, Xiang HJ, Wei S-H (2011a) Controlling doping in graphene through a SiC substrate: A first-principles study. Phys Rev B 83:161405

Huang B, Yu J, Wei S-H (2011b) Strain control of magnetism in graphene decorated by transition-metal atoms. Phys Rev B 84:075415

Huang B, Cao XK, Jiang HX, Lin JY, Wei S-H (2012a) Origin of the significantly enhanced optical transitions in layered boron nitride. Phys Rev B 86:155202

Huang B, Xiang HJ, Yu J, Wei S-H (2012b) Effective control of the charge and magnetic states of transition-metal atoms on single-layer boron nitride. Phys Rev Lett 108:206802

Huang B, Xiang HJ, Xu Q, Wei S-H (2013a) Overcoming the phase inhomogeneity in chemically functionalized graphene: The case of graphene oxides. Phys Rev Lett 110:085501

Huang B, Xiang HJ, Wei S-H (2013b) Chemical functionalization of silicene: Spontaneous structural transition and exotic electronic properties. Phys Rev Lett 111:145502

Huang B, Deng HX, Lee H, Yoon M, Sumpter BG, Liu F, Smith SC, Wei S-H (2014) Exceptional optoelectronic properties of hydrogenated bilayer silicene. Phys Rev X 4:021029

Huang B, Zhuang HL, Yoon M, Sumpter BG, Wei S-H (2015a) Highly stable two-dimensional silicon phosphides: Different stoichiometries and exotic electronic properties. Phys Rev B 91:121401

Huang B, Yoon M, Sumpter BG, Wei S-H, Liu F (2015b) Alloy engineering of defect properties in semiconductors: Suppression of deep levels in transition-metal dichalcogenides. Phys Rev Lett 115:126806

Krasheninnikov AV, Lehtinen PO, Foster AS, Pyykko P, Nieminen RM (2009) Embedding transition-metal atoms in graphene: Structure, bonding, and magnetism. Phys Rev Lett 102:126807

Kresse G, Furthmüller J (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15

Kudin KN, Ozbas B, Schniepp HC, Prud'homme RK, Aksay IA, Car R (2008) Raman spectra of graphite oxide and functionalized graphene sheets. Nano Lett 8:36

Lalmi B, Oughaddou H, Enriquez H, Kara A, Vizzini S, Ealet B, Aufray B (2010) Epitaxial growth of a silicene sheet. Appl Phys Lett 97:223109

Li D, Muller MB, Gilje S, Kaner RB, Wallace GG (2008) Processable aqueous dispersions of graphene nanosheets. Nat Nanotechnol 3:101

Li ZY, Zhang WH, Luo Y, Yang JL, Hou JG (2009) How graphene is cut upon oxidation? J Am Chem Soc 131:6320

Liu JZ, Trimarchi G, Zunger A (2007) strain-minimizing tetrahedral networks of semiconductor alloys. Phys Rev Lett 99:145501

Luo X, Yang J, Liu H, Wu X, Wang Y, Ma Y, Wei S-H, Gong X, Xiang HJ (2011) Prediction of silicon-based layered structures for optoelectronic applications. J Am Chem Soc 133:16285

Mattson EC, Pu H, Cui S, Schofield MA, Rhim S, Lu G, Nasse MJ, Ruoff RS, Weinert M, Gajdardziska-Josifovska M, Chen J, Hirschmug CJ (2011) Evidence of nanocrystalline semiconducting graphene monoxide during thermal reduction of graphene oxide in vacuum. ACS Nano 5:9710

Meng L et al (2013) Buckled silicene formation on Ir(111). Nano Lett 13:685

Novoselov KS, Geim AK, Morozov SV, Jiang D, Katsnelson MI, Grigorieva IV, Dubonos SV, Firsov AA (2005) Two-dimensional gas of massless Dirac fermions in graphene. Nature 438:197

Novoselov KS, McCann E, Morozov SV, Falko VI, Katsnelson MI, Zeitler U, Jiang D, Schedin F, Geim AK (2006) Unconventional quantum Hall effect and Berry's phase of 2ϕ in bilayer graphene. Nat Phys 2:177

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made Simple. Phys Rev Lett 77:3865

Peres NMR (2010) Colloquium: The transport properties of graphene: An introduction. Rev Mod Phys 82:2673

Slijivancanin Z, Andersen M, Hornekar L, Hammer B (2011) Structure and stability of small H clusters on graphene. Phys Rev B 83:205426

Sofo JO, Chaudhari AS, Barber GD (2007) Graphane: A two-dimensional hydrocarbon. Phys Rev B 75:153401

Son Y, Cohen ML, Louie SG (2006) Half-metallic graphene nanoribbons. Nature 444:347

Stankovich S, Dikin DA, Dommett GHB, Kohlhaas KM, Zimney EJ, Stach EA, Piner RD, Nguyen ST, Ruoff RS (2006) Graphene-based composite materials. Nature 442:282

Stojkovic D, Zhang P, Lammert PE, Crespi VH (2003) Collective stabilization of hydrogen chemisorption on graphenic surfaces. Phys Rev B 68:195406

Takeda K, Shiraishi K (1994) Theoretical possibility of stage corrugation in Si and Ge analogs of graphite. Phys Rev B 50:14916

Thompson BC, Frechet JMJ (2008) Polymer-fullerene composite solar cells. Angew Chem Int Ed 47:57

Tsetseris L, Pantelides ST (2012) Hydrogen uptake by graphene and nucleation of graphane. J Mater Sci 47:7571

Tung VC, Allen MJ, Yang Y, Kaner RB (2009) High-throughput solution processing of large-scale graphene. Nat Nanotechnol 4:25

van de Walle A, Asta M, Ceder G (2002) The alloy theoretic automated toolkit: a user guide. CALPHAD Comput Coupling Phase Diagr Thermochem 26:539

Vogt P, De Padova P, Quaresima C, Avila J, Frantzeskakis E, Asensio MC, Resta A, Ealet B, Le Lay G (2012) Silicene: Compelling experimental evidence for graphenelike two-dimensional silicon. Phys Rev Lett 108:155501

Vogt P, Capiod P, Berthe M, Resta A, De Padova P, Bruhn T, Le Lay G, Grandidier B (2014) Synthesis and electrical conductivity of multilayer silicene. Appl Phys Lett 104:021602

Wang X, Tabakman SM, Dai H (2008) Atomic layer deposition of metal oxides on pristine and functionalized graphene. J Am Chem Soc 130:8152

Wang X, Li X, Zhang L, Yoon Y, Weber PK, Wang H, Guo J, Dai H (2009) N-doping of graphene through electrothermal reactions with ammonia. Science 324:768

Wassmann T, Seitsonen AP, Saitta AM, Lazzeri M, Mauri F (2010) Clar's theory, π-electron distribution, and geometry of graphene nanoribbons. J Am Chem Soc 132:3440

Watanabe K, Taniguchi T, Kanda H (2004) Direct-bandgap properties and evidence for ultraviolet lasing of hexagonal boron nitride single crystal. Nat Mater 3:404

Wei S-H, Ferreira LG, Zunger A (1990) First-principles calculation of temperature-composition phase diagrams of semiconductor alloys. Phys Rev B 41:8240

Wei S-H, Zhang SB, Zunger A (1999) Effects of Na on the electrical and structural properties of CuInSe$_2$. J Appl Phys 85:7214

Wilson NR, Pandey PA, Beanland R, Young RJ, Kinloch IA, Gong L, Liu Z, Suenaga K, Rourke JP, York SJ, Sloan J (2009) Graphene oxide: Structural analysis and application as a highly transparent support for electron microscopy. ACS Nano 3:2547

Xiang HJ, Kan EJ, Wei S-H, Whangbo M-H, Yang JL (2009) "Narrow" graphene nanoribbons made easier by partial hydrogenation. Nano Lett 9:4025

Xiang HJ, Kan EJ, Wei S-H, Gong XG, Wangbo M-H (2010a) Thermodynamically stable single-side hydrogenated graphene. Phys Rev B 82:165425

Xiang HJ, Wei S-H, Gong XG (2010b) Structural motifs in oxidized graphene: A genetic algorithm study based on density functional theory. Phys Rev B 82:035416

Xiang HJ, Huang B, Li ZY, Wei S-H, Yang JL, Gong XG (2012) Ordered Semiconducting Nitrogen-Graphene Alloys. Phys Rev X 2:011003

Xiang HJ, Huang B, Kan E, Wei S-H, Gong X-G (2013) Towards direct-gap silicon phases by the inverse band structure design approach. Phys Rev Lett 110:118702

Yan J-A, Xian L, Chou MY (2009) Structural and electronic properties of oxidized graphene. Phys Rev Lett 103:086802

Yang J-H, Zhang Y, Yin W-J, Gong XG, Yakobson BI, Wei S-H (2017a) Two-dimensional SiS layers with promising electronic and optoelectronic properties: Theoretical prediction. Nano Lett 16:1110

Yang J-H, Yuan Q, Deng H-X, Wei S-H, Yakobson BI (2017b) Earth-abundant and non-toxic SiX (X = S, Se) monolayers as highly efficient thermoelectric materials. J Phys Chem C 121:123

Zhang Y, Tan Y-W, Stormer HL, Kim P (2005) Experimental observation of the quantum Hall effect and Berry's phase in graphene. Nature 438:201

Zhang WH, Carravetta V, Li ZY, Luo Y, Yang JL (2009a) Oxidation states of graphene: Insights from computational spectroscopy. J Chem Phys 131:244505

Zhang Y, Tang T-T, Girit C, Hao Z, Martin MC, Zettl A, Crommie MF, Shen YR, Wang F (2009b) Direct observation of a widely tunable bandgap in bilayer graphene. Nature 459:820

Zhang Y, Rubio A, Le Lay G (2017) Emergent elemental two-dimensional materials beyond graphene. J Phys D Appl Phys 50:053004

Zhou J, Wang Q, Sun Q, Chen XS, Kawazoe Y, Jena P (2009) Ferromagnetism in semihydrogenated graphene sheet. Nano Lett 9:3867

Zhu Z, Guan J, Liu D, Tománek D (2015) Designing isoelectronic counterparts to layered group v semiconductors. ACS Nano 9:8284

Spins in Semiconductor Nanocrystals

27

Gustavo Martini Dalpian

Contents

Abstract

Semiconductor nanocrystals are being used as hosts to trap and manipulate
single spins. Spins in nanocrystals can have different properties than their bulk
counterparts, owing both to quantum confinement and surface effects. We will
show that spins can be generated in nanocrystals by the insertion of impurities,
either magnetic or not, or by magnetic polarons present at dangling bonds at their
surfaces. This chapter reports theoretical contributions to this field, where the
density functional theory is used to simulate these functionalized nanocrystals.

1 Introduction

During the last 30 years, the development of nanoscience has led to the possibility
of manipulating nanostructures at the atomic level. Nanocrystals can be grown in

G. M. Dalpian (✉)
Center for Natural and Human Sciences, Federal University of ABC, Santo André, SP, Brazil
e-mail: gustavo.dalpian@ufabc.edu.br

a wide variety of geometries and shapes, and for many of the known compounds, leading to the possibility of fine-tuning their properties. When the *relatively-new* field of nanoscience met with the mature area of magnetism, new phenomena arose. Having magnetic spins in nanocrystals leads to the possibility of controlling and manipulating single spins in an atomic level. Besides the exciting fundamental challenges related to single-spin manipulation, there is also a wide variety of possible applications of this type of technology, going from spin filters (Efros et al. 2001) to single electron transistors (Qu and Vasilopoulos 2006). The ability to manipulate quantum processes at the level of single impurities forms the basis of an area of research called solotronics (Kobak et al. 2014).

Understanding the nature of spins in semiconductor nanocrystals is, consequently, an important task in order to successfully manipulate them. Theoretical and computational tools based on the density functional theory (DFT) and on pseudopotentials methods are a suitable choice for modeling these materials since they combine precision with the possibility of simulating the nanostructures in similar ways as they occur in the laboratory. Modern methodologies can simulate nanostructures with thousands of atoms, both with and without periodic boundary conditions.

In this text, we will qualify two different types of spins in semiconductor nanocrystals and report theoretical, DFT-based results on them. The first type of spin involves the insertion of extrinsic elements into the semiconductor host, in a procedure very similar to what is done for semiconductors (Dietl 2010). The most popular impurities are transition metal atoms, which usually have large exchange splittings, leading to large local magnetic moments. Besides these magnetic elements, it is also possible to observe spins and a magnetic response by inserting nonmagnetic impurities into the nanocrystal, such as carbon or nitrogen in ZnO. This might have advantages since standard elements used for magnetic semiconductors are often toxic and very expensive.

The second way of obtaining spins in nanocrystals is through the intrinsic existence of surfaces in these materials. Dangling bonds present at these surfaces might act as localized spins of magnetic ions, producing a similar effect as the insertion of transition metals into the host nanocrystal.

2   Doped Nanocrystals

Most bulk semiconductors can also be grown as low dimensional nanostructures. Even materials that have only been studied more deeply in the last few years, as the Halide Perovskites, have already been synthesized in the form of nanocrystals (Castañeda et al. 2016). The same is true for other, more traditional, semiconductors such as Si, ZnO, CdTe, CdSe, and so on. In order to broaden the possibility of applications of these nanostructures, it is important to functionalize them. Doping is one of the most conventional ways of functionalizing semiconductors and should also be a suitable path for semiconductor nanocrystals. However,

this can be a formidable task owing to kinetic limitations and self purification mechanisms.

Turnbull, in 1950, was the first one to argue that small droplets would have a smaller concentration of defects when compared to bulk (Turnbull 1950). He was interested in understanding the formation of crystal nuclei in liquid metals. This crystallization is usually induced by the presence of impurities or interfaces. His arguments were basically statistical: the probability of having impurities inside very small droplets should be much smaller than in larger crystals. In 2005, Erwin et al. (2005) argued that the limiting factor for doping a semiconductor nanocrystal was the binding energy of the impurity to the surface of the nanocrystal. In 2006, Dalpian and Chelikowsky (2006) showed that the formation energy of the impurity increases as the size of the nanocrystal decreases. This is a direct consequence of the fact that the impurity levels get deeper as the size of the nanocrystal decreases. This *pinning* of the impurity levels has also been observed experimentally (Norberg et al. 2006). Using DFT calculations, effective mass models, and magnetic circular dichroism spectroscopy, it has been shown that the donor binding energies of Co^{2+} impurities in ZnSe quantum dots are pinned irrespective of the size of the nanocrystal. This occurs owing to the smaller effect of quantum confinement in the localized Co_d levels when compared to the HOMO (highest occupied molecular orbital) and the LUMO (lowest unoccupied molecular orbital) of the nanocrystal.

A lot of advances in the area of doped nanocrystals happened in the last decade, leading to the successful control of doped nanocrystals from an experimental perspective (Mocatta et al. 2011).

From a theoretical perspective, in order to simulate semiconductor nanocrystals, one has to build molecular models that resemble what happens in these nanostructures. In this chapter, nanocrystals are generated by cutting spherical regions from a bulk crystal, as schematically shown in Fig. 1a. This nanocrystal will be defined by its center (that can be a bond, interstitial position or an atom) and its radius. Depending on the definition of the center of the nanocrystal, its symmetry might be changed (Dalpian et al. 2006). The generated structure, shown in Fig. 1b, will present a high density of surface dangling bonds, which will certainly interfere in the final results. Later on this chapter, the importance of these surface dangling bonds will be discussed. However, at this point, we are mostly interested in quantum confinement effects, and we would like to avoid surface states. To do that, the nanocrystal's surface atoms are saturated with pseudo-hydrogen atoms, carefully generated with fractional charges in order to maintain electron counting. Details about this theoretical procedure were thoroughly described by Huang et al. (2005a).

Within the methodology described above, one can easily obtain information about the quantum confinement effects in nanocrystals, such as the increase of the bandgap as the size of the nanocrystals decreases. Quantum confinement effects on doped nanocrystals can also be explored through these methods by inserting impurities into them. This procedure will be described in the next section.
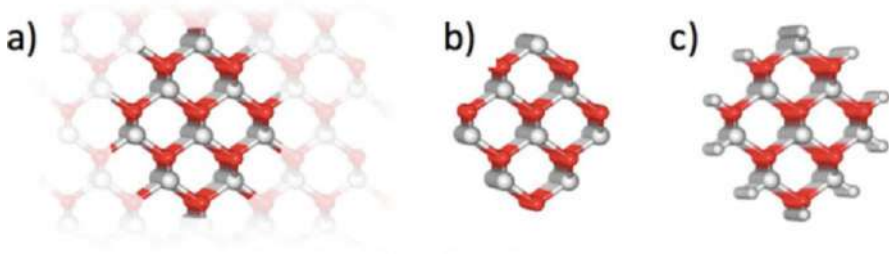
Fig. 1 Procedure used to generate spherical ZnO nanocrystals: (**a**) and (**b**) a spherical region is cut from the bulk crystal; (**c**) the surface dangling bonds are saturated with pseudo-hydrogen atoms. Red and grey circles represent oxygen and Zn atoms, respectively. The small, white circles represent the pseudo-hydrogen saturation

2.1 Magnetic Impurities

To simulate magnetic impurities inside nanocrystals, we substitute one of the internal atoms of the nanocrystal by a transition metal impurity. This process has been successfully done experimentally for a variety of materials and dopants, providing a large quantity of experimental results for comparison. These results include Mn in ZnS (Donegá et al. 2002), Mn in CdS (Nag et al. 2010), Co and Ni in ZnO (Schwartz et al. 2003), Co in ZnSe (Norberg and Gamelin 2005), and many others.

From a computational point of view, one of the first works dedicated to understand the spintronic properties of nanocrystals was performed by Huang, Makmal, Chelikowsky, and Kronik in 2005 (Huang et al. 2005b). They studied the magnetic properties of Mn-doped Ge, GaAs, and ZnSe nanocrystals using real space ab initio pseudopotential methodologies (Chelikowsky et al. 1994), based on the density functional theory. These calculations were performed with the PARSEC code (Kronik et al. 2006). In this chapter, we will not detail the methodology used for performing these calculations, since this has already been discussed in other chapters. However, it is important to mention that real space methods have the advantage of avoiding the need to use large supercells, with large vacuum layers, that are usually needed in plane wave codes. The important parameter in this case is the grid spacing, which might influence in the final calculated properties (Chelikowsky et al. 2011).

When Mn is inserted into these semiconductor nanocrystals, the half-metallicity trends observed in the bulk are preserved. Substituting a Ge atom by Mn will introduce Mn_d orbitals into the band gap of the nanocrystal. Owing to the crystal field, these d levels will split into a tri-degenerated t_2 and a two-degenerated e level. The picture is very similar for GaAs and ZnSe. For Ge, there will be two holes in the t_2 level, whereas substituting Ga by Mn in GaAs will insert one hole. No holes are inserted when Mn substitutes Zn in ZnSe. This effect has been studied for four different sizes of nanocrystals: X_9MnY_{10}, $X_{18}MnY_{19}$, $X_{40}MnY_{41}$, and $X_{64}MnY_{65}$, where X = Ge, As or Zn and Y = Ge, As or Se, respectively, and the index indicates the number of atoms of that type in the nanocrystal. It was observed

that, as the impurity states are highly localized, they are less affected by quantum confinement than the delocalized host states, such as the HOMO and the LUMO. As the size of the nanocrystals decreases, the Mn-related levels become deeper inside the band gap. This causes the ferromagnetic stabilization to be dominated by double exchange mechanisms via localized holes.

The insertion of Mn atoms into II-VI semiconductor does not insert carriers (holes) into the system, since Mn is isovalent to the type-II atom. Consequently, Mn substitution into ZnO nanocrystals does not induce a ferromagnetic response. However, May et al. (2012) have shown that a ferromagnetic response can be obtained by the insertion of external p-type dopants. This can be done by photoexcitation of electrons or by atomic substitution. May suggested that doping ZnO quantum dots with N (at the oxygen site) would lead to a ferromagnetic response in Mn-doped ZnO.

Owing to the localized nature of transition metal impurities, a frequent concern relies on the appropriateness of density functionals to simulate such materials. Whereas the generalized gradient approximation (GGA) or local density approximation (LDA) are very popular amongst chemists and physicists, they should not be applied in a straightforward manner for all systems. In order to elucidate this issue, Badaeva et al. (2008) tested several density functionals to better understand the properties of Co-doped ZnO nanocrystals. They have compared experimental data for the dopant-carrier magnetic exchange interactions with LSDA, GGA, and hybrid PBE1 functionals (Perdew et al. 1997). Hybrid functionals include a fraction of Hartree-Fock exchange, leading to a better description of localized states and of bandgaps. The overall conclusion was that hybrid functionals are usually better suited to study this type of system, although qualitative responses are already obtained with other functionals.

2.2 Nonmagnetic Impurities

In the previous section, we discussed spins in semiconductor nanocrystals that were created by the insertion of transition metal atoms. However, it is also possible to get spins in semiconductor nanocrystals by inserting nonmagnetic atoms into the nanocrystals or by the presence of other types of defects.

One of the first theoretical reports on ferromagnetism in nondoped semiconductors was made by Pemmaraju and Sanvito (2005), in a study about intrinsic point defects in HfO_2. They have observed that oxygen vacancies are spin-polarized in this kind of system and that there are exchange interactions between neighboring vacancies, leading to the possibility of a macroscopic magnetization of the whole crystal. This result was important to help understand experimental observations of magnetism in nondoped samples of HfO_2.

As intrinsic defects, which happen spontaneously in the crystal, can lead to a macroscopic magnetization, one could use similar principles to engineer magnetism into nanocrystals without the need of transition metal atoms. This task can also be obtained by inserting nonmagnetic impurities, such as Carbon or Nitrogen, into quantum dots. To test this idea, Kwak and Cheliwkoswky (2009) studied Carbon-

doped ZnO nanocrystals. In this case, Carbon atoms enter in substitution to oxygen, inserting two holes into the system. The ground state of C-doped ZnO is spin-polarized, with a magnetic moment of $2\mu_B$, originated from its *2p* states. As these are mainly localized states, it has been shown that quantum confinement has limited effect on the defect states, making ferromagnetism robust through different sizes of nanocrystals.

3   Surface Magnetization in Semiconductor Nanocrystals

The surface saturation described previously in Fig. 1c is a theoretical artifact to remove the nanocrystals' surface states from the energy gap. This is only performed to isolate the effects of quantum confinement. Although the effects of quantum confinement are modeled correctly within this model, the surface does not represent what happens in experiment. This type of saturation models a perfect saturation, where the density of dangling bonds at the surface tends to zero. Although this methodology is precise to study quantum confinement effects, we lose the rich nature of the surface of the nanocrystals. In order to analyze the surface effects on nanoparticles and their implications on properties, we have removed all surface saturation and then optimize these structures, allowing all atoms to move to their minimum energy positions. The surface of a nanocrystal can reconstruct in several new arrangements. Owing to these structural changes, and the presence of dangling bonds, the electronic structure of the nanocrystal is also altered, and a magnetic signal, or spin polarization, can also be observed. This magnetization is usually termed as d_0 ferromagnetism, similar to the case of nonmagnetic doping described previously.

From an experimental perspective, there are several observations of ferromagnetic responses in materials that should be nonmagnetic. This *phantom ferromagnetism* (Coey et al. 2008) was first reported in 2008. For nanocrystals, Garcia (Garcia et al. 2007) showed that nondoped ZnO nanoparticles could have a small ferromagnetic response, depending on the way the nanocrystals are saturated. When the nanocrystal is covered with Thiol molecules, a larger magnetization is observed than when the nanocrystals are covered with TOPO. These different signals are explained by different surface saturations, leading consequently to a different density of surface dangling bonds.

The nanocrystals observed in these experiments represent a saturation that is in between the two saturation limits mentioned above: (i) full pseudo-hydrogen saturation and (ii) no saturation. By studying nonsaturated nanocrystals, we will be able to learn about the properties of these dangling bonds and infer about what happens in experiment.

Spins on dangling bonds at the surface of CdSe colloidal nanocrystals have been recently reported and studied (Biadala et al. 2017). These spins have been called "magnetic polarons."

3.1   Surface Reconstructions

To simulate the surface of nanocrystals, after carving a spherical nanocrystal in a bulk region, as shown in Fig. 1a, we minimize the forces on all atoms letting the surfaces reconstruct. We have analyzed ZnO nanocrystals in the zinc blende (ZB) and the wurtzite (WZ) structures and were able to observe several different types of reconstructions and several different structural motifs for ZnO nanocrystals. We have studied different sizes of nanocrystals, with diameters ranging from 0.9 nm to 1.5 nm. The surface of each nanocrystal was reconstructed in a different way, giving us a broad range of different defects, including planar faces, steps, kinks, adislands, adatoms, dimers, and others. We have considered the majority of possible motifs present at any nanocrystal surface.

The most frequent motifs are the formation of planar faces and of dimers. Figure 2 shows a structural model of a saturated ZnO nanocrystal with 147 atoms and of the nonsaturated, optimized counterpart. This planar reconstruction on ZnO has also been reported for thin films (Tushe et al. 2007). Most of these reconstructions occur to reduce the dangling bond density at the surface. The formation of dimers, as also observed in Fig. 2, is very common to several surfaces, including Si(001).

Besides surface reconstructions, removal of bond saturation leads to the striking observation of surface magnetization in these structures. The magnetization is strongly localized at the surfaces of the nanocrystals, with almost no signal at their interior. Figure 3 is an illustration of this phenomenon. The figure shows the radial distribution function of the spin density for different sizes of ZnO nanocrystals. The insets show the spin charge density, i.e., the total charge density for spin up minus total charge density for spin down. This radial distribution clearly shows that the magnetization in these nanostructures is mostly localized at the surface.

Although many motifs that present magnetization were observed, the surface magnetization was not observed in all of our simulated nanocrystals. Our conclusion is that some kind of atomic order inside the nanocrystal is needed for
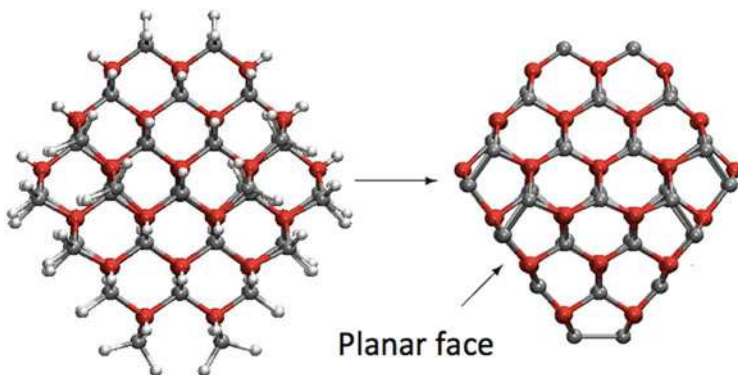
Fig. 2 Illustration of planar faces that occur in some nanocrystals after the passivation is removed. In this figure, the ZnO nanocrystal has 147 Atoms
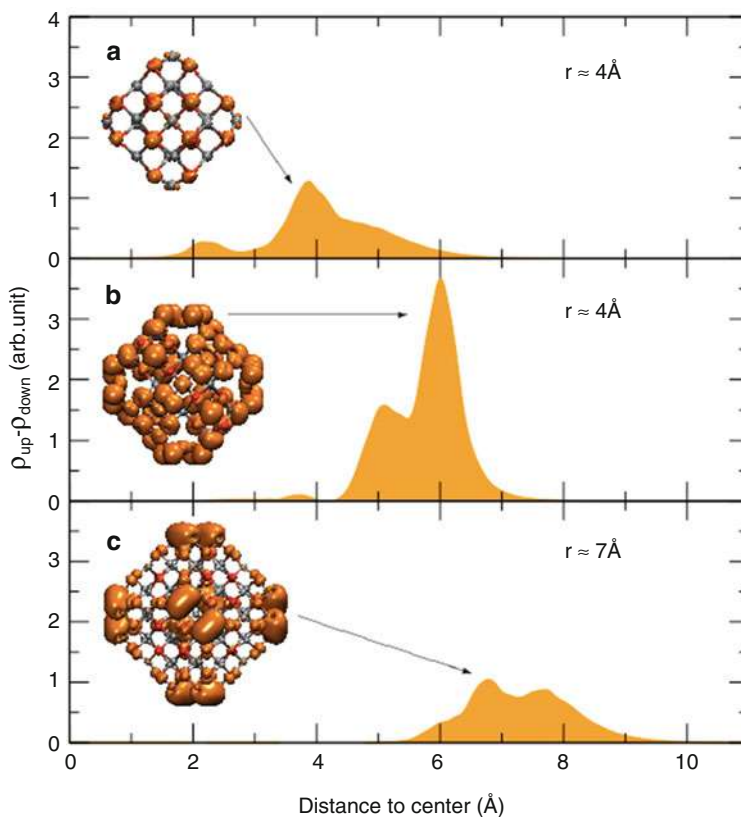
Fig. 3 Radial distribution of the spin density for three sizes of ZnO nanocrystals: (**a**) 4 Å; (**b**) 6 Å; (**c**) 7 Å. The inset shows a real space representation of the spin density distribution within the nanocrystal

the magnetization to be observed. For nanocrystals which structure was strongly reorganized, almost resembling an amorphous structure, no magnetization was observed (Schoenhalz et al. 2009 and 2010). Besides testing different sizes of nanocrystals (with up to 238 atoms), we have also tested different crystal structures. The ground state crystal structure of bulk ZnO is the wurtzite structure, although the energy difference to the zinc blende structure is not large. On Table 1 we report the total magnetization of the studied nanostructures.

Other nanostructured ZnO materials have also been shown to present a ferromagnetic ordering without the need of magnetic impurities. This is the case of ZnO nanowires (Podila et al. 2010), where the surface magnetization can also be used to explain the observed results.

As surfaces are delocalized and spread over the whole material, this effect should suffice to explain the observed magnetization in nondoped ZnO nanostructures (Garcia et al. 2007). Other extended defects such as grain boundaries and dislocations should also be able to hold a macroscopic magnetization.

Table 1 Total magnetization for nonsaturated ZnO nanocrystals (in Bohr magnetons). The number of atoms in the nanocrystal and the structure (ZB = zinc blende and WZ = wurtzite) is also shown (Schoenhalz et al. 2009)

Number of atoms	Structure	Total magnetization
35	ZB	$2\mu_B$
87	ZB	$4\mu_B$
147	ZB	$2\mu_B$
38	ZB	0
86	ZB	$2\mu_B$
238	ZB	0
39	WZ	0
92	WZ	$2\mu_B$
34	WZ	0
88	WZ	0

3.2   Interaction Between Surfaces and Magnetic Impurities

As demonstrated so far, spins in semiconductor nanostructures can be observed when impurities are inserted into the nanocrystal or by the simple fact that they exhibit large surface:volume ratios and that the surface is not perfectly saturated, presenting dangling bonds. A larger degree of complexity appears when both impurity spins and surface magnetizations are considered. This leads to the possibility of an exchange interaction between the delocalized spins at the surface and the localized impurity states.

A prototype case for studying this system is Co-doped ZnO nanocrystals (Schoenhalz and Dalpian 2013). There are plenty of theoretical and experimental results for the bulk system, with strong controversies about its ground state. Co has seven d electrons, and when inserted into ZnO, these d levels will be placed at the band gap. The total magnetic moment per Co atom is $3\mu_B$, because the e levels are lower in energy than the t_2 levels. Figure 4 shows the most important eigenvalues of three different nanocrystals and a comparison to the nondoped nanocrystals. For bulk ZnO, as the band gap is reduced in ab initio calculations, the Co d levels might be incorrectly placed away from the bandgap. Nanocrystals, with larger bandgaps, do not suffer from these problems.

When two impurities are present inside the nanocrystal, the most stable magnetic coupling is antiferromagnetic, since no carriers are present. This leads to a controversy because experimental results report a ferromagnetic configuration for this kind of system. Our results indicate that this ferromagnetism can be understood by the existence of the surfaces, as reported before for pure ZnO nanocrystals.

To simulate the effect of the surfaces on doped nanocrystals, we have used a procedure similar to the one used for nondoped nanocrystals: we remove the surface saturation, insert an impurity inside the nanocrystal, and optimize its structure. A very rich variety of spin configurations has been observed, depending on the size, structure, surfaces, and number of dopants inside the nanocrystal. Besides the usual hybridization and crystals field splittings observed in bulk ZnO and on saturated ZnO nanocrystals, in nonsaturated nanocrystals we also observe a strong interaction

Fig. 4 Comparison of energy levels of a doped and undoped zinc blende NC. The set of eigenvalues in the left are for the pure, undoped nanocrystal, and the second and third sets represent the majority and minority spin channels, respectively. The undoped NC has (**a**) 35, (**b**) 87, and (**c**) 147 atoms of Zn and O. The doped NC had one of the Zn atoms (the most central) substituted by a Co atom for each nanocrystal
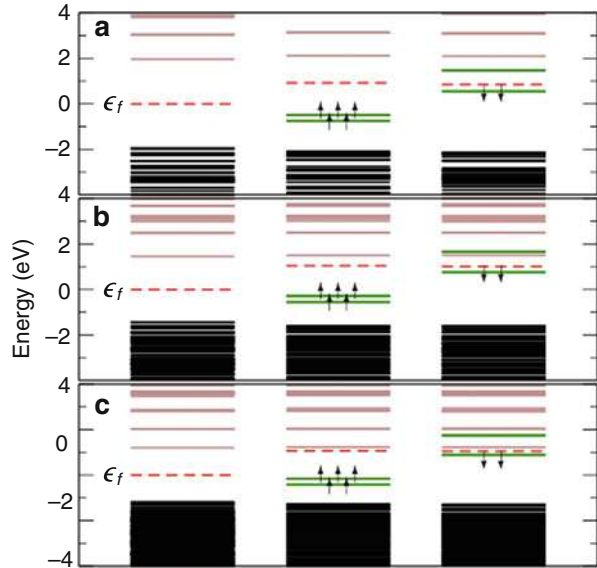
Fig. 5 Radial distribution of the spin density for ZnO nanocrystals. The green line indicates the case of the Co-doped nanocrystal, and the yellow line indicates the case for the bare nanocrystal. The insets show the real space representation of the spin density
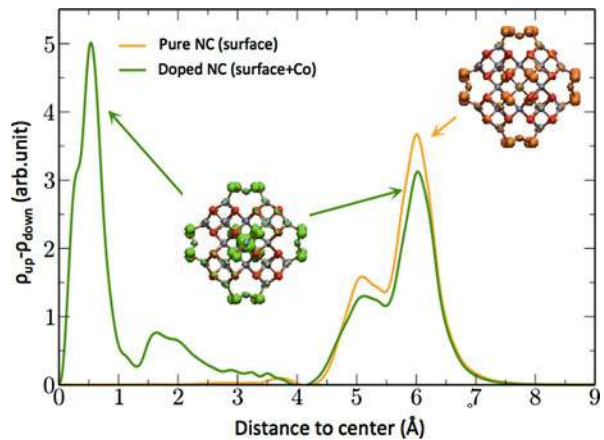
between the Co atom and the surface states, lowering the symmetry and tuning the occupation of the impurity atoms.

Figure 5 presents the radial spin density distribution for a Co-doped ZnO nanocrystal. In this case, there is only one impurity inside the nanocrystal. The curve in green is for the doped nanocrystal, with one Co atom near the center of the nanocrystal. For this case, we observe a magnetization both at the center (Co atom) and at the surface of the nanocrystal. For the sake of comparison, we also show the magnetization for the non-doped nanocrystal. In this case, only the surface magnetization is observed.

For the doped nanocrystal case, the removal of the surface saturation had two main effects. The first one was to induce a surface magnetization, as shown in Fig. 5. The second effect is related to the magnetic coupling between two Co impurities inside the nanocrystal. For the saturated case, as there are no carriers in Co-doped ZnO, the most stable configuration between the spins of the Co atoms is antiparallel, i.e., antiferromagnetic. For unsaturated, reconstructed nanocrystals, the lowest energy configuration is when the spins of the Co atoms are parallel to each other, i.e., a ferromagnetic configuration. This occurs owing to the strong hybridization of the Co atoms with the surfaces, which might act as donors or acceptors changing the magnetic ground state. (Schoenhalz and Dalpian, 2013)

4   Conclusions and Perspectives

The results presented in this chapter show the complex nature of spins in semi-conductor nanocrystals. Owing to quantum confinement and surface effects, the behavior of these spins might be very different from their bulk counterparts. New phenomena are observed. Also, due to the reduced size of the nanocrystals, one has the possibility of inserting single impurities into them, leading to the possibility of single spin manipulation. This has potential applications ranging from quantum computing to spintronic devices.

The results presented in this chapter clearly show the importance of considering both quantum confinement and the surface structure for understanding the ground state properties of doped nanocrystals. As there are limited experimental methods that can precisely track the properties of these surfaces, theoretical methods become indispensable tools to understand their properties.

References

Badaeva E, Feng Y, Gamelin DR, Li X (2008) Investigation of pure and Co^{2+}-doped ZnO quantum dot electronic structures using the density functional theory: choosing the right functional. New J Phys 10:055013

Biadala L et al (2017) Magnetic polaron on dangling-bond spins in CdSe colloidal nanocrystals. Nat Nanotech 12:569

Castañeda JA et al (2016) Efficient biexciton interaction in perovskite quantum dots under weak and strong confinement. ACS Nano 10:8603

Chelikowsky JR, Troullier N, Saad Y (1994) Finite-difference-pseudopotential method: electronic structure calculations without a basis. Phys Rev Lett 72:1240

Chelikowsky JR, Alemany MMG, Chan TL, Dalpian GM (2011) Computational studies of doped nanostructures. Rep Prog Phys 74:1

Coey JMD, Wongsaprom K, Alaria J, Venkatesan M (2008) Charge-transfer ferromagnetism in oxide nanoparticles. J Phys D Appl Phys 41:134012

Dalpian GM, Chelikowsky JR (2006) Self-purification in semiconductor nanocrystals. Phys Rev Lett 96:226802

Dalpian GM, Tiago ML, del Puerto ML, Chelikowsky JR (2006) Symmetry considerations in CdSe nanocrystals. Nano Lett 6:501

Dietl T (2010) A ten-year perspective on dilute magnetic semiconductors and oxides. Nat Mat 9:965

Donegá CM, Bol AA, Meijerink A (2002) Time-resolved luminescence of ZnS: Mn^{2+} nanocrystals. J Lumin 96:87

Efros AL, Rashba EI, Rosen M (2001) Paramagnetic ion-doped nanocrystal as a voltage-controlled spin filter. Phys Rev Lett 87:206601

Erwin SC, Zu L, Efros AL, Kennedy TA, Norris DJ (2005) Doping semiconductor nanocrystals. Nature 436:7047

Garcia MA et al (2007) Magnetic properties of ZnO nanoparticles. Nano Lett 7:1489

Huang X, Lindgren E, Chelikowsky JR (2005a) Surface passivation method for semiconductor nanostructures. Phys Rev B 71:165328

Huang X, Makmal A, Chelikowsky JR, Kronik L (2005b) Size-dependent spintronic properties of dilute magnetic semiconductor nanocrystals. Phys Rev Lett 94:236801

Kobak J et al (2014) Designing quantum dots for solotronics. Nat Commun 5:3191

Kronik L et al (2006) PARSEC–the pseudopotential algorithm for real-space electronic structure calculations: recent advances and novel applications to nano-structures. Phys Status Solidi B 243:1063

Kwak H, Chelikowsky JR (2009) Size-dependent induced magnetism in carbon-doped ZnO nanostructures. Appl Phys Lett 95:263108

May JW, McMorris RJ, Li X (2012) Ferromagnetism in p-type manganese-doped zinc oxide quantum dots. J Phys Chem Lett 3:1374

Mocatta D, Cohen G, Schattner J, Millo O, Rabani E, Banin U (2011) Heavily doped semiconductor nanocrystal quantum dots. Science 332:77

Nag A et al (2010) Size-dependent tuning of Mn^{2+} d emission in Mn^{2+}-doped CdS nanocrystals: bulk vs surface. J Phys Chem C 114:18323

Norberg NS, Gamelin DR (2005) Influence of surface modification on the luminescence of colloidal ZnO nanocrystals. J Phys Chem B 109:20810

Norberg NS, Dalpian GM, Chelikowsky RJ, Gamelin DR (2006) Energetic pinning of magnetic impurity levels in quantum-confined semiconductors. Nano Lett 6:2887

Pemmaraju CD, Sanvito S (2005) Ferromagnetism driven by intrinsic point defects in HfO_2. Phys Rev Lett 94:217205

Perdew JP, Burke K, Ernzerhof M (1997) Generalized gradient approximation made simple. Phys Rev Lett 78:1396

Podila R et al (2010) Origin of FM ordering in pristine micro- and nanostructured ZnO. Nano Lett 10:1383

Qu F, Vasilopoulos P (2006) Spin transport across a quantum dot doped with a magnetic ion. Appl Phys Lett 89:122512

Schoenhalz AL, Dalpian GM (2013) Cobalt-doped ZnO nanocrystals: quantum confinement and surface effects from ab initio methods. Phys Chem Chem Phys 15:15863

Schoenhalz AL, Arantes JT, Fazzio A, Dalpian GM (2009) Surface magnetization in non-doped ZnO nanostructures. Appl Phys Lett 94:162503

Schoenhalz AL, Arantes JT, Fazzio A, Dalpian GM (2010) Surface and quantum confinement effects in ZnO nanocrystals. J Phys Chem C 114:18293

Schwartz DA, Norberg NS, Nguyen QP, Parker JM, Gamelin DR (2003) Magnetic quantum dots: synthesis, spectroscopy, and magnetism of Co^{2+}- and Ni^{2+}-doped ZnO nanocrystals. J Am Chem Soc 125:13205

Turnbull D (1950) Formation of crystal nuclei in liquid metals. J Appl Phys 21:1022

Tusche C, Meyerheim HL, Kirschner J (2007) Observation of depolarized ZnO (0001) monolayers: formation of unreconstructed planar sheets. Phys Rev Lett 99:026102

Excited-State Properties of Thin Silicon Nanowires

28

Li Yang

Contents

Abstract

Calculating many-electron interactions is crucial in determining excited-state properties of low-dimensional structures and is often beyond the capability of density functional theory, which is a ground-state theory. First-principles many-body perturbation theory can better describe many-electron effects and has been successfully employed to study excited-state properties of a wide range of solids. Such properties include quasiparticle energies, excitonic effects, and optical absorption spectra. In one-dimensional ultrathin silicon nanowires, which have been regarded as building blocks of nanoscale devices, many-electron interactions are substantially enhanced because of reduced dimensionality and weak electronic screening. If we employ many-body perturbation theory based on GW calculations for a silicon nanowire with a diameter of 1.2 nm, the quasiparticle band gap is 3.2 eV. This value is twice of that calculated from

L. Yang (✉)
Department of Physics, Institute of Materials Science and Engineering, Washington University, St Louis, MO, USA
e-mail: yangli@wustl.edu; lyang@physics.wustl.edu

density functional theory. Excitonic effects and optical absorption spectra, which cannot be obtained solely from GW calculations, are obtained by solving the Bethe-Salpeter equation. The calculated electron-hole binding energy is 1.2 eV for the same sized silicon nanowire. This binding energy is about two orders of magnitude larger than those found in bulk semiconductors. Our results agree well with recent measurements.

1   Introduction

Many-electron interactions are crucial in understanding excited states and are of fundamental importance in the exploration of new physics and device applications. Bare electrons interacting with other electrons and nuclei form single-particle excitations known as electron "quasiparticles." Excited electrons and holes can attract each other by Coulombic interactions to form two-particle excitations, which are called excitons. Interacting excitons may form four-particle excitations called "biexcitons." An exciton can also interact with an extra electron or hole to form a three-particle excitation called a "trion." These many-body excitations serve as the foundation for a diverse assortment of electric, transport, magnetic, and optical properties of solids.

To date, numerous low-dimensional structures have been successfully fabricated and intensively studied, which include zero-dimensional (0D) quantum dots and nanocrystals; one-dimensional (1D) nanowires (NWs), nanotubes (NTs), and nanoribbons (NRs); and two-dimensional (2D) structures such as graphene, transition metal dichalcogenides, and black phosphorous. Among these structures, single crystalline silicon nanowires (SiNWs) have attracted significant attention as they are potential building blocks for nanoscale devices, such as field-effect transistors, logic circuits, sensors, lasers, and light-emitting diodes (Cui and Lieber 2001; Sirbuly et al. 2005). Consequentially, understanding and predicting excited-state properties in SiNWs are of fundamental interest.

Many-electron interactions are substantially enhanced in nanostructures owing to quantum confinement and weak electronic screening. As a result, excited-state properties, such as quasiparticle band gaps and optical spectra, are significantly different from those of bulk structures. For example, photoluminescence (PL) data reveal a substantial blueshift in semiconducting NWs with decreasing size (Katz et al. 2002). Scanning tunneling spectroscopy (STS) data show a significant increase of the electronic band gap of thin semiconducting SiNWs (Ma et al. 2003).

First-principles approaches based on density functional theory (DFT) (Hohenberg and Kohn 1964; Kohn and Sham 1965) are an extremely successful tool to study atomic and electronic structures of solids (Cohen and Louie 2016). However, DFT is a ground-state theory and cannot accurately capture many-electron effects and excited-state properties. In contrast, many-body perturbation theory (MBPT) (Louie and Rubio 2005) has been proven reliable for calculating and predicting excited-state properties of a broad range of systems. These systems include many molecules, quantum dots, 1D structures, 2D structures, and bulk crystals (Hybertsen

and Louie [1986](#); Rubio et al. [1993](#); Blasé et al. [1995](#); Onida et al. [1995](#); Oğüt et al. [1997](#); Albrecht et al. [1998](#); Rohlfing and Louie [1998](#), [2000](#); Spataru et al. [2004a](#); Park et al. [2006](#); Neaton et al. [2006](#); Yang et al. [2007c](#), [2009](#); Qiu et al. [2013](#); Tran et al. [2014](#)).

The computational framework of MBPT consists of three steps. DFT is employed to calculate the ground-state properties, e.g., to obtain minimum-energy atomic structures and Kohn-Sham eigenvalues and wavefunctions. Then quasiparticle energies are evaluated within the GW approximation to provide band gaps and band-edge energies. Finally, the Bethe-Salpeter equation (BSE) is solved to obtain the two-particle excitation (exciton) energies, exciton wavefunctions, and optical absorption spectra. Thin SiNWs will be used as an example to present details of MBPT simulations and excited-state properties of 1D nanostructures. Our DFT calculations were performed using the Quantum Espresso package (Giannozzi et al. [2009](#)). Our MBPT calculations were performed using the BerkeleyGW package (Deslippe et al. [2012](#)).

2 Excited-State Properties of SiNWs

2.1 DFT Results of SiNWs

Substantial advances have been achieved in fabricating single-crystal SiNWs with diameters as small as 1 nm and lengths of a few tens of μm (Ma et al. [2003](#)). Numerous experiments, such as PL and STS, have shown unique transport and optical properties of SiNWs, which dramatically deviate from those of bulk silicon (Katz et al. [2002](#); Ma et al. [2003](#)). Given the vital role played by silicon in the semiconductor industry and the characteristic size of future microelectronic devices (sub-10 nm), understanding quantum-confinement effects and predicting their impacts on excited-state properties of ultrathin SiNWs are of broad interests for both fundamental science and device applications.

SiNWs synthesized in experiment often grow along specific crystallographic directions. The wires have a roughly cylindrical shape along with a chemically passivated surface (Morales and Lieber [1998](#); Homles et al. [2000](#)). Self-purification processes can also occur in SiNWs. In such processes, impurities or defects spontaneously migrate from the interior of the wire to the surface (Dalpian and Chelikowsky [2009](#)) and improve the single-crystalline purity.

In computational approaches, the cylindrical wire model is constructed with a bulk-like structure in the interior. Dangling bonds on the surface are passivated by hydrogen atoms so that no complex of SiH_3 occurs (Zhao et al. [2004](#)). Shown in Fig. [1a](#) is a ball-and-stick model for a SiNW growing along the [110] direction with a diameter of 1.2 nm. Surface reconstruction can happen in SiNWs and affect its electronic and optical properties. For example, as shown in Fig. [1a](#), a full structural relaxation calculated within DFT in the local density functional approximation (LDA) shows that the direction of $Si-H_2$ bonds on the SiNW surface must be tilted to avoid soft phonon modes (Yang and Chou [2011](#)).
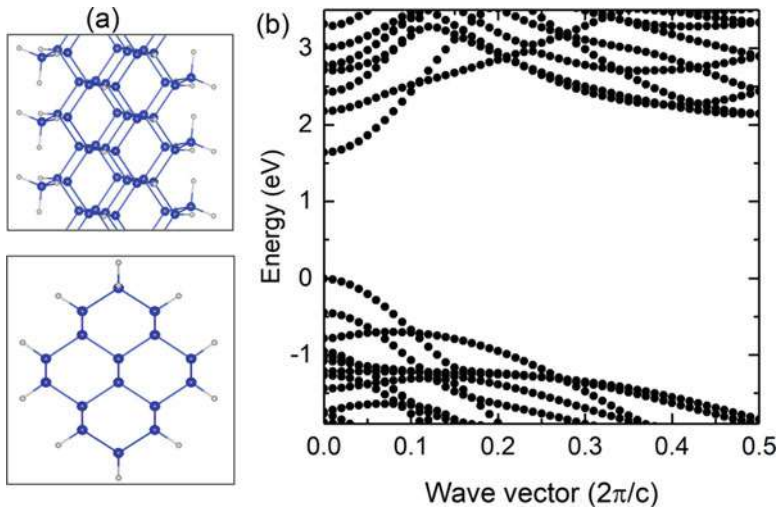
Fig. 1 (**a**) Side and top views of the ball-and-stick model of a H-passivated SiNW along the [110] direction. The diameter is around 1.2 nm. (**b**) DFT/LDA-calculated band structure of the 1.2-nm SiNW. The energy of the top of the valence band is set to be zero

Many new, important features of the electronic structures of SiNWs are found even at the DFT/LDA level. The DFT/LDA band structure of the 1.2-nm SiNW is presented in Fig. 1b. First, quantum confinement significantly increases the band gap: the DFT/LDA value is about 1.6 eV, which is much larger than that of the indirect DFT/LDA band gap of bulk silicon ($\sim$0.5 eV). Second, the band gap is direct in this thin SiNW. This feature may be of great help to broaden the range of silicon applications because the indirect band gap limits the use of bulk silicon in relevant optical applications. Further analysis (Zhao et al. 2004) has shown that this direct band gap is a band-folding effect; the conduction band minimum along the $\Gamma - X$ direction of bulk silicon is folded into the zone center of SiNWs. Interestingly, not all thin SiNWs exhibit the direct-gap feature. Other first-principles calculations predict that the direct/indirect band gap feature is sensitive to the orientation of SiNWs. For instance, the band gap can be indirect for thin SiNWs along the [111] direction (Yan et al. 2007). Finally, the degeneracy of light and heavy holes of the valence band maximum (VBM) of bulk silicon is eliminated. This is a quantum confinement effect because the energy levels of a free particle in a quantum well are inversely proportional to the particle mass. As a result, light holes and heavy holes experience different quantum conferment effects, and the degeneracy of energy levels is eliminated. Similar features have been predicted for the electronic structure of other types of ultrathin NWs, such as germanium NWs (Beckman et al. 2007).

In addition, studies based on DFT calculations have suggested various approaches for tuning the band gap and the absolute band-edge energies (referred to the vacuum level) of SiNWs. For example, applying axial strain can modulate the band gap, direct/indirect band-gap transitions, and band offsets of SiNWs

(Wu et al. 2009). Surface passivation and shape engineering can be utilized to control the band structure of SiNWs. For instance, OH passivation can enhance the band gap by a few hundred meV (Nolan et al. 2007). Tapered structures have been proposed to engineer quantum confinement and tune the absolute band-edge energies to realize charge separation within a single SiNW (Wu et al. 2008), which is useful for photovoltaic applications. Finally, motivated by studies of gated graphene nanoribbons (Son et al. 2006), an electrostatic gate field has been reported to be able to efficiently control the band structure of SiNWs (Nduwimana and Wang 2014).

2.2   Quasiparticle Band Structure and Band Gap of SiNWs

The GW approximation within MBPT (see Chap. 15, "Modeling Excited States of Confined Systems" of MTM by Hung and Öğüt) has proven successful in providing reliable quasiparticle band gaps of solids. Hedin suggested using MBPT to expand the self-energy into a product of the screened Coulomb interaction, W, and the single-particle Green's function, G, namely $\Sigma \approx iGW$, which is called the GW approximation (Hedin 1965). In the 1980s, first-principles GW calculations were established by Hybertsen and Louie (Hybertsen and Louie 1986). Since this pioneering work, GW theory has been successfully applied for obtaining reliable quasiparticle energy gaps of a wide range of solids (Hybertsen and Louie 1986; Shirley and Louie 1993; Blasé et al. 1994, 1995; Onida et al. 1995; Öğüt et al. 1997; Delerue et al. 2000; Zhao et al. 2004; Neaton et al. 2006; Yang et al. 2007c; Qiu et al. 2013)

We employ a non-self-consistent G_0W_0 calculation to calculate quasiparticle band gaps of a free-standing 1.2-nm SiNW. The computation setups of such GW calculations are critical to obtain converged quasiparticle energies (Shih et al. 2010; Peelaers et al. 2011; Ping et al. 2012 Qiu et al. 2013). The k-point sampling grid in the reciprocal space is set to be $1 \times 1 \times 32$. The static polarizability is calculated within the random-phase approximation (RPA), and the dynamic screening is obtained by the general plasmon pole (GPP) model (Hybertsen and Louie 1986). Usually the calculations of static polarizability and the Coulomb-hole part of self-energy require a sum of all unoccupied (conduction) states. Here a modified static-remainder approximation is applied to achieve the converged summation with a finite number of conduction states (Deslippe et al. 2013). Finally, a slab Coulomb truncation scheme is utilized to remove artificial interactions between neighboring periodic cells to mimic a free-standing SiNW (Ismail-Beig 2006; Rozzi et al. 2006). It is worth noting that the drawback of G_0W_0 is its dependence on the starting point.

In Fig. 2a, GW-calculated quasiparticle energies are presented according to their DFT/LDA energies. The most important feature is that many-electron interactions induce a significant self-energy correction: a 1.62 eV correction to the 1.6-eV DFT/LDA band gap, nearly a 100% enhancement. This self-energy correction is significantly larger than that of bulk silicon, which is about 0.7 eV. Such an enhanced GW correction agrees with other studies of 1D CNTs (Spataru et al. 2004a) and h-BN nanotubes (Park et al. 2006), in which the self-energy correction is also
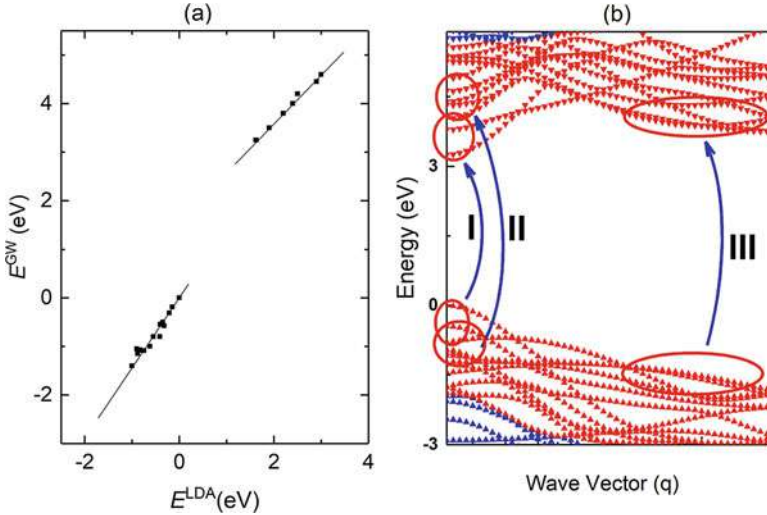
Fig. 2 (**a**) The quasiparticle energy vs the DFT/LDA energy. The linear fits are applied to those of valence states and conduction states, respectively. (**b**) The quasiparticle band structure of the 1.2-nm SiNW along the [110] direction. Those interband transitions, which are prominent in the optical absorption spectrum, are marked by "I," "II," and "III"

1∼3 eV. The reduced dimensionality and weak electronic screening of free-standing SiNWs are reasons for the enhancement of self-energy corrections. In fact, the dielectric function of free-standing SiNW is very different from that of bulk silicon. The diagonal components of the dielectric function of free-standing SiNWs are substantially smaller than those of bulk silicon (Yang et al. 2007a). Particularly within the long-wave limit, the dielectric function of free-standing thin SiNWs approaches the vacuum value (1). This is substantially smaller than the known static permittivity of bulk silicon, which is about 12.

The self-energy correction is an approximately linear function of the DFT/LDA energy in Fig. 2a. This behavior is similar to the one calculated in bulk semiconductors and 1D CNTs (Hybertsen and Louie 1986; Spataru et al. 2004a). As a result, linear functions can be used to fit GW corrections to the DFT/LDA band energies. With these fitted results, the whole quasiparticle band structure of a 1.2-nm SiNW can be plotted in Fig. 2b. Compared with the DFT/LDA band structure (Fig. 1b), the GW correction mainly increases of the band gap, and the important direct band-gap feature is kept.

Computing the quasiparticle band gap as a function of the diameter of SiNWs is of broad interest. Unfortunately, the simulation cost of GW calculations is proportional to N^4, where N is the number of electrons per unit cell. This is substantially more expensive than that of DFT computations, which scales as N^3 in principle, although slightly less than that in practice. With contemporary computation facilities, GW calculations of SiNWs with significantly larger diameters are formidable. An interpolation approach can be applied to approximate a solution to this problem. For instance, the quasiparticle band gap can be fit by

$$E_g = E_g^{bulk} + {A}\Big/{d^\alpha} \; , \tag{1}$$

where d is the diameter of SiNWs, and A and α are fitting parameters. The exponent α is usually about 1.4~1.8 (Zhao et al. 2004; Yan et al. 2007), which depends on the orientation of SiNWs, and smaller than 2, indicating that many-electron energy does not follow the simple scaling law of kinetic energy of a single particle within an infinite quantum well.

Additionally, the quasiparticle band gaps of SiNWs growing along different orientations have been calculated (Yan et al. 2007). Enhanced quasiparticle band gaps were identified. Interestingly, despite different orientations, GW corrections of SiNWs with similar diameters are found to be nearly identical.

The absolute quasiparticle band-edge energies referred to the vacuum level are important for broad applications. For example, SiNWs can be potentially useful for renewable energy applications such as water splitting (Zhang et al. 2011a), in which the absolute band edge energies are crucial for electron transfers and redox reactions. GW calculations for SiNWs with a diameter less than 2 nm result in absolute energy of the valence band maximum (VBM) lower than -5.5 eV, making such wires useful for water splitting (Yan et al. 2007).

GW quasiparticle band gaps also agree well with measurements. The STS measurements (Ma et al. 2003) showed that the band gap of ultrathin SiNWs with a diameter around 1 nm can reach 3.4 eV, which is very close to the GW-calculated value of 3.2 eV of 1.2-nm [110] SiNW (Zhao et al. 2004).

2.3   Optical Absorption Spectra and Excitonic Effects

In semiconductors, an electron can be promoted from a valence band to a conduction band by an incident photon, resulting in optical absorption. The excitation creates a positively charged "hole" in the valence band, which interacts with the excited electron owing to the screened Coulombic interaction. The interacting e-h pair forms an exciton. Excitonic effects are important for understanding optical properties of semiconductors and insulators. For example, as illustrated in Fig. 3a, without considering excitonic effect, the E_1 peak around 3.5 eV is notably absent in the calculated single-particle spectrum. When excitonic effects are included, the E_1 peak is resolved, and the overall optical absorption spectrum agrees well with experiment (Albrecht et al. 1998; Rohlfing and Louie 2000). Excitonic effects are also crucial for applications. The photovoltaic performance of solar cells is strongly influenced by exciton generation, diffusion, and dissociation (Gregg 2003). Previous work has demonstrated the importance of excitons in carbon and h-BN nanotubes (Spataru et al. 2004a; Park et al. 2006), semiconductor quantum dots, and nanocrystals (Oğüt et al. 1997). In the following, optical absorption spectra, excitonic effects, and fine structures of e-h pairs of thin SiNWs will be presented.

Single-particle results:

The single-particle optical absorption spectrum can be obtained within the RPA:
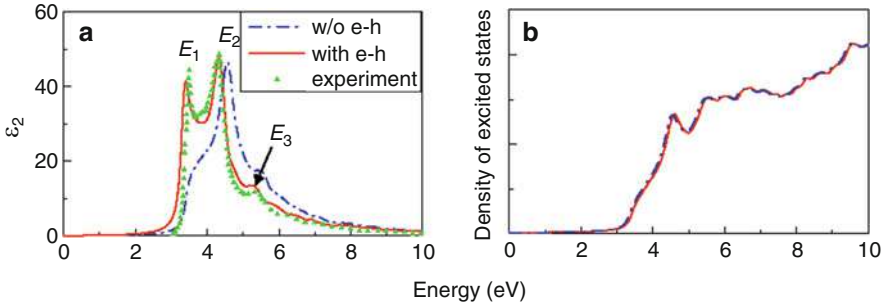
Fig. 3 (**a**) Optical absorption spectra and (**b**) density of excited states of bulk silicon with and without excitonic effects included. Experiment data (Philipp 1972; Aspnes and Theeten 1980; Lautenschlager et al. 1987) are included in (**a**). A Lorentzian broadening with a width of 0.10 eV is applied

$$\varepsilon_2 \left(\omega \right) = \frac{16\pi e^2}{\omega^2} \sum_{v,c} \left| \vec{\lambda} \cdot \left\langle v \middle| \vec{v} \middle| c \right\rangle \right|^2 \delta \left(\omega - (E_c - E_v) \right), \tag{2}$$

where $\vec{\lambda}$ is the polarization vector of incident light, v and c are valence and conduction states, respectively, and $\vec{v}$ is the velocity operator. E_c and E_v are the quasiparticle energies of conduction states and valence states, respectively. Only the interband transitions between conduction and valence states with the same crystal momentum are considered because the momentum of incident photon is usually much smaller than the size of the first Brillouin zone.

The single-particle optical absorption spectrum of the 1.2-nm SiNW along the [110] direction is illustrated in Fig. 4a. Twelve valence bands and 12 conduction bands are included in interband transitions for the converged absorption spectrum up to 8 eV. Although the direct band gap is around 3.2 eV, the single-particle optical absorption is not significant until 4.5 eV. This feature indicates that the optical interband transition between the VBM and CBM is very weak, and the direct band gap of SiNWs remains optically inactive. In fact, the interband transitions between band edges are forbidden by the face-centered cubic symmetry of silicon lattices, and the band-folding-induced direct band gap does not change the corresponding optical selection rule. As a result, in addition to the fundamental band gap, a symmetry gap exists in the optical absorption of SiNWs.

When compared with bulk silicon, new features emerge in the single-particle optical absorption spectrum of SiNWs. As marked in Fig. 4a, there are three dominant peaks (P_1, P_2, and P_3), which are different from the known double-peak (E_1 and E_2) feature of bulk silicon (Fig. 3a). These significant absorption peaks in SiNWs have complex origins associated with interband transitions. For example, the P_1 peak arises mainly from those transitions marked by "I" in Fig. 2b and the P_2 peak comes mainly from those transitions marked by "II" and "III."
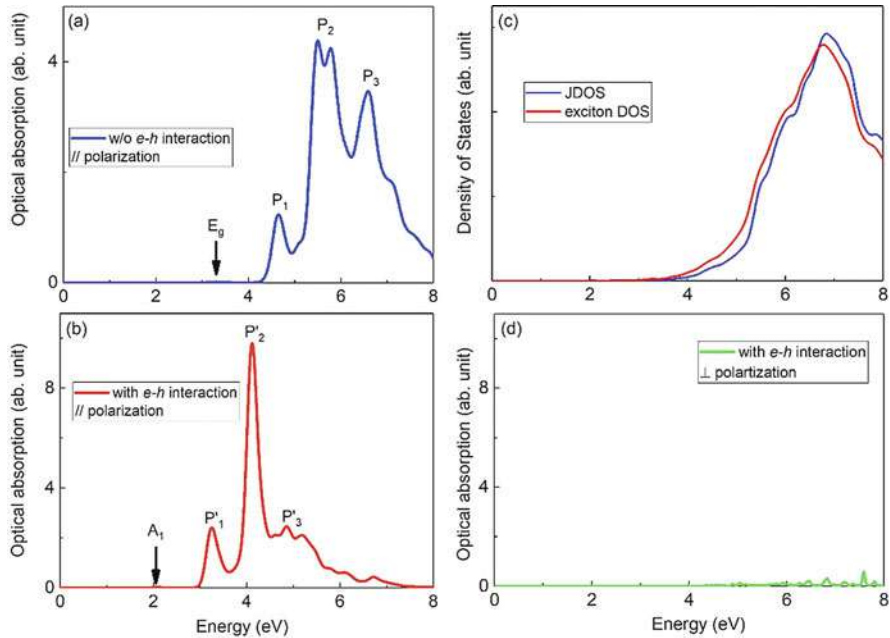
Fig. 4 (**a**) The single-particle optical absorption spectrum of the 1.2-nm SiNW along the [110] direction. (**b**) That with e-h interaction included. Both spectra are for incident photo polarized along the axial direction. The band gap (E_g) and the lowest-energy singlet exciton (A_1) are marked by arrows, respectively. (**c**) The joint density of single-particle states and the density of excitonic states. (**d**) The optical absorption spectrum of photon polarized vertically to the axial direction of the SiNW. Excitonic effects are included. A Lorentzian broadening with a width of 0.10 eV is applied to all these figures

A more detailed analysis (Yang et al. 2007a) shows that the transitions of II and III correspond to the E_1 peak in bulk silicon at 3.5 eV, which is blueshifted, enhanced, and split in thin SiNWs. The bulk E_1 peak is owing to transitions in the neighborhood of the L point in the bulk Brillouin zone between the top valence band and the nearly parallel lowest conduction band along the [111] direction. In the [110] SiNWs, the bulk L points map to either the Γ point or the zone boundary. Therefore, the corresponding transitions appear around $k = 0$ transitions labeled by II and between parallel bands in the 1D Brillouin zone (transitions labeled by III). The peak marked by P_3 relates to the bulk absorption peak E_2. It consists of many complicated interband transitions in the SiNW, which are not labeled in Fig. 2. Similar results of shifted and split absorption peaks have been observed in other semiconducting nanostructures, e.g., InP nanowires (Yu et al. 2003) and nanocrystals (Vasiliev et al. 2002).

Two-particle results:

Excitons are crucial for understanding the optical properties of semiconductors and insulators. Exciton eigenvalues and wavefunctions can be obtained by solving

the corresponding Dyson's equation. Under the Tamm-Dancoff approximation (Fetter and Walecka 1971), the wavefunction of exciton S in a crystal can be written as

$$\chi_S\left(\overrightarrow{x}_e, \overrightarrow{x}_h\right) = \sum_k \overset{\text{hole elec}}{\sum_v \sum_c} A^S_{vck} \psi_{ck}\left(\overrightarrow{x}_e\right) \psi^*_{vk}\left(\overrightarrow{x}_h\right), \tag{3}$$

where A^S_{vck} is the exciton amplitude, and $\psi_{ck}\left(\overrightarrow{x}_e\right)$ and $\psi_{vk}\left(\overrightarrow{x}_h\right)$ are the electron and hole wavefunctions, respectively. These single-particle wavefunctions are usually obtained by DFT (LDA or GGA) calculations. k is the wave vector of periodic systems. The Dyson's equation for the two-body Green's function can be written as

$$\left(E_{ck} - E_{vk}\right) A^S_{vck} + \sum_{k'v'c'} \langle vck| K^{eh} \left|v'c'k'\right\rangle A^S_{v'c'k'} = \Omega^S A^S_{vck}, \tag{4}$$

which is called the BSE. Many-body effects of *e-h* interactions are included into the interaction kernel K^{eh} that is a four-point energy-dependent function (Rohlfing and Louie 2000) and contains two types of *e-h* interactions, the bare *e-h* exchange interaction and the direct (screened) attractive interaction. The GW-calculated dielectric screening will be used in constructing the *e-h* interaction kernel K^{eh}. For most intrinsic semiconducting materials, since the plasmon energy is much larger than the *e-h* interaction energy, the static approximation is employed to eliminate the energy dependence of the interaction kernel (Rohlfing and Louie 2000), which substantially simplifies the simulation to solve BSE.

Once the BSE (Eq. 4) is solved, the two-body optical absorption spectrum can be obtained by

$$\varepsilon_2\left(\omega\right) = \frac{16\pi e^2}{\omega^2} \sum_S \left|\overrightarrow{\lambda} \cdot \sum_{vck} A^S_{vck} \langle vk| \overrightarrow{v} |ck\rangle\right|^2 \delta\left(\omega - \Omega_S\right), \tag{5}$$

The BSE approach has been used to compute excitonic states and corresponding optical absorption spectra for a wide range of nanostructures and bulk crystals (Onida et al. 1995; Albrecht et al. 1998; Rohlfing and Louie 1998, 1999, 2000; Spataru et al. 2004a; Capaz et al. 2006; Neaton et al. 2006; Park et al. 2006; Yang et al. 2007b, 2009; Qiu et al. 2013; Tran et al. 2014). The calculation of *e-h* interaction kernel is very time-consuming, so an interpolation scheme is usually adopted (Rohlfing and Louie 2000; Spataru et al. 2004b). The *e-h* interaction kernel is first calculated on a coarse k-grid in the reciprocal space and interpolated onto a finer k-grid, to improve the convergence of exciton eigenvalues and eigenfunctions. For the 1.2-nm SiNW a $1 \times 1 \times 32$ coarse k-grid is used to obtain the *e-h* interaction kernel, and an interpolated $1 \times 1 \times 128$ fine k-grid is used for the convergence of exciton energies (to within 0.1 eV). Truncation of the Coulomb interaction is also necessary to mimic a free-standing nanostructure (Ismail-Beig 2006; Rozzi et al. 2006).

Enhanced excitonic effects have been calculated in ultrathin SiNWs (Yang et al. 2007a; Bruno et al. 2007) and Si/Ge NWs (Palummo et al. 2010a) using the described BSE approach. The two-particle optical absorption spectrum of a 1.2-nm SiNW is presented in Fig. 4b, in which the polarization of incident light is along the axial direction. *e-h* interactions drastically change the single-particle absorption spectrum. All three main peaks are generally redshifted by about 1∼1.5 eV after inclusion of e-h interactions. This excitonic effect is substantially different from that observed in the spectrum of the bulk. There the *e-h* interaction mainly enhances the E_1 peak while the whole optical absorption spectrum does not significantly shift. In contrast, *e-h* interactions dramatically shift and reshape optical spectra in semiconducting nanostructures. If excitonic effects are neglected, the calculated optical spectra of nanostructures are not qualitatively correct in either energies or profiles.

Electron-hole binding energy is a direct measure of excitonic effects. As marked in Fig. 4b, the lowest-energy exciton (A_1) is located at 2 eV. Given that the quasiparticle band gap is ∼3.2 eV, the *e-h* binding energy is about 1.2 eV, which is about two orders of magnitude larger than that of bulk silicon (∼14 meV). Two reasons contribute to this enhanced *e-h* interaction: (1) The result in Fig. 4b is for a suspended thin SiNW. In such a free-standing 1D structure, most electric field lines are outside. Consequently, the overall dielectric screening is much weaker than that of bulk structures. (2) The thin SiNW effectively confines electrons and holes within a 1D geometry that physically enhances their proximity. For example, 1D hydrogen model with a bare Coulomb interaction will give a bound state with infinite binding energy. This has been noticed in understanding enhanced excitonic effects of free-standing CNTs (Spataru et al. 2004a).

When both the joint density of states (JDOS) of single particles and the DOS of excitons are plotted in Fig. 4c, the excitonic effect shifts the DOS by about 0.3 eV, which is different from the DOS in bulk silicon (Fig. 3b). Compared with the enhanced *e-h* binding energy of the exciton between the VBM and CBM, the small shift of DOS indicates that, despite enhanced *e-h* interactions, most excitonic states do not have a large *e-h* binding energy. From the substantial change of optical absorption spectrum shown in Fig. 4a, b, one can infer that the optical absorptions spectrum is dictated by a small number of excitons having both large *e-h* binding energies and significant optical oscillator strength.

SiNWs have lower symmetries than bulk Si. In contrast to the isotropic optical absorption spectrum of bulk silicon, anisotropic optical spectra of SiNWs have been reported at the level of single-particle absorption (Zhao et al. 2004). A notable excitonic effect is that the *e-h* interactions can enhance the anisotropic response. As shown in Fig. 4d, the optical absorption spectra of the incident light with different polarization directions vary dramatically. The absorption spectrum for the incident light polarized vertically to the axial direction is strongly suppressed (Bruneval et al. 2005; Yang et al. 2007a). This feature is known as a "local field effect." A strong depolarization field is developed vertically to the axial direction because the wavelength of the incident photon is much larger than the diameter of nanostructures (Spataru et al. 2004b). The *e-h* exchange interaction is essential to capture this local field effect, as shown in Fig. 4d. On the other hand, such a highly polarized optical

response enhances the use of SiNWs and many other 1D semiconductors for optical and radio applications (Gudiksen et al. 2002; Dresselhaus 2004).

Because of the **strong** *e-h* interaction, the quasiparticle band gap and the energy of the optical absorption edge (the so-called optical band gap) are significantly different. In bulk semiconductors with direct band gaps, the optical band gap is nearly the same as the quasiparticle (transport) band gap owing to the weak *e-h* binding energy. However, in this 1.2-nm SiNW, the quasiparticle band gap is about 3.2 eV while the "optical band gap" is about 2.0 eV (the energy of the lowest-energy exciton). Therefore, it is necessary to differentiate these two concepts before studying **the** electronic structure of nanomaterials.

The electronic self-energy correction increases the quasiparticle band gap while enhanced *e-h* interactions contribute to large exciton binding energies, which substantially redshift the optical absorption spectrum. Both effects are significant, but they affect the final optical spectrum in opposite directions, which results in some cancellation. Significant absorption peaks remain around the 3∼5-eV range for a SiNW with a diameter of as small as 1.2 nm (Yang et al. 2007a). This energy range of the significant optical absorption spectrum calculated by the GW-BSE approach is similar to the DFT/LDA result, although the profiles of optical spectra and the physics behind are very different.

2.4 Exciton Spectrum and Fine Structure

An analysis of the sequence of *e-h* binding energies forms the foundation for constructing simple models that are widely employed to understand excitonic effects in nanostructures. For example, the *e-h* binding energy of monolayer WS_2 can be extrapolated from the measured sequence of exciton peaks according to a non-hydrogenic energy series (Chernikov et al. 2014). The role of quantum confinement on exciton spectra and corresponding *e-h* models of 1D SiNWs (Zhang et al. 2011b) are of fundamental interest.

The most interesting excitons are those associated with transitions between the CBM and VBM around the Γ point in the SiNW. Since the CBM and VBM have quadratic dispersions, a hydrogenic model based on the effective mass approximation should accurately describe the sequence of band-edge excitons. First-principles BSE calculations have identified four bound excitons which are labeled as A_1, A_2, A_3, and A_4. They are associated with transitions around the band edges at the zone center. The binding energies of these excitons are plotted in Fig. 5a and also a fitting using a hydrogenic-like model (dashed line), namely **scaling** as $1/n^x$, where *n* is the main quantum number. The fitting exponent x turns out to be 1.8 (Yang et al. 2007a), which is surprisingly close to the value **2** of the hydrogenic model. This hydrogen-like exciton spectrum is helpful for the modeling of e-h pairs in other 1D semiconducting structures thus avoiding expensive first-principles simulations.

In Fig. 5b, excitonic wavefunctions of these four states are plotted in real space. According to Eq. (3), an exciton is a two-body excitation with six spatial degrees
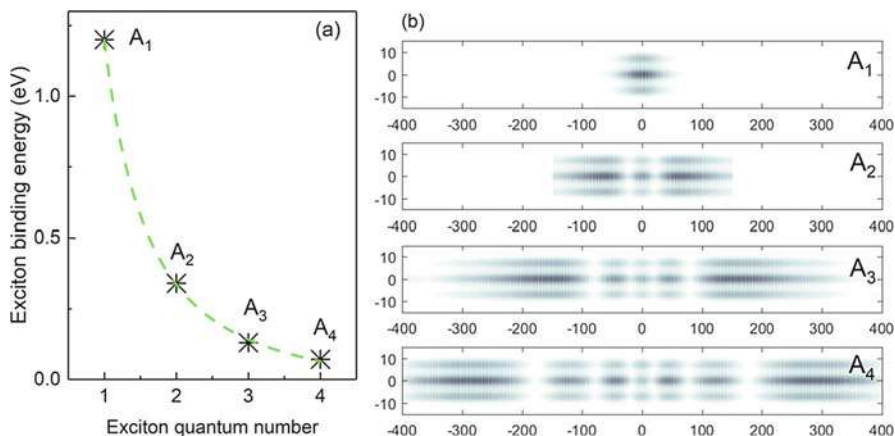
Fig. 5 (**a**) The *e-h* binding energy of excitons relative to the band edge around the Γ point. The green dash line is the fitting curve. (**b**) The real-space excitonic wavefunctions. The hole is fixed at the center of each plot. The length unit is the Bohr radius

of freedom. In Fig. 5b, the hole position is fixed at the center of the figures, and the corresponding electron wavefunction is plotted. As the quantum number increases, more nodal structures develop along the wire direction, in a way similar to the 1 s, 2 s, 3 s, and 4 s of hydrogenic states. This similarity confirms that the sequence of excitonic levels can be described by hydrogen-like models, despite the role of quantum confinement.

Enhanced singlet-triplet splitting: When an exciton **forms**, the spins of the electron and the hole may have various configurations. If the total spin is zero, the exciton is in the spin-singlet state; if the total spin is one, the exciton is in the spin-triplet state. The energy difference between singlet and triplet excitons is determined by the bare-exchange interaction between the electron and hole wavefunctions. The singlet-triplet (exchange) splitting dictates many important optical properties of nanomaterials (An et al. 2007; Srivastava et al. 2008). For instance, the spin-singlet exciton can be optically bright and exhibits a shorter radiative lifetime whereas the spin-triplet exciton is optically dark and exhibits a longer radiative lifetime. In SiNWs with strong quantum confinement, the *e-h* exchange interaction is increased by the enhanced overlap between electron and hole wavefunctions. Using a first-principles BSE approach, the enhancement of the singlet-triplet energy splitting is calculated to be as large as 90 meV (Palummo et al. 2010b) in thin SiNWs, while that of bulk silicon is less than 5 meV.

2.5   Extrinsic Factors and "Cancellation" Effect

Substrates and encapsulations used in fabricating nanostructures may change the role of excitonic effects in nanostructures. Studies have shown that these supporting

Beckman SP, Han J, Chelikowsky JR (2007) Quantum confinement effects in Ge [110] nanowires. Phys Rev B 74:165314

Blasé X, Rubio A, Louie SG, Cohen ML (1994) Stability and band gap constancy of boron nitride nanotubes. Europhys Lett 28:335

Blasé X, Rubio A, Louie SG, Cohen M (1995) Quasiparticle band structure of bulk hexagonal boron nitride and related systems. Phys Rev B 51:6868

Bruneval F, Botti S, Reining L (2005) Comment on "quantum confinement and electronic properties of silicon nanowires". Phys Rev Lett 94:219701

Bruno M, Palummo M, Marini A, Del Sole R, Ossicini S (2007) From Si nanowires to porous silicon: the role of excitonic effects. Phys Rev Lett 98:036807

Capaz RB, Spataru CD, Ismail-Beigi S, Louie SG (2006) Diameter and chirality dependence of exciton properties in carbon nanotubes. Phys Rev B 74:121401

Chernikov A, Berkelbach TC, Hill HM, Rigosi A, Li Y, Aslan OB, Reichman DR, Hybertsen MS, Heinz TF (2014) Exciton binding energy and nonhydrogenic Rydberg series in monolayer WS_2. Phys Rev Lett 113:076802

Cohen ML, Louie SG (2016) Fundamentals of condensed matter physics. Cambridge University Press, Cambridge

Cui Y, Lieber CM (2001) Functional nanoscale electronic devices assembled using silicon nanowire building blocks. Science 291:851–853

Dalpian GM, Chelikowsky JR (2009) Self-purification in semiconductor nanocrystals. Phys Rev Lett 96:226802

Delerue C, Lannoo M, Allan G (2000) Excitonic and quasiparticle gaps in Si nanocrystals. Phys Rev Lett 84:2457

Deslippe J, Samsonidze G, Strubbe DA, Jain M, Cohen ML, Louie SG, Berkeley GW (2012) A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures. Comput Phys Commun 183:1269

Deslippe J, Samsonidze G, Jain M, Cohen ML, Louie SG (2013) Coulomb-hole summations and energies for GW calculations with limited number of empty orbitals: a modified static remainder approach. Phys Rev B 87:165124

Dresselhaus M (2004) Applied physics: nanotube antennas. Nature 432:959–960

Fetter A, Walecka JD (1971) Quantum theory of many particle systems. McGraw-Hill, San Francisco, p 538

Gao S, Liang Y, Spataru CD, Yang L (2016) Dynamical excitonic effects in doped two-dimensional semiconductors. Nano Lett 16:5568–5573

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Corso AD, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502

Gregg BA (2003) Excitonic Solar cells. J Phys Chem B 107:4688–4698

Gudiksen MS, Lauhon LJ, Wang J, Smith DC, Lieber CM (2002) Growth of nanowire superlattice structures for nanoscale photonics and electronics. Nature 415:617–620

Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139:A796–A823

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864

Homles JD, Johnston KP, Doty RC, Korgel BA (2000) Control of thickness and orientation of solution-grown silicon nanowires. Science 287:1471–1473

Hybertsen MS, Louie SG (1986) Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. Phys Rev B 34:5390–5413

Ismail-Beig S (2006) Truncation of periodic image interactions for confined systems. Phys Rev B 73:233103

Katz D, Wizansky T, Millo O, Rothenberg E, Mokari T, Banin U (2002) Size-dependent tunneling and optical spectroscopy of CdSe quantum rods. Phys Rev Lett 89:086801

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133

Lautenschlager P, Garriga M, Via L, Cardona M (1987) Temperature dependence of the dielectric function and interband critical points in silicon. Phys Rev B 36:4821

Liang Y, Yang L (2015) Carrier plasmon induced nonlinear band gap renormalization in two-dimensional semiconductors. Phys Rev Lett 114:063001

Louie SG, Rubio A (2005) Quasiparticle and optical properties of solids and nanostructures: the GW-BSE approach. In: Handbook of materials modeling. Springer, Dordrecht/Berlin/Heidelberg/New York, pp 215–240

Ma DDD, Lee CS, Au FCK, Tong SY, Lee ST (2003) Small-diameter silicon nanowire surfaces. Science 299:1874–1877

Morales AM, Lieber CM (1998) A laser ablation method for the synthesis of crystalline semiconductor nanowires. Science 279:208–211

Nduwimana A, Wang XQ (2014) Tunable electronic properties of silicon nanowires under strain and electric bias. AIP Adv 4:077122

Neaton JB, Hybertsen MS, Louie SG (2006) Renormalization of molecular electronic levels at metal-molecule interfaces. Phys Rev Lett 97:216405

Nolan M, O'Callaghan S, Fagas G, Greer JC (2007) Silicon nanowire band gap modification. Nano Lett 7:34–38

Oğüt S, Chelikowsky JR, Louie SG (1997) Quantum confinement and optical gaps in Si nanocrystals. Phys Rev Lett 79:1770

Onida G, Reining L, Godby RW, Del Sole R, Andreoni W (1995) Ab initio calculations of the quasiparticle and absorption spectra of clusters: the sodium tetramer. Phys Rev Lett 75:818

Palummo M, Amato M, Ossicini A (2010a) Ab initio optoelectronic properties of SiGe nanowires: role of many-body effects. Phys Rev B 82:073305

Palummo M, Iori F, Del Sole R, Ossicini S (2010b) Giant excitonic exchange splitting in Si nanowires: first-principles calculations. Phys Rev B 81:121303(R)

Park C-H, Spataru CD, Louie SG (2006) Excitons and many-electron effects in the optical response of single-walled boron nitride nanotubes. Phys Rev Lett 96:126105

Peelaers H, Partoens B, Giantomassi M, Rangel T, Goossens E, Rignanese GM, Gonze X, Peeters FM (2011) Convergence of quasiparticle band structures of Si and Ge nanowires in the GW approximation and the validity of scissor shifts. Phys Rev B 83:045306

Philipp HP (1972) Influence of oxide layers on the determination of the optical properties of silicon. J Appl Phys 43:2836

Ping Y, Rocca D, Lu D, Galli G (2012) Ab initio calculations of absorption spectra of semiconducting nanowires within many-body perturbation theory. Phys Rev B 85:035316

Qiu DY, Felipe H, Louie SG (2013) Optical spectrum of MoS_2: many-body effects and diversity of exciton states. Phys Rev Lett 111:216805

Rohlfing M, Louie SG (1998) Excitonic effects and the optical absorption spectrum of hydrogenated Si clusters. Phys Rev Lett 80:3320

Rohlfing M, Louie SG (1999) Excitons and optical spectrum of the Si (111)−(2×1) surface. Phys Rev Lett 83:856

Rohlfing M, Louie SG (2000) Electron-hole excitations and optical spectra from first principles. Phys Rev B 62:4927

Rozzi CA, Varsano D, Marini A, Gross EKU, Rubio A (2006) Exact Coulomb cutoff technique for supercell calculations. Phys Rev B 73:205119

Rubio A, Corkil JL, Cohen M, Shirley EL, Louie SG (1993) Quasiparticle band structure of AlN and GaN. Phys Rev B 48:11810

Shih B-C, Xue Y, Zhang P, Cohen ML, Louie SG (2010) Quasiparticle band gap of ZnO: high accuracy from the conventional G_0W_0 approach. Phys Rev Lett 105:146401

Shirley EL, Louie SG (1993) Electron excitations in solid C 60: energy gap, band dispersions, and effects of orientational disorder. Phys Rev Lett 71:133

Sirbuly DJ, Law M, Yan H, Yang P (2005) Semiconductor nanowires for subwavelength photonics integration. J Phys Chem B 109:15190–15213

Son Y-W, Cohen ML, Louie SG (2006) Half-metallic graphene nanoribbons. Nature 444: 347–349

Spataru CD, Léonard F (2010) Tunable band gaps and excitons in doped semiconducting carbon nanotubes made possible by acoustic plasmons. Phys Rev Lett 104:177402

Spataru CD, Ismail-Beigi S, Benedict LX, Louie SG (2004a) Excitonic effects and optical spectra of single-walled carbon nanotubes. Phys Rev Lett 92:077402

Spataru CD, Ismail-Beigi S, Benedict LX, Louie SG (2004b) Quasiparticle energies, excitonic effects and optical absorption spectra of small-diameter single-walled carbon nanotubes. Appl Phys A Mater Sci Process 78:1129

Srivastava A, Htoon H, Klimov VI, Kono J (2008) Direct observation of dark excitons in individual carbon nanotubes: inhomogeneity in the exchange splitting. Phys Rev Lett 101:087402

Tran V, Soklaski R, Liang Y, Yang L (2014) Layer-controlled band gap and anisotropic excitons in few-layer black phosphorus. Phys Rev B 89:235319

Ugeda MM, Bradley AJ, Shi SF, da Jornada FH, Zhang Y, Qiu DY, Ruan W, Mo SK, Hussain Z, Shen ZX, Wang F, Louie SG, Crommie MF (2014) Giant bandgap renormalization and excitonic effects in a monolayer transition metal dichalcogenide semiconductor. Nat Mater 13: 1091–1095

Vasiliev I, Ogut S, Chelikowsky JR (2002) First-principles density-functional calculations for optical spectra of clusters and nanocrystals. Phys Rev B 65: 115416

Wu Z, Neaton JB, Grossman JC (2008) Quantum confinement and electronic properties of tapered silicon nanowires. Phys Rev Lett 100:246804

Wu Z, Neaton JB, Grossman JC (2009) Charge separation via strain in silicon nanowires. Nano Lett 9:2418–2422

Yan JA, Yang L, Chou MY (2007) Size and orientation dependence in the electronic properties of silicon nanowires. Phys Rev B 76:115319

Yang L, Chou MY (2011) Lattice vibrational modes and their frequency shifts in semiconductor nanowires. Nano Lett 11:2618–2621

Yang L, Spataru CD, Louie SG, Chou MY (2007a) Enhanced electron-hole interaction and optical absorption in a silicon nanowire. Phys Rev B 75:201304(R)

Yang L, Cohen ML, Louie SG (2007b) Excitonic effects in the optical spectra of graphene nanoribbons. Nano Lett 7:3112–3115

Yang L, Park C-H, Son Y-W, Cohen ML, Louie SG (2007c) Quasiparticle energies and band gaps in graphene nanoribbons. Phys Rev Lett 99:186801

Yang L, Deslippe J, Park C-H, Cohen ML, Louie SG (2009) Excitonic effects on the optical response of graphene and bilayer graphene. Phys Rev Lett 103:186802

Yu H, Li J, Loomis RA, Wang LW, Buhro WE (2003) Two- versus three-dimensional quantum confinement in indium phosphide wires and dots. Nat Mater 2:517–520

Zhang RQ, Liu XM, Wen Z, Jiang Q (2011a) Prediction of silicon nanowires as photocatalysts for water splitting: band structures calculated using density functional theory. J Phys Chem C 115:3425–3428

Zhang L, Luo JW, Franceschetti A, Zunger A (2011b) Excitons and excitonic fine structures in Si nanowires: prediction of an electronic state crossover with diameter changes. Phys Rev B 84:075404

Zhao X, Wei CM, Yang L, Chou MY (2004) Quantum confinement and electronic properties of silicon nanowires. Phys Rev Lett 92:236805

Interlayer Interactions in Low-Dimensional Layered Hetero-structures: Modeling and Applications

29

Oded Hod

Contents

Abstract

The field of nanoscale material science has experienced a true revolution over the past 30 years with the discovery of the quasi-zero- and one-dimensional cage-like structures of fullerenes and nanotubes. The successful isolation of graphene about a century ago has further triggered an avalanche of studies unraveling its unique physical and chemical properties. This, in turn, has led to numerous breakthroughs of basic science nature as well as diverse potential technological applications in the fields of nanoscale electronics, flexible displays, solar cells, DNA sequencing, chemical sensing, composite materials, solid lubrication, and many others.

Following extensive studies of the properties of graphene, much attention was recently paid to other members of the two-dimensional (2D) hexagonal layered materials family including hexagonal boron nitride (*h*-BN) and several transition metal dichalcogenides. Interestingly, stacking individual layers of these materials to form homogeneous and heterogeneous structures results in unique

O. Hod (✉)

Department of Physical Chemistry, School of Chemistry, The Raymond and Beverly Sackler Faculty of Exact Sciences and The Sackler Center for Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv, Israel
e-mail: odedhod@tau.ac.il

physical characteristics that depend on the specific chemical composition of the system and can be tuned via the application of external perturbations. This opens numerous opportunities for combinatorial materials design with versatile structure-function relations. Understanding what determines the properties of such complex structures and how to control them remains a challenge yet to be met in order for these materials to fulfill their full potential.

Theory and computation may offer a valuable microscopic perspective toward achieving this goal. The reduced-dimensions of materials at the nanoscale allow for a unique interplay between theory, computation, and experiment. Here, accurate fully atomistic simulations can complement experiments both in analyzing and rationalizing experimental results and in providing reliable predictions that can minimize the need for demanding trial and error experimental efforts.

Such theoretical treatments of 2D materials require special attention to their anisotropic nature characterized by a strong in-plane covalent bonding network and weaker interlayer interactions. While state-of-the-art quantum mechanical approaches can simultaneously describe these interactions with high accuracy, their computational demand often limits their applicability to relatively small systems. An efficient alternative can be provided by carefully tailored classical force-fields. When appropriately parameterized against experimental results or high accuracy calculations of small model systems, these can provide a reliable description of the structural, mechanical, tribological, and heat transport properties of realistic nanoscale systems with atomic scale resolution. Such force-fields that provide a proper description of intralayer interactions in a variety of 2D materials have been developed over the years and are accessible via standard molecular dynamics simulation codes. Surprisingly, despite the great scientific interest in 2D layered materials, complementary interlayer force-fields that can accurately capture both their binding and sliding energy landscapes are currently available for a very limited set of systems.

In the present chapter, I provide a brief review of recent developments of reliable, efficient, and transferable anisotropic interlayer force-fields for homogeneous and heterogeneous low-dimensional hexagonal layered materials. To demonstrate the performance of these methods, a few applications to the study of the structural and tribological properties of quasi-one- and quasi-two dimensional layered structures will be presented.

1   Introduction

The successful isolation of graphene, a one atom thick hexagonal carbon layer, by Novoselov and Geim in 2004 marked the opening of a new era in materials science (Novoselov et al. 2004; Geim and Novoselov 2007; Li and Kaner 2008; Geim 2009; Rao et al. 2009). The study of its unique physical properties including exceptional mechanical strength (Lee et al. 2008), superlubric tribological characteristics (Dienwiebel et al. 2004; Koren et al. 2015), remarkable room temperature electron mobilities (Novoselov et al. 2005a; Morozov et al. 2008), high thermal

conductivity (Balandin et al. 2008; Cai et al. 2010; Xu et al. 2014), quasi-relativistic electronic structure (Geim and Novoselov 2007), and controllable optical properties is flourishing even more than a decade later. Nevertheless, despite this wealth of outstanding material characteristics, the semi-metallic nature of graphene limits its applicability in switchable nanoscale electronic devices. This has triggered the scientific community to search for graphene alternatives within the diverse family of layered inorganic materials (Novoselov et al. 2005b, 2016; Deng et al. 2016).

Currently, two-dimensional (2D) inorganic materials are at the forefront of scientific research with an ever-growing number of chemical compositions including hexagonal boron nitride (*h*-BN), germanene, silicene, phosphorene, transition metal dichalcogenides, such as molybdenum and tungsten disulfide (*2H*-MoS_2 and *2H*-WS_2) and diselenide (*2H*-$MoSe_2$ and *2H*-WSe_2), among others (Heine 2015; Li and Zhu 2015). Similar to graphene, these materials exhibit diverse physical properties and appear in many structural forms ranging from planar configurations to nanotubes, scrolls, onions, and cones.

The recently demonstrated ability to stack such 2D materials in homogeneous as well as heterogeneous few-layered structures (Novoselov et al. 2016; Deng et al. 2016; Dean et al. 2010; Geim and Grigorieva 2013; Wang et al. 2014; Das et al. 2015) opens a vast combinatorial material space enabling the design of novel structures exhibiting desired physical properties. Furthermore, electromechanical manipulation of such structures can be used to gain control over their physical properties – a basic requirement for their future technological application (Koren et al. 2015, 2016; Leven et al. 2013).

Here, theory and computation play a prominent role in efficiently scanning the vast material space, identifying potential candidates that are expected to exhibit the desired physical properties, and studying their sensitivity towards external perturbations. This can focus the demanding experimental efforts on a selected group of predesigned structures and increase the probability of obtaining the required 2D material functionality. Moreover, a synergic cooperation between theory, computation, and experiment, where experimental data are used to refine theoretical models and computational algorithms that, in turn, provide more accurate predictions, can formulate a feedback cycle that will considerably enhance the efficiency of 2D materials design.

An accurate theoretical account of the physical and chemical properties of layered (and other) materials requires a full quantum mechanical treatment of their electronic structure. Computational methodologies that may provide such ultimate accuracy include wave function-based configuration interaction or coupled clusters methods as well as quantum Monte-Carlo procedures (Nightingale and Umrigar 1999; Pang 2016). These methods, however, are quite unfavorable in terms of the scalability of their computational burden with system size and are hence limited to very small molecular structures. Green's function-based methods that can efficiently sum high-order terms in a perturbative expansion scale better with system size and hence can treat somewhat larger systems (Sakai et al. 2015). Density functional theory (DFT) (Hohenberg and Kohn 1964; Kohn and Sham 1965) is currently probably the most popular method to study the physical properties of large systems

(Sakai et al. 2015). While being exact in principle, its practical approximations provide a good balance between computational efficiency and accuracy. Nowadays, DFT calculations of material properties involving hundreds of atoms are being routinely performed in many research groups (Barone et al. 2011). If one desires to treat even larger systems on a fully atomistic and quantum mechanical basis, accuracy is often further compromised to gain computational efficiency. Here, methods like the variational Hartree-Fock formalism (Szabo and Ostlund 1989), semi-empirical approaches (Pople 1970), density functional based tight-binding approximations (Porezag et al. 1995; Kwon and Tománek 2000; Miró et al. 2013; Zahid et al. 2013), or empirical tight-binding (Szabados et al. 2006; Carlson and Dumitrică 2007) can be used to provide approximate descriptions of 2D material properties.

Often, however, a fully quantum mechanical treatment is not required and a much simplified classical picture is sufficient to efficiently and accurately describe many material properties. Specifically, when considering static structural and mechanical properties of various materials (Chen et al. 2013; Butz et al. 2014; van Wijk et al. 2015) as well as their dynamic tribological (Guo et al. 2007; Guerra et al. 2010; Reguzzoni et al. 2012a; Vanossi et al. 2013; van Wijk et al. 2013; Xu et al. 2013; Ze 2014; Kang and Lee 2014) and heat transport characteristics (Jiang 2014a), Newtonian dynamics with carefully designed force-fields (FF) can provide valuable insights (Frenkel and Smit 2002; Allen and Tildesley 2009; Rapaport 2009). Obviously, this results in considerable computational gain enabling the treatment of extremely large atomistic models including tens of thousands of atoms for relatively long timescales. Furthermore, it allows for devising multiscale approaches, where the information gained from fully atomistic calculations serves to design coarse-grained and continuum models. Nevertheless, this comes at the expense of neglecting explicit quantum mechanical effects and even more importantly the need to design system specific interatomic potentials with limited transferability.

With this respect, the formulation of dedicated force-fields for 2D materials and their layered constructs has to account for their inherent structural anisotropy. The individual layers can be viewed as a network of covalently bonded atoms organized in a periodic hexagonal lattice structure. Upon stacking, each layer interacts with its neighboring counterparts via long-range dispersion and electrostatic interactions. Hence, a strategy involving a separate treatment of intralayer and interlayer interactions is often invoked.

2   Intralayer Interactions

The intralayer force-field term usually involves six main ingredients including (see Fig. 1): (http://cbio.bmt.tue.nl/pumma/index.php/Theory/Potentials; Rappe et al. 1992) (i) bonded two-body interactions describing the bond stretching and compression energy; (ii) bonded three-body interactions accounting for angle bending penalties; bonded four-body interactions depicting (iii) improper and (iv) dihedral
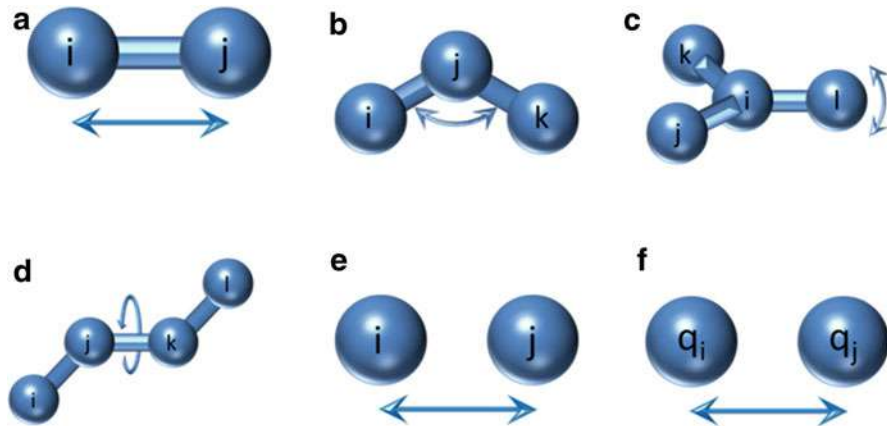
Fig. 1 Illustration of several intralayer force-field term contributions: (**a**) Bond stretching and compression; (**b**) angle bending; (**c**) improper angle deformation; (**d**) torsional angle deformation; (**e**) van der Waals interactions; (**f**) electrostatic interactions

torsional deformations; (v) nonbonded van der Waals (vdW) interactions; and (vi) nonbonded electrostatic interactions.

The first two components are often represented by simple harmonic springs of the form $V_{ij}^{(i)}\left(r_{ij}\right) = k_{ij}\left(r_{ij} - r_{ij}^0\right)^2$ and $V_{ijk}^{(ii)}\left(\theta_{ijk}\right) = k_{ijk}\left[\cos\left(\theta_{ijk}\right) - \cos\left(\theta_{ijk}^0\right)\right]^2$. Here, r_{ij} is the distance between atomic center i and j, θ_{ijk} is the angle defined by atomic centers i, j and k (see Fig. 1a, b), r_{ij}^0 and θ_{ijk}^0 are the equilibrium ij pair interatomic distance and ijk triad angle, respectively, and k_{ij} and k_{ijk} are the corresponding force constants. Improper torsion is often also described by a harmonic potential of the form $V_{ijkl}^{(iii)}\left(\delta_{ijkl}\right) = k_{ijkl}^{(iii)}\left(\delta_{ijkl} - \delta_{ijkl}^0\right)^2$, δ_{ijkl} being the improper angle (see Fig. 1c) and $k_{ijkl}^{(iii)}$ and δ_{ijkl}^0 being the corresponding force constant and equilibrium improper angle, respectively. Dihedral angle torsion can be depicted by a periodic potential of the form $V_{ijkl}^{(iv)}\left(\phi_{ijkl}\right) = k_{ijkl}^{(iv)}\left[1 + \cos\left(\eta_{ijkl}\phi_{ijkl} - \phi_{ijkl}^0\right)\right]$, where ϕ_{ijkl} is the dihedral angle (see Fig. 1d), $k_{ijkl}^{(iv)}$ is the corresponding force constant, ϕ_{ijkl}^0 sets the equilibrium dihedral angle, and η_{ijkl} is the multiplicity factor signifying the number of minima that this potential term possesses. The nonbonded vdW term is often described by the two-body Lennard-Jones (LJ) potential of the form $V_{ij}^{(v)}\left(r_{ij}\right) = 4\epsilon_{ij}\left[\left(\sigma_{ij}/r_{ij}\right)^{12} - \left(\sigma_{ij}/r_{ij}\right)^6\right]$ where ϵ_{ij} and $\sqrt[6]{2}\sigma_{ij}$ are the binding energy and equilibrium distance of the ij atomic pair, respectively (see Fig. 1e). Note that this expression accounts not only for the long-range attractive vdW interactions but also for short-range repulsions. Finally, interatomic monopolar electrostatic interactions are calculated using Coulombs law $V_{ij}^{(vi)}\left(r_{ij}\right) = q_i q_j/\left(4\pi\epsilon_0 r_{ij}\right)$, where ϵ_0 is the vacuum permittivity and q_k is the effective partial charge of atom k (see Fig. 1f).

It should be noted that while the expressions described above are commonly used to describe the intralayer interactions in layered materials they are, by no means, the only ones available. For example, if a reactive force-field is required, the harmonic bonded two-body terms can be replaced by an anharmonic potential that allows for bond breaking and formation. Furthermore, the force-field parameters can be varied according to the local chemical environment, thus providing flexibility to describe different bond-orders and accounting for various chemical binding schemes.

The various force-field parameters have to be determined separately for each material. The fitting procedure can be performed against experimentally measured material properties, such as the Young modulus, the shear modulus, lattice constants, phonon spectra, and heats of formation. Alternatively, they can be calibrated to fit results of higher accuracy methods such as state-of-the-art DFT calculations. Dedicated intralayer force-field parameterizations are available for a variety of layered materials including graphene and its derivatives (Al-Jishi and Dresselhaus 1982; Tersoff 1988; Brenner 1988, 1990; Stuart et al. 2000; Brenner et al. 2002; Los and Fasolino 2003; Ghiringhelli et al. 2005a, 2008; Los et al. 2005; Perebeinos and Tersoff 2009; Lindsay and Broido 2010; Jiang 2015; O'Connor et al. 2015), h-BN (Sevik et al. 2011), and $2H$-MoS$_2$ (Jiang 2015; Wakabayashi et al. 1975; Jiménez Sandoval et al. 1991; Brunier et al. 1992; Morita et al. 2008; Liang et al. 2009; Varshney et al. 2010; Jiang et al. 2013; Nicolini and Polcar 2016), many of which are implemented in open source or commercial software packages. These have been successfully used to describe a variety of material properties including the structural and mechanical behavior (Tersoff 1992; Lu 1997; Hertel et al. 1998; Fasolino et al. 2007; Bets and Yakobson 2009; Costamagna et al. 2012; Singh et al. 2013, 2015; Bucholz and Sinnott 2013; Jin-Wu et al. 2013; Jiang 2014b; Jiang and Park 2014), thermal transport characteristics (Varshney et al. 2010; Che et al. 2000; Mohamed and Deepak 2001; González Noya et al. 2004; Yao et al. 2005; Lindsay and Broido 2011), crack propagation (Xiaonan et al. 2015), and melting dynamics of 2D monolayers (Singh et al. 2015).

3   Interlayer Interactions

The interlayer force-field term usually involves three main contributions: long-range attractive vdW interactions, short-range Pauli repulsions, and electrostatic interactions. The latter, which is required when atoms in different layers have a sizable partial charge, can often be described by a simple Coulombic term similar to the $V_{ij}^{(vi)}$ intralayer expression given above (see Fig. 1f). To describe the attractive vdW and repulsive Pauli interactions, the LJ (or Morse (Los and Fasolino 2003; Los et al. 2005; Ghiringhelli et al. 2005b; Karssemeijer and Fasolino 2011)) potential is often chosen similar to the intralayer $V_{ij}^{(v)}$ expression (see Fig. 1e) (Nicolini and Polcar 2016; Lebedeva et al. 2011a; Shibuta and Elliott 2011; Ye et al. 2012, 2015; Jayasena et al. 2013; Liu et al. 2015; Kushima et al. 2015). With appropriate parameterization, this expression can provide a good description of the interlayer
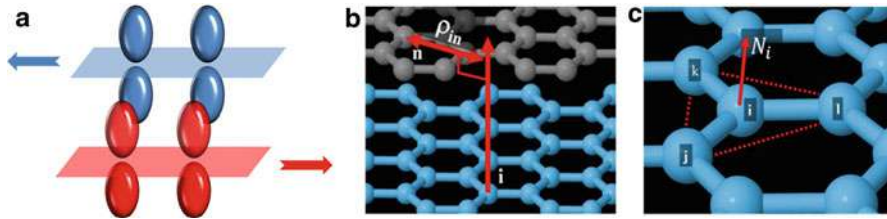
Fig. 2 Anisotropic Pauli repulsion term: (**a**) illustration of the anisotropic p_z orbital overlap variations during interlayer sliding; (**b**) definition of the lateral interatomic distance; (**c**) definition of the local surface normal. (Reprinted with permission from Oz et al. (2016). Copyright (2016) American Chemical Society)

binding energy curve (Szabados et al. 2006; Jiang 2015; Brunier et al. 1992; Varshney et al. 2010; Jiang et al. 2013; Lu 1997; Bucholz and Sinnott 2013; Jiang and Park 2014; Singh et al. 2015; Girifalco and Lad 1956; Green et al. 1974; Weiss and Phillips 1976; Kuzuba et al. 1985; Lu et al. 1992; Tersoff and Ruoff 1994; Kwon and Tománek 1998; Buldum and Lu 1999; Girifalco et al. 2000; Girifalco and Hodak 2002; Qian et al. 2003; Michel and Verberck 2011; Tolga et al. 2016). However, using the same parameters to study relative interlayer sliding fails to reproduce the corrugated sliding energy landscape and results in too shallow energy barriers (Carlson and Dumitrică 2007; Stuart et al. 2000; Palser 1999; Kolmogorov and Crespi 2000, 2005; Shtogun and Woods 2010; Lebedeva et al. 2011b; Reguzzoni et al. 2012b; Neek-Amal and Peeters 2014; Strutyński et al. 2014; Korhonen and Koskinen 2015; Jiang and Park 2015). Hence, the ability of LJ type potentials to reliably describe many structural, mechanical, and tribological properties of layered materials may be considerably hindered.

The origin of this deficiency in the LJ potential was identified by Kolmogorov and Crespi (KC) to be the use of isotropic expressions that depend only on interatomic distances to describe an anisotropic material property (Kolmogorov and Crespi 2000, 2005; Lebedeva et al. 2011b). Focusing on graphene, KC realized that when two layers slide upon each other, electron clouds associated with the p_z orbitals around carbon atoms residing on adjacent layers overlap as the atoms pass each other (see Fig. 2a). This, in turn, induces Pauli repulsion variations resulting in a corrugated sliding energy landscape. Since the electron density overlap depends on the lateral distance between the crossing atoms, an interlayer pair-potential of the following form was suggested (Kolmogorov and Crespi 2005):

$$V^{KC}(r_{in}, \rho_{in}) = e^{-\lambda(r_{in}-z_0)}\left[C + f(\rho_{in}) + f(\rho_{ni})\right] - A(r_{in}/z_0)^{-6}. \qquad (1)$$

This expression includes a LJ type long-range attractive term scaled by $A = 10.238$; an isotropic Morse-like short-range exponential repulsive term with the equilibrium interlayer distance $z_0 = 3.34$ Å, exponent $\lambda = 3.629$ Å^{-1}, and scaling factor $C = 3.030$; and an anisotropic short-range Gaussian repulsive term of the form $f(\rho) = e^{-(\rho/\delta)^2} \Sigma_n c_{2n}(\rho/\delta)^{2n}$ that depends on the lateral distance, ρ,

between each pair of atoms on adjacent layers scaled by $\delta = 0.578$ Å. The nonzero coefficients of the polynomial multiplying the Gaussian are given by $c_0 = 15.71$, $c_2 = 12.29$, and $c_4 = 4.933$. The lateral distance, ρ_{in}, is defined as the shortest distance between atom n on one layer and the surface normal at atom i residing on an adjacent layer (see Fig. 2b). For a hexagonal lattice, the simplest definition of the local surface normal at a given atomic position is the unit vector perpendicular to the triangle formed by its three nearest neighbors (see Fig. 2c). Since for curved systems generally $\rho_{in} \neq \rho_{ni}$ the pair potential is symmetrized to fulfill Newton's third law. With the inclusion of the anisotropic repulsive term, the KC potential was shown to simultaneously provide a good description of both interlayer binding and sliding energy landscapes (see Fig. 3) (Kolmogorov and Crespi 2000, 2005; Reguzzoni et al. 2012b).

Despite this notable success and the ever growing interest in 2D materials, the KC approach (and related methods (Lebedeva et al. 2011b, 2012; Popov et al. 2011)) has not been extended to describe other layered materials for more than a decade since its inception. Recently, however, the transferability of the method was demonstrated where a KC type potential was applied to the graphene/h-BN hetero-structure (van Wijk et al. 2014; Woods et al. 2016). In this study, the FF parameters have been set to reproduce experimentally observed moiré superstructures arising from the 1.8% intralayer lattice constant mismatch between graphene and h-BN. This indicated the ability of the KC force-field to describe the structural properties of homogeneous and heterogeneous junctions of 2D materials.

To harness the full predictive capabilities of KC-type FFs to describe the interlayer interactions in a variety homogeneous and heterogeneous layered materials structures, a slightly modified functional form, termed interlayer potential (ILP), was recently suggested. The ILP accounts for monopolar electrostatic interactions and allows for simple parameterization against state-of-the-art first-principles calculations based on pair-wise (Tkatchenko and Scheffler 2009; Marom et al. 2011) and many-body dispersion corrected (Tkatchenko et al. 2012; Ambrosetti et al. 2014) screened-hybrid (Heyd et al. 2003; Heyd and Scuseria 2004a, b) density functional theory calculations (Leven et al. 2014, 2016a; Maaravi et al. 2017; https://pubs.acs.org/doi/10.1021/acs.nanolett.8b02848).

The ILP consists of a modified KC **short-range anisotropic repulsive term** of the form:

$$V_{Rep}(r_{in}, \rho_{in}) = \text{Tap}(r_{in}) \times e^{\alpha_{in}\left(1 - \frac{r_{in}}{\beta_{in}}\right)} \left[\varepsilon_{in} + C_{in} \left(e^{-\left(\frac{\rho_{in}}{\gamma_{in}}\right)^2} + e^{-\left(\frac{\rho_{ni}}{\gamma_{in}}\right)^2} \right) \right], \qquad (2)$$

where ε_{in} and C_{in} are constants setting the energy scales associated with the isotropic and anisotropic repulsion, respectively, β_{in} and γ_{in} set the corresponding interaction ranges, and α_{in} is a parameter defined to set the steepness of the isotropic repulsion term. Note that for computational efficiency, the 4th degree polynomial multiplying the anisotropic Gaussian repulsion term appearing in the original KC expression (see Eq. (1) above)
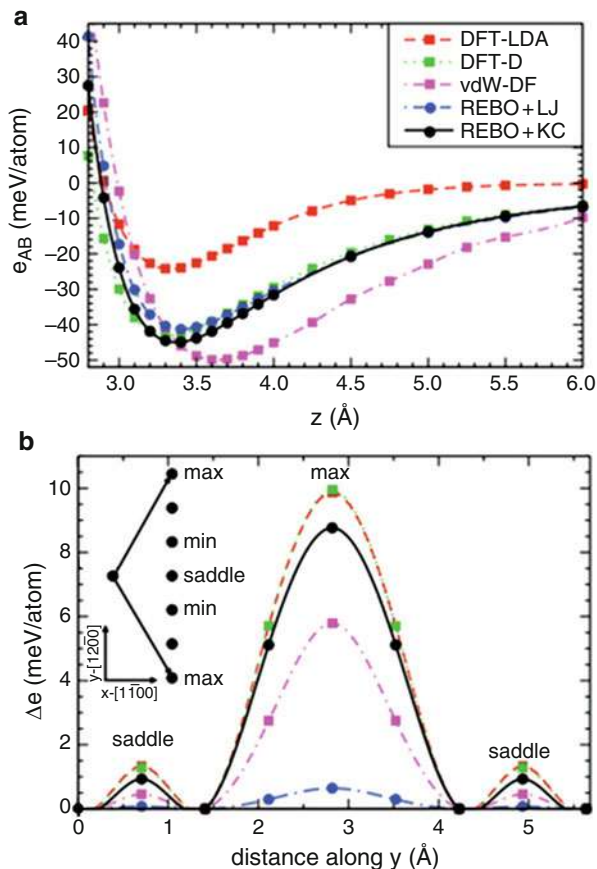
Fig. 3 Interlayer (**a**) binding and (**b**) sliding energy curves of bilayer graphene calculated using the LJ (full blue circles) and KC (full black circles) interlayer potentials compared to a reference dispersion corrected DFT calculation (full green squares). (Reprinted with permission from Reguzzoni et al. (2012b). Copyright (2012) by the American Physical Society)

is replaced by a constant factor. Furthermore, a taper function of the form $\text{Tap}(r_{in}) = 20(r_{in}/R_{cut})^7 - 70(r_{in}/R_{cut})^6 + 84(r_{in}/R_{cut})^5 - 35(r_{in}/R_{cut})^4 + 1$, which provides a continuous (up to third derivative) cutoff for interatomic separations exceeding R_{cut}, is used to damp the long-range contribution of the repulsive term. These modifications simplify the FF expressions and reduce the computational cost while providing a satisfactory description of the interlayer interactions (Leven et al. 2014, 2016a; Maaravi et al. 2017).

To treat the **long-range attractive interactions,** the Tkatchenko and Scheffler (TS) dispersion correction scheme to DFT is considered (Tkatchenko and Scheffler 2009). Within this approach, widely used semi-local or hybrid exchange-correlation density functional approximations that are known to provide an inadequate

description of long-range van der Waals (vdW) interactions are augmented by a term that decays asymptotically with the interatomic distance as r^{-6}. This term is damped at short distances to avoid double-counting of short-range correlation effects. In the ILP, the following pair-wise long-range attractive term is adopted:

$$V_{Att}\left(r_{in}\right) = \text{Tap}\left(r_{in}\right)\left\{-\left[1 + e^{-d\left[\left(r_{in}/\left(S_R \cdot r_{in}^{eff}\right)\right)-1\right]}\right]^{-1} \cdot \frac{C_{6,in}}{r_{in}^6}\right\}. \tag{3}$$

Here, r_{in}^{eff} is the sum of effective equilibrium vdW atomic radii of atom i and atom n residing on different layers, $C_{6,in}$ is the pair-wise dispersion coefficient of the two atoms in the solid-state environment, and d and S_R are unit-less parameters defining the steepness and onset of the short-range Fermi-Dirac type damping function. Similar to the repulsive term, long-range taper damping is implemented to reduce computational burden. This specific form is chosen as it allows the extraction of the various parameters directly from first-principles calculations avoiding the need for experimental data that is macroscopic by nature, very difficult to obtain, and mostly unavailable for many 2D layered materials and hetero-structures thereof.

In cases where atoms residing on the interacting layers bear sizable effective charges, electrostatic contributions should be taken into account. To this end, the formalism implemented in the ReaxFF scheme can be utilized (van Duin et al. 2001). Within this approach, a shielded Coulomb potential term of the form:

$$V_{Coul}\left(r_{in}\right) = \text{Tap}\left(r_{in}\right) \times \left[kq_iq_n/\sqrt[3]{r_{in}^3 + (1/\lambda_{in})^3}\right] \tag{4}$$

is used. Here, k is Coulomb's constant and λ_{in} is a shielding parameter introduced to eliminate the short-range singularity of the classical monopolar electrostatic interaction expression. This shielding takes effect in regions where Pauli repulsions between overlapping electron clouds dominate the interlayer potential and hence has minor influence on the results. When considering periodic systems, the long-range taper cutoff can be replaced by Ewald summation techniques (Ewald 1921; Toukmaji and Board 1996) in order to avoid conditionally convergent sums (Hod 2012a).

In many cases, the effective ionic charges, q_i and q_n, can be treated as constant values throughout the simulation. Nevertheless, in order to provide a general description, they can be dynamically evaluated using the electronegativity equalization method (EEM) (van Duin et al. 2001; Mortier et al. 1986; Njo et al. 1998; Bultinck et al. 2002). This method relies on a principle formulated by Sanderson stating that when a molecule or a solid is formed, the electronegativities of the constituent atoms equalize to give a global electronegativity of the entire system (Sanderson 1951, 1983). Hence, the electronegativity of a given atom within the molecular or solid environment (χ_i) is written in terms of the corresponding isolated atom electronegativity (Parr et al. 1978) $\left(\chi_i^0\right)$ and hardness (Parr and Pearson 1983; Sanderson 1976) $\left(\eta_i^0\right)$ as $\chi_i = \left(\chi_i^0 + \Delta\chi_i\right) + 2\left(\eta_i^0 + \Delta\eta_i\right)q_i +$

$\sum_{j\neq i} kq_j / \sqrt[3]{r_{ij}^3 + (1/\lambda_{ij})^3}$. Here, $\Delta\chi_i$ and $\Delta\eta_i$ represent the electronegativity and hardness variations due to the embedding molecular or solid environment and the last term represents the electrostatic potential induced by all other atoms in the system (Weismiller et al. 2010). The effective atomic charges can then be obtained by enforcing the guiding principle that within the molecular or solid environment all atomic electronegativities should equal the equilibrated molecular electronegativity $\chi_{i=1...N} = \chi_{eq}$. To this end, the following matrix equation is solved:

$$
\begin{pmatrix}
2\left(\eta_1^0 + \Delta\eta_1\right) & k/\sqrt[3]{r_{12}^3 + (1/\lambda_{12})^3} & \cdots & k/\sqrt[3]{r_{1N}^3 + (1/\lambda_{1N})^3} & -1 \\
k/\sqrt[3]{r_{21}^3 + (1/\lambda_{21})^3} & 2\left(\eta_2^0 + \Delta\eta_2\right) & \cdots & k/\sqrt[3]{r_{2N}^3 + (1/\lambda_{2N})^3} & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
k/\sqrt[3]{r_{N1}^3 + (1/\lambda_{N1})^3} & k/\sqrt[3]{r_{N2}^3 + (1/\lambda_{N2})^3} & \cdots & 2\left(\eta_N^0 + \Delta\eta_N\right) & -1 \\
1 & 1 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
q_1 \\ q_2 \\ \vdots \\ q_N \\ \chi_{eq}
\end{pmatrix}
= -
\begin{pmatrix}
\chi_1^0 + \Delta\chi_1 \\ \chi_2^0 + \Delta\chi_2 \\ \vdots \\ \chi_N^0 + \Delta\chi_N \\ -Q
\end{pmatrix},
$$

(5)

where the isolated atomic electronegativities and hardnesses, χ_i^0 and η_i^0, their corresponding molecular or solid environment variations $\Delta\chi_i$ and $\Delta\eta_i$, and the total charge, Q, are provided as input. The latter is dictated by the modeled system, whereas the former are used as fitting parameters.

While originally parameterized for homogeneous h-BN and graphene junctions (Leven et al. 2014; Maaravi et al. 2017), as well as for heterogeneous junctions thereof (Leven et al. 2016a), with appropriate parameterization the ILP should be transferable to a variety of layered materials such as different graphene allotropes including pentaheptite, haeckelite, dimerite, and octite (David et al. 2010); graphane; graphyne; germanene; silicene; stanene; phosphorene; and members of the transition metal dichalcogenides family such as molybdenum and tungsten disulfide and diselenide. As demonstrated in the next section, such an extension will make the ILP a versatile and powerful simulation tool for the study of the structural, mechanical, dynamic, and heat transport properties of homogeneous and heterogeneous layered material structures.

4   Applications

An illustrative demonstration of the performance of classical ILPs for studying interlayer interactions in layered materials is the case of the graphene/h-BN heterojunction. This heterojunction has gained great attention from the scientific community in recent years (Geim and Grigorieva 2013; Wang et al. 2014; Das et al. 2015; Leven et al. 2016a) due to its rich physical properties that include a controllable electronic band-gap (Chen et al. 2014; Fan et al. 2011; Jung et al. 2015; Yankowitz et al. 2012), ultrahigh electron mobility (Dean et al. 2010; Levendorf et al. 2012; Tang et al. 2013), the demonstration of Hofstadter's butterfly

phenomenon (Ponomarenko et al. 2013; Dean et al. 2013), and the manifestation of controllable hyperbolic meta-material characteristics (Dai et al. 2015).

Owing to the intrinsic lattice-constant mismatch between graphene and *h*-BN, the former tends to deform when placed on a multilayer *h*-BN substrate in order to adapt to the underlying crystal lattice (van Wijk et al. 2014; Yankowitz et al. 2012; Dean et al. 2013; Yang et al. 2013; Woods et al. 2014). The resulting moiré superstructures alter graphene's electronic and optical properties thus opening new opportunities for technological applications (Jung et al. 2015; Kumar et al. 2015). Naturally, modeling such superstructures from first-principles is a prohibitively demanding task. This, in turn, calls for the utilization of classical ILPs.

Figure 4 presents the fully relaxed structure of a free-standing graphene/*h*-BN bilayer supercell calculated using the graphene/*h*-BN ILP (Leven et al. 2016a). A clear moiré superstructure is obtained (see panel (d)), where regions of optimal interlayer registry (blue regions in panel (a)) and distance (blue regions in panel (b)) and intralayer bond lengths (red regions in panel (c)) are separated by sharp domain walls characterized by increased interlayer distance and compressed intralayer bond lengths in accordance with experimental findings (Yankowitz et al. 2012; Woods et al. 2014). Furthermore, due to the lack of rigid support, the calculation predicts that the entire structure should become highly corrugated. This latter observation was recently verified experimentally (Argentero et al. 2017).

Such moiré superstructures are expected to have considerable effect on the tribological properties of the heterojunctions. To investigate this, the graphene/*h*-BN ILP (Leven et al. 2016a) has been recently used to simulate the frictional
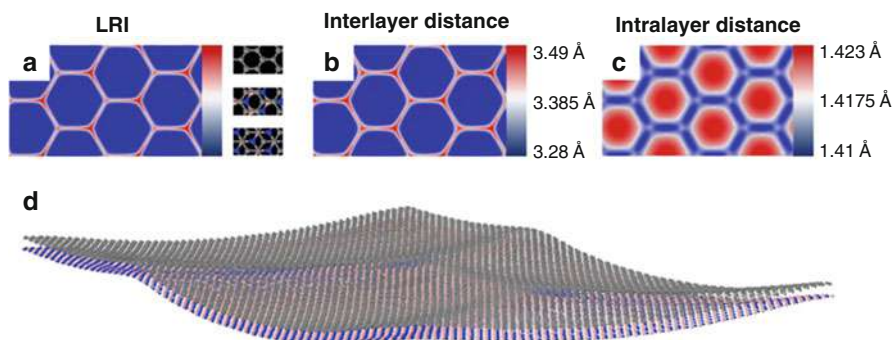
Fig. 4 Fully relaxed graphene/*h*-BN free-standing bilayer. (**a**) Local registry index map, where blue and red colors represent optimal and worst interlayer stacking, respectively. The calculation is performed using the registry index method that defines a purely geometric quantitative measure of interlattice commensurability (Leven et al. 2013, 2016a; Koren et al. 2016; Marom et al. 2010; Hod 2010, 2012b, 2013; Blumberg et al. 2012; Garel et al. 2012; Oz et al. 2016; Leven et al. 2016b); (**b**) interlayer distance map; (**c**) graphene intralayer bond-length map; and (**d**) overall structure of a fully relaxed, free-standing, graphene/*h*-BN bilayer. (Reprinted with permission from Leven et al. (2016a). Copyright (2016) American Chemical Society)

behavior of the heterojunction (Mandelli et al. 2017). It was found that for the aligned interface (where the lattice vectors of the two layers are parallel), with contact size below the characteristic lateral dimension of the elevated superstructure ridges, the junction behaves like its homogeneous counterparts with friction forces that grow linearly with the contact area. Superlubricity sets in due to the progressive appearance of moiré patterns resulting in a transition to collective stick-slip motion of the ridges that eventually turns into smooth soliton-like sliding with increasing contact size (see Fig. 5). For angularly misaligned contacts (where the lattice vectors of the two layers are rotated with respect to each other), incommensurability effects, also appearing in homogeneous junctions, are enhanced and the friction coefficients further drop by orders of magnitude and remain extremely low, even under external loads. These simulations indicated the potential of achieving robust superlubricity, independent of the relative interfacial orientation, and sustainable under external loads, in practical applications utilizing two-dimensional layered materials heterojunctions.

A similar phenomenon is found to occur in homogeneous multiwalled nanotubes, where the lattice mismatch results from the curvature difference between adjacent nanotube walls. Due to the frustrated geometry of the tube's circumference, this may result in a pronounced faceted superstructure, where the circular cross-section is deformed into a more stable polygonal one (Garel et al. 2012, 2014; Liu and Cowley 1994; Gogotsi et al. 2000; Zhang et al. 2003, 2005; Celik-Aktas et al. 2005a; Golberg et al. 2007a, b; Nigues et al. 2014). For narrow nanotubes, the excessive stain involved in the formation of sharp vertices is not compensated by the energy gain of forming the planar facets; hence, faceting is suppressed. However, above a certain nanotube diameter, typically of 12 nm (Garel et al. 2012), the standard circular cross section can be strongly distorted and faceting may appear (see Fig. 6).

Here, as well, modeling such systems from first-principles is impractical and classical ILPs serve as a viable alternative for studying their structural and mechanical properties. Hence, to decipher the mechanism that underlies facet formation and dictates their character, elaborate geometry relaxations of a wide set of double-walled carbon and boron nitride nanotubes were performed (Leven et al. 2016b). It was found that regardless of the nanotube identity (namely, diameter, and chemical composition), chiral angle matching between adjacent layers is a necessary
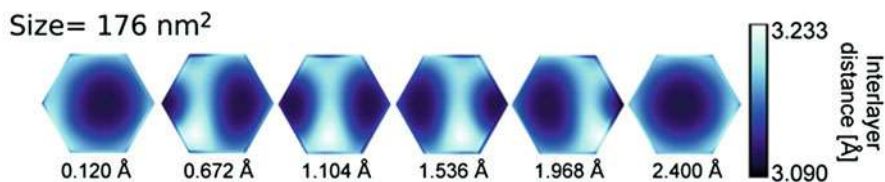
Fig. 5 Snapshots of the soliton-like smooth sliding motion of the moiré superstructure ridges occurring upon shearing of a graphene/*h*-BN junction. (Reprinted with permission from Mandelli et al. (2017). Copyright (2017) American Chemical Society)
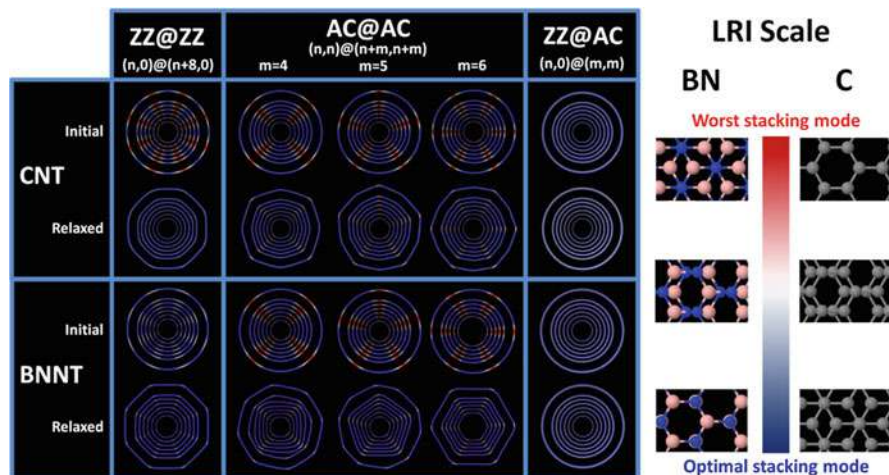
Fig. 6 Schematic representation of achiral double-walled carbon (top two rows) and boron-nitride (bottom two rows) nanotubes, showing their structure and local registry patterns prior to and after geometry relaxation. Each frame shows seven double-walled nanotubes with diameters in the range of 5–20 nm. Five groups of double-walled nanotubes are presented (from left to right): ZZ@ZZ $(n,0)@(n + 8,0)$; AC@AC $(n,n)@(n + 4,n + 4)$, $(n,n)@(n + 5,n + 5)$ and $(n,n)@(n + 6,n + 6)$; and ZZ@AC. For the ZZ@ZZ systems the $n = 55, 80, 105, 130, 155, 180$, and 243 nanotubes are chosen. For the AC@AC systems the $n = 31, 46, 60, 75, 89, 104$, and 140 are chosen. For the ZZ@AC systems, the following set is considered: $(54,0)@(36,36)$, $(80,0)@(51,51)$, $(104,0)@(65,65)$, $(130,0)@(80,80)$, $(154,0)@(94,94)$, $(179,0)@(108,108)$, and $(241,0)@(144,144)$. The local registry index color bar (right) ranges from blue to red, representing the optimal and worst stacking modes of graphene and h-BN, respectively. Grey, pink, and blue spheres on the right-hand-side represent carbon, boron, and nitrogen atoms, respectively. (Reprinted with permission from Leven et al. (2016b). Copyright (2016) by the American Physical Society)

condition for the formation of faceting. Considering first achiral nanotubes, where both nanotube walls are of the same type (either zigzag (ZZ) or armchair (AC)), it was found that above a critical diameter that corresponds well with experimental observations (Garel et al. 2012), evenly spaced extended axial facets form (see Fig. 6). Notably, the number of facets equals the interlayer difference in number of circumferential unit-cells. Interestingly, similar axial facets are formed also for mono-chiral double-walled nanotubes, where both nanotube walls are chiral but share the same chiral angle (see left column in Fig. 7). Elongated helical facets, similar to those commonly observed in experiment, appear in nanotubes that exhibit small interlayer chiral angle mismatch (see middle columns in Fig. 7). These are gradually suppressed with increasing interwall chiral angle difference resulting in outer layer corrugation, which is induced by the moiré superlattice, in agreement with experiments (Schouteden et al. 2013). Since multiwalled boron-nitride nanotubes exhibit better interwall chiral angle matching (Celik-Aktas et al.
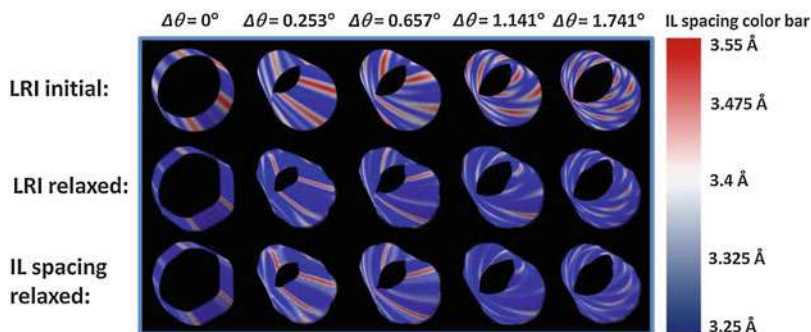
Fig. 7 Schematic representation of (120,100)@(126,105) (leftmost column), (60,60)@(66,65) (second column), (70,70)@(77,74) (third column), (68,68)@(75,70) (fourth column), and (71,71)@(80,72) (rightmost column) double-walled boron-nitride nanotubes showing their local interlayer registry patterns before (top row) and after (middle row) geometry relaxation. The color bar on the right refers to the interlayer spacing of the different systems presented in the bottom row. The colors used in the local registry index patterns are the same as in Fig. 6 above. The chiral angle difference between the inner and outer shells is indicated at the top of each column. (Reprinted with permission from Leven et al. (2016b). Copyright (2016) American Chemical Society)

2005a; Golberg et al. 1999, 2000; Celik-Aktas et al. 2005b) than their carbon counterparts (Schouteden et al. 2013; Zuo et al. 2003; Li et al. 2003; Koziol et al. 2005; Hashimoto et al. 2005; Xu et al. 2006; Ducati et al. 2006; Hirahara et al. 2006; Gao et al. 2006; Guo and Guo 2007; Guan et al. 2008), these findings rationalize the experimental observation of relative abundance of faceting in the former (Celik-Aktas et al. 2005a, b; Golberg et al. 1999, 2000).

In resemblance to the effect of moiré structures on the interlayer sliding friction of planar interfaces, nanotube circumferential faceting is expected to have a notable impact on the interwall sliding friction. When the relative axial or angular alignment of adjacent nanotube walls is varied, the corresponding interwall lattice registry changes. If the nanotube is faceted, then this modification results in rearrangement of the faceted superstructure (Guerra et al. 2017). In the case of achiral or mono-chiral double-walled nanotubes, where the facets are axially aligned, a series of unfaceting and refaceting events occurs during the interwall motion accompanied by facet rotation and reconfiguration (see Fig. 8a, b). For bi-chiral systems, where helical facets appear, interwall telescoping results in rotation of the entire faceted superstructure akin to a rotating Archimedean screw (see Fig. 8c). These global structural variations introduce new dissipative channels that enhance interwall friction (Guerra et al. 2017). This, along with the above-mentioned fact that faceting is more abundant in multiwalled boron-nitride nanotubes than in their carbon counterparts, may rationalize recent experimental measurements showing that the former exhibit enhanced interwall friction (Nigues et al. 2014).
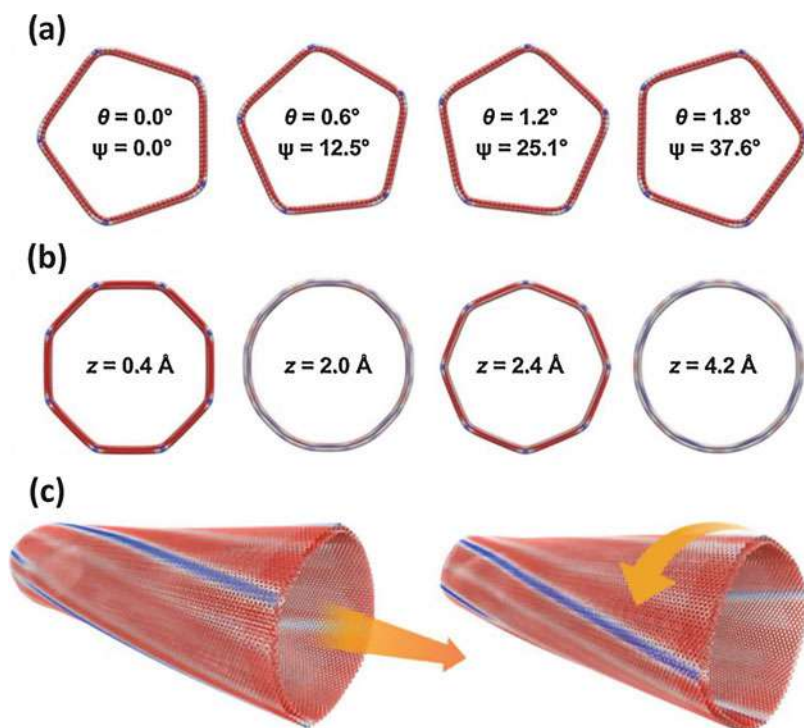
Fig. 8 Facet reconfiguration during interwall pull-out and rotation in (**a**) armchair (104,104)@(109,109); (**b**) zigzag (180,0)@(188,0); and (**c**) bichiral (70,70)@(77,74) double-walled boron-nitride nanotubes. θ, z, and Ψ values appearing in panels (**a**) and (**b**) indicate the relative angular and axial positions of the outer and inner walls and the corresponding facet rotation angle, respectively. The two configurations appearing in panel (**c**) have relative interwall angular and axial orientations of 0.2°/2.4 Å (left) and 0.2°/3.2 Å (right). These correspond to configurations close to the maximum and minimum of the interwall sliding-rotation potential energy surface, respectively. Red, white, and blue atom false coloring indicate low, average, and high atomic interlayer energy, respectively. (Reprinted with permission from Guerra et al. (2017). Copyright (2017) American Chemical Society)

5   Summary

Two-dimensional hexagonal layered materials are of the most promising systems in the field of material science to deliver new technological breakthroughs. The large variety of members in this family and their diverse chemical and physical properties form a vast playground for the design of homogeneous and heterogeneous systems presenting novel structure-function relations. The small dimensions characterizing this field of research call for a synergic interplay between experiment, theory, and computation. With this respect, carefully tailored classical force-field may provide valuable insights on the atomistic nature of different structural, mechanical,

dynamical, and tribological phenomena found in these systems. In the present chapter, a review of recent advances in the field of anisotropic force-fields for describing the interlayer interactions in hexagonal layered materials has been provided. The performance of such force-fields was demonstrated via several applications that address the structural and frictional properties of planar hetero-junctions of graphene and *h*-BN as well as faceted double-walled nanotubes. The good agreement obtained with several experimental observations indicates the predictive power of the developed interlayer force-fields and their ability to rationalize experimental results. The extension of such dedicated force-fields to treat other layered materials is an ongoing effort that, when fulfilled, is expected to enable the discovery of novel low-dimensional structures. Furthermore, the utilization of such force-fields within the framework of multiscale modeling as a source of reliable atomistic information for coarse grained and continuum treatments (Leven et al. 2013; Koren et al. 2016; Leven et al. 2016a; Jung et al. 2015; Kumar et al. 2015; Marom et al. 2010; Hod 2010; Hod 2012b; Blumberg et al. 2012; Garel et al. 2012; Hod 2013; Oz et al. 2016; Leven et al. 2016b; Girifalco 1992; Last et al. 1999; Shenoy et al. 2008; Lu et al. 2009; Yang et al. 2011; Chenxi et al. 2014; Ruiz et al. 2015; Ward 2016; Lebedev et al. 2016) may lead to the development of new nano- and microelectromechanical systems with diverse functionalities based on the promising concepts of 2D layered materials.

References

Al-Jishi R, Dresselhaus G (1982) Lattice-dynamical model for graphite. Phys Rev B 26:4514–4522

Allen MP, Tildesley DJ (2009) Computer simulation of liquids. Oxford University Press, New York

Ambrosetti A, Reilly AM, DiStasio RA, Tkatchenko A (2014) Long-range correlation energy calculated from coupled atomic response functions. J Chem Phys 140:18A508

Argentero G, Mittelberger A, Reza Ahmadpour Monazam M, Cao Y, Pennycook TJ, Mangler C, Kramberger C, Kotakoski J, Geim AK, Meyer JC (2017) Unraveling the 3D atomic structure of a suspended graphene/hBN van der Waals heterostructure. Nano Lett 17:1409–1416

Balandin AA, Ghosh S, Bao W, Calizo I, Teweldebrhan D, Miao F, Lau CN (2008) Superior thermal conductivity of single-layer graphene. Nano Lett 8:902–907

Barone V, Hod O, Peralta JE, Scuseria GE (2011) Accurate prediction of the electronic properties of low-dimensional graphene derivatives using a screened hybrid density functional. Acc Chem Res 44:269–279

Bets KV, Yakobson BI (2009) Spontaneous twist and intrinsic instabilities of pristine graphene nanoribbons. Nano Res 2:161–166

Blumberg A, Keshet U, Zaltsman I, Hod O (2012) Interlayer registry to determine the sliding potential of layered metal dichalcogenides: the case of 2H-MoS_2. J Phys Chem Lett 3:1936–1940

Brenner DW (1988) Tersoff-type potentials for carbon, hydrogen and oxygen. Mat Res Soc Symp Proc 141:59–64

Brenner DW (1990) Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. Phys Rev B 42:9458–9471

Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys Condens Matter 14:783

Brunier TM, Drew MGB, Mitchell PCH (1992) Molecular mechanics studies of molybdenum disulphide catalysts parameterisation of molybdenum and sulphur. Mol Sim 9:143–159

Bucholz EW, Sinnott SB (2013) Structural effects on mechanical response of MoS_2 nanostructures during compression. J Appl Phys 114:034308

Buldum A, Lu JP (1999) Atomic scale sliding and rolling of carbon nanotubes. Phys Rev Lett 83:5050–5053

Bultinck P, Langenaeker W, Lahorte P, De Proft F, Geerlings P, Waroquier M, Tollenaere JP (2002) The electronegativity equalization method I: parametrization and validation for atomic charge calculations. J Phys Chem A 106:7887–7894

Butz B, Dolle C, Niekiel F, Weber K, Waldmann D, Weber HB, Meyer B, Spiecker E (2014) Dislocations in bilayer graphene. Nature 505:533–537

Cai W, Moore AL, Zhu Y, Li X, Chen S, Shi L, Ruoff RS (2010) Thermal transport in suspended and supported monolayer graphene grown by chemical vapor deposition. Nano Lett 10:1645–1651

Carlson A, Dumitrică T (2007) Extended tight-binding potential for modelling intertube interactions in carbon nanotubes. Nanotechnology 18:065706

Celik-Aktas A, Zuo JM, Stubbins JF, Tang CC, Bando Y (2005a) Double-helix structure in multiwall boron nitride nanotubes. Acta Crystallogr A 61:533–541

Celik-Aktas A, Zuo JM, Stubbins JF, Tang C, Bando Y (2005b) Structure and chirality distribution of multiwalled boron nitride nanotubes. Appl Phys Lett 86:133110

Che J, Çağın T, Deng W, Goddard WA (2000) Thermal conductivity of diamond and related materials from molecular dynamics simulations. J Chem Phys 113:6888–6900

Chen Y, Shen Z, Xu Z, Hu Y, Xu H, Wang S, Guo X, Zhang Y, Peng L, Ding F, Liu Z, Zhang J (2013) Helicity-dependent single-walled carbon nanotube alignment on graphite for helical angle and handedness recognition. Nat Commun 4:2205

Chen Z-G, Shi Z, Yang W, Lu X, Lai Y, Yan H, Wang F, Zhang G, Li Z (2014) Observation of an intrinsic bandgap and Landau level renormalization in graphene/boron-nitride heterostructures. Nat Commun 5:4461

Chenxi Z, Jizhou S, Qingda Y (2014) Periodic buckling patterns of graphene/hexagonal boron nitride heterostructure. Nanotechnology 25:445401

Costamagna S, Neek-Amal M, Los JH, Peeters FM (2012) Thermal rippling behavior of graphane. Phys Rev B 86:041408

Dai S, Ma Q, Liu MK, Andersen T, Fei Z, Goldflam MD, Wagner M, Watanabe K, Taniguchi T, Thiemens M, Keilmann F, Janssen GCAM, Zhu SE, Jarillo Herrero P, Fogler MM, Basov DN (2015) Graphene on hexagonal boron nitride as a tunable hyperbolic metamaterial. Nat Nano 10:682–686

Das S, Robinson JA, Dubey M, Terrones H, Terrones M (2015) Beyond graphene: progress in novel two-dimensional materials and van der Waals solids. Ann Rev Mater Res 45:1–27

David JA, Lincoln DC, Mark TL (2010) Embedded ribbons of graphene allotropes: an extended defect perspective. New J Phys 12:125006

Dean CR, Young AF, Meric I, Lee C, Wang L, Sorgenfrei S, Watanabe K, Taniguchi T, Kim P, Shepard KL, Hone J (2010) Boron nitride substrates for high-quality graphene electronics. Nat Nano 5:722–726

Dean CR, Wang L, Maher P, Forsythe C, Ghahari F, Gao Y, Katoch J, Ishigami M, Moon P, Koshino M, Taniguchi T, Watanabe K, Shepard KL, Hone J, Kim P (2013) Hofstadter's butterfly and the fractal quantum hall effect in moiré superlattices. Nature 497:598–602

Deng D, Novoselov KS, Fu Q, Zheng N, Tian Z, Bao X (2016) Catalysis with two-dimensional materials and their heterostructures. Nat Nano 11:218–230

Dienwiebel M, Verhoeven GS, Pradeep N, Frenken JWM, Heimberg JA, Zandbergen HW (2004) Superlubricity of graphite. Phys Rev Lett 92:126101

Ducati C, Koziol K, Friedrichs S, Yates TJV, Shaffer MS, Midgley PA, Windle AH (2006) Crystallographic order in multi-walled carbon nanotubes synthesized in the presence of nitrogen. Small 2:774–784

Ewald PP (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann Phys 369:253–287

Fan Y, Zhao M, Wang Z, Zhang X, Zhang H (2011) Tunable electronic structures of graphene/boron nitride heterobilayers. Appl Phys Lett 98:083103

Fasolino A, Los JH, Katsnelson MI (2007) Intrinsic ripples in graphene. Nat Mater 6:858–861

Frenkel D, Smit B (2002) Understanding molecular simulation. Academic, San Diego

Gao M, Zuo JM, Zhang R, Nagahara LA (2006) Structure determinations of double-wall carbon nanotubes grown by catalytic chemical vapor deposition. J Mater Sci 41:4382–4388

Garel J, Leven I, Zhi C, Nagapriya KS, Popovitz-Biro R, Golberg D, Bando Y, Hod O, Joselevich E (2012) Ultrahigh torsional stiffness and strength of boron nitride nanotubes. Nano Lett 12:6347–6352

Garel J, Zhao C, Popovitz-Biro R, Golberg D, Wang W, Joselevich E (2014) BCN nanotubes as highly sensitive torsional electromechanical transducers. Nano Lett 14:6132–6137

Geim AK (2009) Graphene: status and prospects. Science 324:1530–1534

Geim AK, Grigorieva IV (2013) Van der Waals heterostructures. Nature 499:419–425

Geim AK, Novoselov KS (2007) The rise of graphene. Nat Mater 6:183–191

Ghiringhelli LM, Los JH, Meijer EJ, Fasolino A, Frenkel D (2005a) Modeling the phase diagram of carbon. Phys Rev Lett 94:145701

Ghiringhelli LM, Los JH, Fasolino A, Meijer EJ (2005b) Improved long-range reactive bond-order potential for carbon. II. Molecular simulation of liquid carbon. Phys Rev B 72:214103

Ghiringhelli LM, Valeriani C, Los JH, Meijer EJ, Fasolino A, Frenkel D (2008) State-of-the-art models for the phase diagram of carbon and diamond nucleation. Mol Phys 106:2011–2038

Girifalco LA (1992) Molecular properties of fullerene in the gas and solid phases. J Phys Chem 96:858–861

Girifalco LA, Hodak M (2002) Van der Waals binding energies in graphitic structures. Phys Rev B 65:125404

Girifalco LA, Lad RA (1956) Energy of cohesion, compressibility, and the potential energy functions of the graphite system. J Chem Phys 25:693–697

Girifalco LA, Hodak M, Lee RS (2000) Carbon nanotubes, buckyballs, ropes, and a universal graphitic potential. Phys Rev B 62:13104–13110

Gogotsi Y, Libera JA, Kalashnikov N, Yoshimura M (2000) Graphite polyhedral crystals. Science 290:317–320

Golberg D, Han W, Bando Y, Bourgeois L, Kurashima K, Sato T (1999) Fine structure of boron nitride nanotubes produced from carbon nanotubes by a substitution reaction. J Appl Phys 86:2364–2366

Golberg D, Bando Y, Kurashima K, Sato T (2000) Ropes of BN multi-walled nanotubes. Solid State Commun 116:1–6

Golberg D, Mitome M, Bando Y, Tang CC, Zhi CY (2007a) Multi-walled boron nitride nanotubes composed of diverse cross-section and helix type shells. Appl Phys A Mater Sci 88:347–352

Golberg D, Costa PMFJ, Lourie O, Mitome M, Bai X, Kurashima K, Zhi C, Tang C, Bando Y (2007b) Direct force measurements and kinking under elastic deformation of individual multiwalled boron nitride nanotubes. Nano Lett 7:2146–2151

González Noya E, Srivastava D, Chernozatonskii LA, Menon M (2004) Thermal conductivity of carbon nanotube peapods. Phys Rev B 70:115416

Green JF, Bolland TK, Bolland JW (1974) Lennard-Jones interaction for hexagonal layered crystals. J Chem Phys 61:1637–1646

Guan L, Suenaga K, Iijima S (2008) Smallest carbon nanotube assigned with atomic resolution accuracy. Nano Lett 8:459–462

Guerra R, Tartaglino U, Vanossi A, Tosatti E (2010) Ballistic nanofriction. Nat Mater 9:634–637

Guerra R, Leven I, Vanossi A, Hod O, Tosatti E (2017) Smallest Archimedean screw: facet dynamics and friction in multiwalled nanotubes. Nano Lett 17:5321–5328

Guo W, Guo Y (2007) Energy optimum chiralities of multiwalled carbon nanotubes. J Am Chem Soc 129:2730–2731

Guo Y, Guo W, Chen C (2007) Modifying atomic-scale friction between two graphene sheets: a molecular-force-field study. Phys Rev B 76:155429

Hashimoto A, Suenaga K, Urita K, Shimada T, Sugai T, Bandow S, Shinohara H, Iijima S (2005) Atomic correlation between adjacent graphene layers in double-wall carbon nanotubes. Phys Rev Lett 94:045504

Heine T (2015) Transition metal chalcogenides: ultrathin inorganic materials with tunable electronic properties. Acc Chem Res 48:65–72

Hertel T, Walkup RE, Avouris P (1998) Deformation of carbon nanotubes by surface van der Waals forces. Phys Rev B 58:13870–13873

Heyd J, Scuseria GE (2004a) Assessment and validation of a screened Coulomb hybrid density functional. J Chem Phys 120:7274–7280

Heyd J, Scuseria GE (2004b) Efficient hybrid density functional calculations in solids: assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional. J Chem Phys 121:1187–1192

Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened Coulomb potential. J Chem Phys 118:8207–8215

Hirahara K, Kociak M, Bandow S, Nakahira T, Itoh K, Saito Y, Iijima S (2006) Chirality correlation in double-wall carbon nanotubes as studied by electron diffraction. Phys Rev B 73:195420

Hod O (2010) Quantifying the stacking registry matching in layered materials. Isr J Chem 50:506–514

Hod O (2012a) Graphite and hexagonal boron-nitride have the same interlayer distance. Why? J Chem Theory Comput 8:1360–1369

Hod O (2012b) Interlayer commensurability and superlubricity in rigid layered materials. Phys Rev B 86:075444

Hod O (2013) The registry index: a quantitative measure of materials′ interfacial commensurability. ChemPhysChem 14:2376–2391

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871

Jayasena B, Reddy CD, Subbiah S (2013) Separation, folding and shearing of graphene layers during wedge-based mechanical exfoliation. Nanotechnology 24:205301

Jiang J-W (2014a) Registry effect on the thermal conductivity of few-layer graphene. J Appl Phys 116:164313

Jiang JW (2014b) The buckling of single-layer MoS_2 under uniaxial compression. Nanotechnology 25:355402

Jiang J-W (2015) Graphene versus MoS_2: a short review. Front Phys 10:287–302

Jiang J-W, Park HS (2014) Mechanical properties of MoS_2/graphene heterostructures. Appl Phys Lett 105:033108

Jiang J-W, Park HS (2015) A Gaussian treatment for the friction issue of Lennard-Jones potential in layered materials: application to friction between graphene, MoS_2, and black phosphorus. J Appl Phys 117:124304

Jiang J-W, Park HS, Rabczuk T (2013) Molecular dynamics simulations of single-layer molybdenum disulphide (MoS_2): Stillinger-Weber parametrization, mechanical properties, and thermal conductivity. J Appl Phys 114:064307

Jiménez Sandoval S, Yang D, Frindt RF, Irwin JC (1991) Raman study and lattice dynamics of single molecular layers of MoS_2. Phys Rev B 44:3955–3962

Jin-Wu J, Zenan Q, Harold SP, Timon R (2013) Elastic bending modulus of single-layer molybdenum disulfide (MoS_2): finite thickness effect. Nanotechnology 24:435705

Jung J, DaSilva AM, MacDonald AH, Adam S (2015) Origin of band gaps in graphene on hexagonal boron nitride. Nat Commun 6:6308

Kang JW, Lee KW (2014) Molecular dynamics simulation of square graphene-nanoflake oscillator on graphene nanoribbon. J Nanosci Nanotechnol 14:9158–9164

Karssemeijer LJ, Fasolino A (2011) Phonons of graphene and graphitic materials derived from the empirical potential LCBOPII. Surf Sci 605:1611–1615

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138

Kolmogorov AN, Crespi VH (2000) Smoothest bearings: interlayer sliding in multiwalled carbon nanotubes. Phys Rev Lett 85:4727–4730

Kolmogorov AN, Crespi VH (2005) Registry-dependent interlayer potential for graphitic systems. Phys Rev B 71:235415

Koren E, Lörtscher E, Rawlings C, Knoll AW, Duerig U (2015) Adhesion and friction in mesoscopic graphite contacts. Science 348:679–683

Koren E, Leven I, Lörtscher E, Knoll A, Hod O, Duerig U (2016) Coherent commensurate electronic states at the interface between misoriented graphene layers. Nat Nano 11:752–757

Korhonen T, Koskinen P (2015) Peeling of multilayer graphene creates complex interlayer sliding patterns. Phys Rev B 92:115427

Koziol K, Shaffer M, Windle A (2005) Three-dimensional internal order in multiwalled carbon nanotubes grown by chemical vapor deposition. Adv Mater 17:760–763

Kumar H, Er D, Dong L, Li J, Shenoy VB (2015) Elastic deformations in 2D van der Waals heterostructures and their impact on optoelectronic properties: predictions from a multiscale computational approach. Sci Rep 5:10872

Kushima A, Qian X, Zhao P, Zhang S, Li J (2015) Ripplocations in van der Waals layers. Nano Lett 15:1302–1308

Kuzuba T, Sato T, Ishii T, Arai T (1985) Interlayer binding of hexagonal boron nitride in the rigid-layer approximation. Phys Rev B 32:1230–1236

Kwon Y-K, Tománek D (1998) Electronic and structural properties of multiwall carbon nanotubes. Phys Rev B 58:R16001–R16004

Kwon Y-K, Tománek D (2000) Orientational melting in carbon nanotube ropes. Phys Rev Lett 84:1483–1486

Last JA, Hooks DE, Hillier AC, Ward MD (1999) The physicochemical origins of coincident epitaxy in molecular overlayers: lattice modeling vs potential energy calculations. J Phys Chem B 103:6723–6733

Lebedev AV, Lebedeva IV, Knizhnik AA, Popov AM (2016) Interlayer interaction and related properties of bilayer hexagonal boron nitride: Ab initio study. RSC Adv 6:6423–6435

Lebedeva IV, Knizhnik AA, Popov AM, Ershova OV, Lozovik YE, Potapkin BV (2011a) Diffusion and drift of graphene flake on graphite surface. J Chem Phys 134:104505

Lebedeva IV, Knizhnik AA, Popov AM, Lozovik YE, Potapkin BV (2011b) Interlayer interaction and relative vibrations of bilayer graphene. Phys Chem Chem Phys 13:5687–5695

Lebedeva IV, Knizhnik AA, Popov AM, Lozovik YE, Potapkin BV (2012) Modeling of graphene-based NEMS. Phys E 44:949–954

Lee C, Wei X, Kysar JW, Hone J (2008) Measurement of the elastic properties and intrinsic strength of monolayer graphene. Science 321:385–388

Leven I, Krepel D, Shemesh O, Hod O (2013) Robust superlubricity in graphene/*h*-BN heterojunctions. J Phys Chem Lett 4:115–120

Leven I, Azuri I, Kronik L, Hod O (2014) Inter-layer potential for hexagonal boron nitride. J Chem Phys 140:104106

Leven I, Maaravi T, Azuri I, Kronik L, Hod O (2016a) Interlayer potential for graphene/*h*-BN heterostructures. J Chem Theor Comput 12:2896–2905

Leven I, Guerra R, Vanossi A, Tosatti E, Hod O (2016b) Multiwalled nanotube faceting unravelled. Nat Nano 11:1082–1086

Levendorf MP, Kim C-J, Brown L, Huang PY, Havener RW, Muller DA, Park J (2012) Graphene and boron nitride lateral heterostructures for atomically thin circuitry. Nature 488:627–632

Li D, Kaner RB (2008) Graphene-based materials. Science 320:1170–1171

Li F, Chou SG, Ren W, Gardecki JA, Swan AK, Ünlü MS, Goldberg BB, Cheng H-M, Dresselhaus MS (2003) Identification of the constituents of double-walled carbon nanotubes using Raman spectra taken with different laser-excitation energies. J Mater Res 18:1251–1258

Li X, Zhu H (2015) Two-dimensional MoS_2: properties, preparation, and applications. J Mater 1:33–44

Liang T, Phillpot SR, Sinnott SB (2009) Parametrization of a reactive many-body potential for Mo-S systems. Phys Rev B 79:245110

Lindsay L, Broido DA (2010) Optimized Tersoff and Brenner empirical potential parameters for lattice dynamics and phonon thermal transport in carbon nanotubes and graphene. Phys Rev B 81:205441

Lindsay L, Broido DA (2011) Enhanced thermal conductivity and isotope effect in single-layer hexagonal boron nitride. Phys Rev B 84:155421

Liu B, Meng F, Reddy CD, Baimova JA, Srikanth N, Dmitriev SV, Zhou K (2015) Thermal transport in a graphene-MoS_2 bilayer heterostructure: a molecular dynamics study. RSC Adv 5:29193–29200

Liu M, Cowley JM (1994) Structures of the helical carbon nanotubes. Carbon 32:393–403

Los JH, Fasolino A (2003) Intrinsic long-range bond-order potential for carbon: performance in Monte Carlo simulations of graphitization. Phys Rev B 68:024107

Los JH, Ghiringhelli LM, Meijer EJ, Fasolino A (2005) Improved long-range reactive bond-order potential for carbon. I. Construction. Phys Rev B 72:214102

Lu JP (1997) Elastic properties of carbon nanotubes and nanoropes. Phys Rev Lett 79:1297–1300

Lu JP, Li XP, Martin RM (1992) Ground state and phase transitions in solid C_{60}. Phys Rev Lett 68:1551–1554

Lu WB, Liu B, Wu J, Xiao J, Hwang KC, Fu SY, Huang Y (2009) Continuum modeling of van der Waals interactions between carbon nanotube walls. Appl Phys Lett 94:101917

Maaravi T, Leven I, Azuri I, Kronik L, Hod O (2017) Interlayer potential for homogeneous graphene and hexagonal boron nitride systems: reparametrization for many-body dispersion effects. J Phys Chem C 121:22826–22835. https://pubs.acs.org/doi/10.1021/acs.nanolett.8b02848

Mandelli D, Leven I, Hod O, Urbakh M (2017) Sliding friction of graphene/hexagonal–boron nitride heterojunctions: a route to robust superlubricity. Sci Rep 7:10851

Marom N, Bernstein J, Garel J, Tkatchenko A, Joselevich E, Kronik L, Hod O (2010) Stacking and registry effects in layered materials: the case of hexagonal boron nitride. Phys Rev Lett 105:046801

Marom N, Tkatchenko A, Rossi M, Gobre VV, Hod O, Scheffler M, Kronik L (2011) Dispersion interactions with density-functional theory: benchmarking semiempirical and interatomic pairwise corrected density functionals. J Chem Theor Comput 7:3944–3951

Michel KH, Verberck B (2011) Phonon dispersions and piezoelectricity in bulk and multilayers of hexagonal boron nitride. Phys Rev B 83:115328

Miró P, Ghorbani-Asl M, Heine T (2013) Spontaneous ripple formation in MoS_2 monolayers: electronic structure and transport effects. Adv Mater 25:5473–5475

Mohamed AO, Deepak S (2001) Temperature dependence of the thermal conductivity of single-wall carbon nanotubes. Nanotechnology 12:21

Morita Y, Onodera T, Suzuki A, Sahnoun R, Koyama M, Tsuboi H, Hatakeyama N, Endou A, Takaba H, Kubo M, Del Carpio CA, Shin-yoshi T, Nishino N, Suzuki A, Miyamoto A (2008) Development of a new molecular dynamics method for tribochemical reaction and its application to formation dynamics of MoS_2 tribofilm. Appl Surf Sci 254:7618–7621

Morozov SV, Novoselov KS, Katsnelson MI, Schedin F, Elias DC, Jaszczak JA, Geim AK (2008) Giant intrinsic carrier mobilities in graphene and its bilayer. Phys Rev Lett 100:016602

Mortier WJ, Ghosh SK, Shankar S (1986) Electronegativity-equalization method for the calculation of atomic charges in molecules. J Am Chem Soc 108:4315–4320

Neek-Amal M, Peeters FM (2014) Graphene on hexagonal lattice substrate: stress and pseudo-magnetic field. Appl Phys Lett 104:173106

Nicolini P, Polcar T (2016) A comparison of empirical potentials for sliding simulations of MoS_2. Comput Mater Sci 115:158–169

Nightingale MP, Umrigar CJ (eds) (1999) Quantum Monte Carlo methods in physics and chemistry. Springer, Netherlands

Nigues A, Siria A, Vincent P, Poncharal P, Bocquet L (2014) Ultrahigh interlayer friction in multiwalled boron nitride nanotubes. Nat Mater 13:688–693

Njo SL, Fan J, van de Graaf B (1998) Extending and simplifying the electronegativity equalization method. J Mol Catal A 134:79–88

Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA (2004) Electric field effect in atomically thin carbon films. Science 306:666–669

Novoselov KS, Geim AK, Morozov SV, Jiang D, Katsnelson MI, Grigorieva IV, Dubonos SV, Firsov AA (2005a) Two-dimensional gas of massless Dirac fermions in graphene. Nature 438:197–200

Novoselov KS, Jiang D, Schedin F, Booth TJ, Khotkevich VV, Morozov SV, Geim AK (2005b) Two-dimensional atomic crystals. Proc Natl Acad Sci U S A 102:10451–10453

Novoselov KS, Mishchenko A, Carvalho A, Castro Neto AH (2016) 2D materials and van der Waals heterostructures. Science 353:aac9439

O'Connor TC, Andzelm J, Robbins MO (2015) AIREBO-M: a reactive model for hydrocarbons at extreme pressures. J Chem Phys 142:024903

Oz I, Leven I, Itkin Y, Buchwalter A, Akulov K, Hod O (2016) Nanotube motion on layered materials: a registry perspective. J Phys Chem C 120:4466–4470

Palser AHR (1999) Interlayer interactions in graphite and carbon nanotubes. Phys Chem Chem Phys 1:4459–4464

Pang T (2016) An introduction to quantum Monte Carlo methods. Morgan & Claypool Publishers, San Rafael

Parr RG, Pearson RG (1983) Absolute hardness: companion parameter to absolute electronegativity. J Am Chem Soc 105:7512–7516

Parr RG, Donnelly RA, Levy M, Palke WE (1978) Electronegativity: the density functional viewpoint. J Chem Phys 68:3801–3807

Perebeinos V, Tersoff J (2009) Valence force model for phonons in graphene and carbon nanotubes. Phys Rev B 79:241409

Ponomarenko LA, Gorbachev RV, Yu GL, Elias DC, Jalil R, Patel AA, Mishchenko A, Mayorov AS, Woods CR, Wallbank JR, Mucha-Kruczynski M, Piot BA, Potemski M, Grigorieva IV, Novoselov KS, Guinea F, Fal'ko VI, Geim AK (2013) Cloning of dirac fermions in graphene superlattices. Nature 497:594–597

Pople JA (1970) Approximate molecular orbital theory (advanced chemistry). McGraw-Hill, New York

Popov AM, Lebedeva IV, Knizhnik AA, Lozovik YE, Potapkin BV (2011) Molecular dynamics simulation of the self-retracting motion of a graphene flake. Phys Rev B 84:245437

Porezag D, Frauenheim T, Köhler T, Seifert G, Kaschner R (1995) Construction of tight-binding-like potentials on the basis of density-functional theory: application to carbon. Phys Rev B 51:12947–12957

Qian D, Liu WK, Ruoff RS (2003) Load transfer mechanism in carbon nanotube ropes. Compos Sci Technol 63:1561–1569

Rao CNR, Sood AK, Subrahmanyam KS, Govindaraj A (2009) Graphene: the new two-dimensional nanomaterial. Angew Chem Int Ed (English) 48:7752–7777

Rapaport DC (2009) The art of molecular dynamics simulation. Cambridge University Press, New York

Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Amer Chem Soc 114:10024–10035

Reguzzoni M, Fasolino A, Molinari E, Righi MC (2012a) Friction by shear deformations in multilayer graphene. J Phys Chem C 116:21104–21108

Reguzzoni M, Fasolino A, Molinari E, Righi MC (2012b) Potential energy surface for graphene on graphene: Ab initio derivation, analytical description, and microscopic interpretation. Phys Rev B 86:245434

Ruiz L, Xia W, Meng Z, Keten S (2015) A coarse-grained model for the mechanical behavior of multi-layer graphene. Carbon 82:103–115

Sakai Y, Saito S, Cohen ML (2015) First-principles study on graphene/hexagonal boron nitride heterostructures. J Phys Soc Jap 84:121002

Sanderson RT (1951) An interpretation of bond lengths and a classification of bonds. Science 114:670–672

Sanderson RT (1976) Chemical bonds and bond energy, 2nd edn. Academic, New York

Sanderson RT (1983) Polar covalence. Academic, New York

Schouteden K, Volodin A, Li Z, Van Haesendonck C (2013) Atomically resolved moiré-type superstructures in double-walled carbon nanotubes. Carbon 61:379–385

Sevik C, Kinaci A, Haskins JB, Çağın T (2011) Characterization of thermal transport in low-dimensional boron nitride nanostructures. Phys Rev B 84:085409

Shenoy VB, Reddy CD, Ramasubramaniam A, Zhang YW (2008) Edge-stress-induced warping of graphene sheets and nanoribbons. Phys Rev Lett 101:245501

Shibuta Y, Elliott JA (2011) Interaction between two graphene sheets with a turbostratic orientational relationship. Chem Phys Lett 512:146–150

Shtogun YV, Woods LM (2010) Many-body van der Waals interactions between graphitic nanostructures. J Phys Chem Lett 1:1356–1362

Singh SK, Neek-Amal M, Costamagna S, Peeters FM (2013) Thermomechanical properties of a single hexagonal boron nitride sheet. Phys Rev B 87:184106

Singh SK, Neek-Amal M, Costamagna S, Peeters FM (2015) Rippling, buckling, and melting of single- and multilayer MoS_2. Phys Rev B 91:014101

Strutyński K, Gomes JANF, Melle-Franco M (2014) Accuracy of dispersion interactions in semiempirical and molecular mechanics models: the benzene dimer case. J Phys Chem A 118:9561–9567

Stuart SJ, Tutein AB, Harrison JA (2000) A reactive potential for hydrocarbons with intermolecular interactions. J Chem Phys 112:6472–6486

Szabados Á, Biró LP, Surján PR (2006) Intertube interactions in carbon nanotube bundles. Phys Rev B 73:195404

Szabo A, Ostlund NS (1989) Modern quantum chemistry: introduction to advanced electronic structure theory. Dover Publications, Mineola

Tang S, Wang H, Zhang Y, Li A, Xie H, Liu X, Liu L, Li T, Huang F, Xie X, Jiang M (2013) Precisely aligned graphene grown on hexagonal boron nitride by catalyst free chemical vapor deposition. Sci Rep 3:2666

Tersoff J (1988) Empirical interatomic potential for carbon, with applications to amorphous carbon. Phys Rev Lett 61:2879–2882

Tersoff J (1992) Energies of fullerenes. Phys Rev B 46:15546–15549

Tersoff J, Ruoff RS (1994) Structural properties of a carbon-nanotube crystal. Phys Rev Lett 73:676–679

Tkatchenko A, Scheffler M (2009) Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. Phys Rev Lett 102:073005

Tkatchenko A, DiStasio RA, Car R, Scheffler M (2012) Accurate and efficient method for many-body van der Waals interactions. Phys Rev Lett 108:236402

Tolga A, Jeremy KM, Hakan E (2016) A new interlayer potential for hexagonal boron nitride. J Phys Condens Matter 28:385401

Toukmaji AY, Board JA (1996) Ewald summation techniques in perspective: a survey. Comp Phys Comm 95:73–92

van Duin ACT, Dasgupta S, Lorant F, Goddard WA (2001) Reax FF: a reactive force field for hydrocarbons. J Phys Chem A 105:9396–9409

van Wijk MM, Dienwiebel M, Frenken JWM, Fasolino A (2013) Superlubric to stick-slip sliding of incommensurate graphene flakes on graphite. Phys Rev B 88:235423

van Wijk MM, Schuring A, Katsnelson MI, Fasolino A (2014) Moiré patterns as a probe of interplanar interactions for graphene on h-BN. Phys Rev Lett 113:135504

van Wijk MM, Schuring A, Katsnelson MI, Fasolino A (2015) Relaxation of moiré patterns for slightly misaligned identical lattices: graphene on graphite. 2D Mater 2:034010

Vanossi A, Manini N, Urbakh M, Zapperi S, Tosatti E (2013) Colloquium: modeling friction: from nanoscale to mesoscale. Rev Mod Phys 85:529–552

Varshney V, Patnaik SS, Muratore C, Roy AK, Voevodin AA, Farmer BLMD (2010) Simulations of molybdenum disulphide (MoS_2): force-field parameterization and thermal transport behavior. Comput Mater Sci 48:101–108

Wakabayashi N, Smith HG, Nicklow RM (1975) Lattice dynamics of hexagonal MoS_2 studied by neutron scattering. Phys Rev B 12:659–663

Wang H, Liu F, Fu W, Fang Z, Zhou W, Liu Z (2014) Two-dimensional heterostructures: fabrication, characterization, and application. Nanoscale 6:12250–12272

Ward MD (2016) Soft crystals in flatland: unraveling epitaxial growth. ACS Nano 10:6424–6428

Weismiller MR, Duin ACT v, Lee J, Yetter RA (2010) Reax FF reactive force field development and applications for molecular dynamics simulations of ammonia borane dehydrogenation and combustion. J Phys Chem A 114:5485–5492

Weiss K, Phillips JM (1976) Calculated specific surface energy of molybdenite (MoS2). Phys Rev B 14:5392–5395

Woods CR, Britnell L, Eckmann A, Ma RS, Lu JC, Guo HM, Lin X, Yu GL, Cao Y, Gorbachev RV, Kretinin AV, Park J, Ponomarenko LA, Katsnelson MI, Gornostyrev YN, Watanabe K, Taniguchi T, Casiraghi C, Gao HJ, Geim AK, Novoselov KS (2014) Commensurate-incommensurate transition in graphene on hexagonal boron nitride. Nat Phys 10:451–456

Woods CR, Withers F, Zhu MJ, Cao Y, Yu G, Kozikov A, Ben Shalom M, Morozov SV, van Wijk MM, Fasolino A, Katsnelson MI, Watanabe K, Taniguchi T, Geim AK, Mishchenko A, Novoselov KS (2016) Macroscopic self-reorientation of interacting two-dimensional crystals. Nat Commun 7:10800

Xiaonan W, Alireza T, Douglas ES (2015) Fracture mechanics of monolayer molybdenum disulfide. Nanotechnology 26:175703

Xu Z, Bai X, Wang ZL, Wang E (2006) Multiwall carbon nanotubes made of monochirality graphite shells. J Am Chem Soc 128:1052–1053

Xu Z, Li X, Yakobson BI, Ding F (2013) Interaction between graphene layers and the mechanisms of graphite's superlubricity and self-retraction. Nanoscale 5:6736–6741

Xu X, Pereira LFC, Wang Y, Wu J, Zhang K, Zhao X, Bae S, Tinh Bui C, Xie R, Thong JTL, Hong BH, Loh KP, Donadio D, Li B, Özyilmaz B (2014) Length-dependent thermal conductivity in suspended single-layer graphene. Nat Commun 5:3689

Yang B, Barsoum MW, Rethinam RM (2011) Nanoscale continuum calculation of basal dislocation core structures in graphite. Philoso Mag 91:1441–1463

Yang W, Chen G, Shi Z, Liu C-C, Zhang L, Xie G, Cheng M, Wang D, Yang R, Shi D, Watanabe K, Taniguchi T, Yao Y, Zhang Y, Zhang G (2013) Epitaxial growth of single-domain graphene on hexagonal boron nitride. Nat Mater 12:792–797

Yankowitz M, Xue J, Cormode D, Sanchez-Yamagishi JD, Watanabe K, Taniguchi T, Jarillo-Herrero P, Jacquod P, LeRoy BJ (2012) Emergence of superlattice dirac points in graphene on hexagonal boron nitride. Nat Phys 8:382–386

Yao Z, Wang J-S, Li B, Liu G-R (2005) Thermal conduction of carbon nanotubes using molecular dynamics. Phys Rev B 71:085417

Ye Z, Tang C, Dong Y, Martini A (2012) Role of wrinkle height in friction variation with number of graphene layers. J Appl Phys 112:116102

Ye Z, Otero-de-la-Roza A, Johnson ER, Martini A (2015) Oscillatory motion in layered materials: graphene, boron nitride, and molybdenum disulfide. Nanotechnology 26:165701

Zahid F, Liu L, Zhu Y, Wang J, Guo H (2013) A generic tight-binding model for monolayer, bilayer and bulk MoS2. AIP Adv 3:052111

Ze L (2014) The diversity of friction behavior between bi-layer graphenes. Nanotechnology 25:075703

Zhang G, Jiang X, Wang E (2003) Tubular graphite cones. Science 300:472–474

Zhang GY, Bai XD, Wang EG, Guo Y, Guo W (2005) Monochiral tubular graphite cones formed by radial layer-by-layer growth. Phys Rev B 71:113411

Zuo JM, Vartanyants I, Gao M, Zhang R, Nagahara LA (2003) Atomic resolution imaging of a carbon nanotube from diffraction intensities. Science 300:1419–1421

Emergence of Functionalized Properties in Semiconductor Nanostructures

30

James R. Chelikowsky

Contents

Abstract

The role of functionalizing properties of semiconductor materials in the bulk by doping is reasonably well understood. However, as the length scale of merit approaches the nano-regime, the interplay of restricted dimensionality and quantum confinement can complicate the role dopants play in altering electronic properties. In particular, well-known rules for describing defects in the bulk may become inoperative at the nanoscale. We illustrate how computational methods can be used to predict the emergence of nanoscale electronic properties in semiconductors and how these properties change as the size of the system is decreased.

J. R. Chelikowsky (✉)
Center for Computational Materials, Oden Institute for Computational Engineering and Sciences, Departments of Physics and Chemical Engineering, The University of Texas at Austin, Austin, TX, USA
e-mail: jrc@utexas.edu

1   Introduction

"Functionalization" of material properties is accomplished by changing the chemical makeup of a material so that it possesses a desirable property. In organic materials this can be accomplished by adding functional groups. A similar process occurs in semiconductor materials by the controlled addition of impurities, i.e., doping. Indeed, without precise control of the electronic properties using dopants, solid-state devices would not exist.

A classic example of functionalized electronic materials is the construction of a *pn* junction diode (Kittel 2005). A *pn* junction relies on the existence of an interface between *p*-doped and *n*-doped materials. The *p*-doped material can be created by adding boron atoms to bulk silicon, and the *n*-doped material can be created by adding phosphorous atoms to bulk silicon. The addition of these doping atoms results in profound differences in the silicon host material. For example, one boron atom for every hundred thousand silicon atoms increases the conductivity of pure silicon by a thousandfold at room temperature (Kittel 2005). In a *pn* junction, the electronic states of the host are modified so that an internal field occurs. When an external field is applied, the internal field can be increased or decreased to control the flow of electrons (or holes) so that the resulting current is rectified.

The process and control of such impurity atoms like boron and phosphorous account for the phenomenal progress in the electronics industry with the creation of devices ranging from a simple transistor some 50 years ago to contemporary nanoscale integrated circuits containing billions of transistors. As a consequence, the computational power present in a typical cell phone or handheld personal data device vastly exceeds that of the first electronic computers.

However, as we place more transistors on a chip, the physical dimensions of a transistor become smaller. At a certain point, the size of a transistor will start to affect its physical property. Specifically, size-dependent quantum mechanical phenomena will occur. This was recognized some time ago. In 1959, Richard Feynman presented a seminar (Feynman 1992) in which he stated: "When we get to the very, very small world—say circuits of seven atoms—we have a lot of new things that would happen that represent completely new opportunities for design. Atoms on a small scale behave like nothing on a large scale, for they satisfy the laws of quantum mechanics. So, as we go down and fiddle around with the atoms down there, we are working with different laws, and we can expect to do different things. We can manufacture in different ways. We can use, not just circuits, but some system involving the quantized energy levels, or the interactions of quantized spins, etc." Feynman's comments had little impact in directing new work (Toumey 2005) as the technological framework for nanoscale control took time to implement. Today his comments appear prescient.

The design rules that work successfully at macroscopic dimensions may not be operative at nanoscale dimensions as the physical properties of nanoscale materials can be significantly different than the bulk material. For example, an optically inactive material such as silicon can become optically active at the nanoscale (Canham 2000; Shirley et al. 2000).

The use of "size" to modify the physical properties of materials, without changing their chemical composition technological implications in a number of areas: sensors for bioimaging (Michalet et al. 2005), energy devices such as solar cells (Gur et al. 2005; Ellingson et al. 2005), and wavelength tunable lasers (Klimov et al. 2007).

In general, functionalization at the nanoscale whether by exploiting quantum confinement or by doping is of great importance, but the nanoscale presents new and challenging issues. The inclusion of just one dopant atom in a thousand-atom nanocrystal might exceed the solubility expected for a bulk crystal; in the case of P-doped Si, a doping concentration of one part in a hundred thousand atoms would be considered extremely heavy doping.

Physically, it seems plausible that dopants would interact differently with nanocrystal hosts as a function of size (Beaulac et al. 2010). As the nanocrystal decreases in size, the ratio of surface to bulk atoms grows. The interaction of the dopant with the surface can allow a decrease in any strain energies as the surface atoms are more readily displaced than the bulk atoms, owing to their lower coordination number. Moreover, as the nanocrystal becomes smaller, quantum confinement increases the bandgap and can effectively turn a small bandgap semiconductor into a large gap semiconductor, which is known to be difficult to dope (Chadi 1994; Neumark 1997; Troparevsky and Chelikowsky 2001).

Understanding the doping process at the nanoscale requires a better picture of both thermodynamics and kinetics. Computational methods can be used to solve the electronic structure problem and address some of these issues. In the case of Si, computational work has accurately predicted many properties, and there is no reason to believe that these studies would not describe the nanoscale with similar accuracy. In the early 1980s, it was conclusively demonstrated that one could determine accurate structural energies of silicon. For example, the energetics of a number of different phases of silicon were examined (Yin and Cohen 1980) with the equations of states for dense phases carefully mapped out and found to be in agreement with high-pressure experiments. Moreover, these studies predicted the existence of superconductivity in a high-pressure form of Si. This prediction was confirmed by the experiment within a few years (Chang et al. 1985). Shortly after the realization that accurate structural energies could be computed, methods for ab initio molecular dynamics were introduced (Car and Parrinello 1985).

While these computational methods have greatly enhanced our understanding of bulk materials, their application to the nanoscale is often problematic. Nanostructures are still "large," in a quantum sense. The number of atoms contained within a "smallish" 10 nm nanocrystal of Si is roughly 50,000. Current methods for electronic structure descriptions are ill-suited for such systems as one would have to solve for over 100,000 energy levels. The memory required and the computationally intensive nature of such a problem are often beyond the state of the art, although such systems are becoming within our computational grasp. Here we illustrate how computational approaches and advances can be used to illuminate the quantum behavior and the physical properties of dopants at the nanoscale

for systems approaches containing thousands of atoms using electronic structure methods discussed in Volume 1 of this handbook (Chap. 24 "Extending the Scale with Real-Space Methods for the Electronic Structure Problem" of the MTM set by Chelikowsky).

2   Defect Levels in Doped Nanocrystals

Phosphorus atoms added to bulk silicon represent the leading example of an *n*-type semiconductor. Phosphorous induces defect energy states very close in energy to the conduction band. For such *shallow* donor states, i.e., states within a few kT of the conduction band, electrons can be easily thermally excited. These carriers greatly enhance the conductivity of the original pure semiconductor.

As one considers making device materials at smaller length scales, a central question is whether this picture will continue to be valid (Pavesi and Turan 2010). Quantum confinement in this size regime is expected to alter the electronic properties of doped Si nano-crystals (Yoffe 2001; Melnikov and Chelikowsky 2004; Chan et al. 2008; Chelikowsky et al. 2011). Important questions arise. For example, will the defect energy levels continue to shallow or not. Also, is there a length scale where device construction based on macroscopic guidelines fails? These are the questions that Feynman intuited in the late 1950s.

In Si nanocrystals, the bandgap, E_G, increases as the crystal size decreases owing to quantum confinement as depicted in Fig. 1a. After doping with P, a defect energy level is created within the bandgap. Since P atom has one more electron than a Si atom, the extra electron creates a donor state. The binding energy of the donor within a bulk crystal can be defined as the energy difference between the defect energy level and the conduction band minimum. Within a confined system, this definition is problematic because the electron in the lowest unoccupied state remains
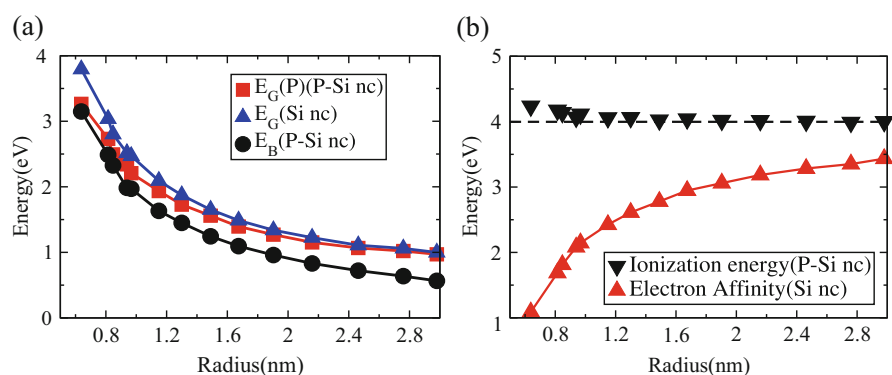
Fig. 1 (**a**) The dependence on the nanocrystal radius of the energy gap E_G in Si nanocrystal(▲), the binding energy E_B(●), the energy difference between the defect level and the highest doubly occupied level $E_G(P)$ in P-doped Si nanocrystal(■). (**b**) Ionization energy of P-doped Si nanocrystal(▼) and electron affinity of pure Si nanocrystal(▲) plotted as a function of nanocrystal radius

physically localized near the P atom. The corresponding definition of the binding energy at the nanoscale is the energy required to ionize a P-doped Si nanocrystal by removing an electron minus the energy gained by adding the electron to a pure Si nanocrystal.

One can calculate the ionization energy I_d for P-doped Si nanocrystals as

$$I_d = E_P(n-1) - E_P(n), \tag{1}$$

where n is the number of electrons in a neutral P-doped Si nanocrystals. $E_P(n)$ and $E_P(n-1)$ are the ground-state total energy of a neutral and positively charged P-doped Si nanocrystals, respectively. The electron affinity of pure Si nanocrystals can be determined by

$$A_p = E_{Si}(n-1) - E_{Si}(n), \tag{2}$$

where $E_{Si}(n-1)$ and $E_{Si}(n)$ are the ground-state total energy of a neutral and negatively charged Si nanocrystals, respectively. The binding energy E_B for the donor electron is the difference between these two quantities:

$$E_B = I_d - A_p. \tag{3}$$

The calculated ionization energy and electron affinity as a function of nanocrystal size are illustrated in Fig. 1b. The electron affinity of pure Si nanocrystal decreases with size owing to the increase of kinetic energy of the added electron by quantum confinement, which can be considered a consequence of the uncertainty principle. We note that the ionization energy of P-doped Si nanocrystal depends weakly on nanocrystal size, i.e., it is effectively pinned to the vacuum level, and remains close to the bulk value of ~ 4 eV.

In Fig. 1a, the size dependence of the energy difference ($E_G(P)$) between the singly occupied defect energy level and the highest doubly occupied level for the doped Si nanocrystal is plotted. The binding energy E_B of P-doped Si nanocrystal as a function of size is also shown. As the size of the nanocrystal increases, the sum $E_G(P)$ for the doped nanocrystal and the binding energy E_B should converge to the bulk Si bandgap.

In nanocrystals, the defect electron is found to be more tightly bound as nanocrystal size decreases. This is largely due to the weak screening (Öğüt et al. 1997, 2003) in quantum dots. For nanocrystal up to 6 nm in diameter, the binding energy is significantly larger than kT at room temperature. A fit to the calculated binding energy E_B scales as $R^{-1.1}$. By extrapolating to a large radius, E_B tends to a limiting "bulk" value of 0.06 eV, which is somewhat higher than the experimental value of 0.044 eV; however, the difference is within the fitting error and likely fortuitously close since density functional theory was used to perform the computations.

2.1   Characterization of the Defect Wave Function

In Fig. 2a, the defect-state charge density is plotted along the [100] direction. As
the nanocrystal size increases, the defect wave function becomes more delocalized.
The maxima of the charge density at around 0.2 nm correspond to the bond length
between the P at the origin and its first Si neighbors. We can remove these atomic
details by spherically averaging the defect wave function as shown in Fig. 2b.
The defect wave function exponentially decays from the origin in Fig. 2c, and
the decay corresponds very well to the conventional understanding of defects in
semiconductors: the defect ion and the defect electron form a hydrogen-like system
with the wave function described as $\psi \sim \exp\left(-r/a_B^{\text{eff}}\right)$.

From the decay of the defect wave function, we can obtain an effective Bohr
radius a_B^{eff} and its dependence on nanocrystal size as plotted in Fig. 2d. The effective
Bohr radius varies nearly linearly with nanocrystal radius, R, up to 3 nm where R
is approximately five times a_B^{eff}. Nonlinearity can be observed for very small Si
nanocrystals, and a_B^{eff} appears to converge to ~ 0.2 nm as $R \to 0$. The limit tends
to the size of a phosphorus atom while retaining an sp^3 configuration. The effective
Bohr radius a_B^{eff} is not expected to remain linear with the radius R for large Si
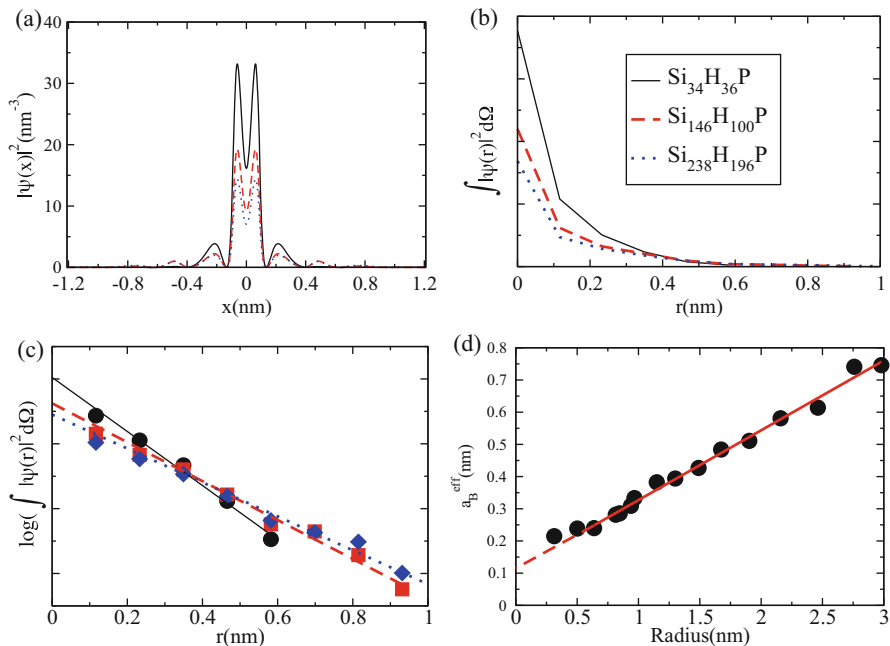
Fig. 2 (**a**) Charge density for the dopant electron along the [100] direction for three P-doped Si
nanocrystals with different radii. x is the coordinate along that direction. Their spherically averaged
charge densities are plotted in (**b**). A log-linear plot of (**b**) is illustrated in (**c**). Effective Bohr radius
a_B^{eff} corresponding to the dopant electron is plotted in (**d**) as a function of nanocrystal radius

nanocrystals. In the bulk limit, a_B^{eff} is ~2.3 nm assuming a dielectric constant of 11.4 and effective electron mass of 0.26 times the free electron mass. The dependence on R should achieve this bulk limit for large Si nanocrystals, which are removed from the size regime as shown in our figure.

2.2 Hyperfine Splitting of Defects

If the defect wave functions are known, one can compare to the experiment by computing the hyperfine splitting (HFS), which arises from an interaction of an electron spin with a nuclear spin (Feher 1959). The isotropic HFS constant a can be calculated from

$$a = \frac{2}{3}\mu_0 g \mu_B g_N \mu_N |\psi(0)|^2, \tag{4}$$

where μ_0 is the permeability of free space; g and g_N are the electron and the P nucleus g-factor, respectively; μ_B and μ_N are the Bohr and nuclear magneton; and $|\psi(0)|^2$ is the defect charge density evaluated at the P position (Weil and Bolton 2007). Pseudopotential calculations do not yield the true core charge density, which can differ strongly from the all-electron potentials within the core region. This deficiency can be overcome using a scheme from Van de Walle and Blöchl that utilizes the pseudopotential wave functions as input (Van de Walle and Blochl 1993). The correct amplitude of the wave function at the P core (positioned at the origin) can be obtained from

$$|\psi(0)|^2 = |\psi_{ps}(0)|^2 \frac{|\phi_{AE}^s(0)|^2}{|\phi_{ps}^s(0)|^2}, \tag{5}$$

where $\psi_{ps}(0)$ is the defect wave function evaluated at the P position using pseudopotential and $\phi_{AE}^s(0)$ and $\phi_{ps}^s(0)$ are the all-electron and pseudo-s-wave for an isolated P atom in free space. This formalism assumes that the impurity state is localized on the impurity atom, which need not be the case as discussed later in this review.

The calculated HFS is plotted in Fig. 3 for P atom located at the center of Si nanocrystal. Very good agreement between experimental data (Fujii et al. 2002) and the theoretical calculations is observed (Chan et al. 2008). For nanocrystals with radius between 2 to 3 nm, the HFS is around 100 G, which considerably exceeds the bulk value of 42 G. Moreover, the HFS continues to increase as the nanocrystal size decreases. This is a consequence of quantum confinement as illustrated in Fig. 2. The defect wave function becomes more localized at the P site as the radius decreases, leading to a higher amplitude of the wave function at the P nucleus. For small nanocrystals, the HFS can be significantly changed by atomic relaxation owing to the influence of a relatively large surface area.

Fig. 3 Calculated hyperfine splitting for P-doped Si nanocrystals as a function of nanocrystal radius is plotted together with experimental data(▲) from Fujii et al. (2002). Calculated results corresponding to structures arising from an unrelaxed exact bulk-like geometry(●) and to structures relaxed to a local energy minimum(♦) are shown in the figure
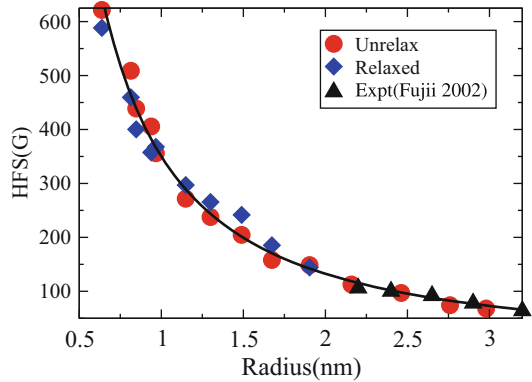
Effective mass theory has yielded similar results (Debernardi and Fanciulli 2009), although such methods cannot readily determine structural energies of the P substitutional sites. The most favorable structural position of the P impurity in a Si nanocrystal is problematic. Studies have verified that the symmetry of the P away from the surface is substitutional with the tetrahedral symmetry unbroken. After substituting the center Si atom by P, one can intentionally perturb the position of the P atom randomly such that the symmetry of the system was broken. However, on relaxation to a local energy minimum, the four P-Si bond lengths are preserved, and a tetrahedral geometry has been observed (Chan et al. 2008). Depending on the size of the nanocrystal, the P-Si bond varies from 0.236 to 0.238 nm (the calculated bulk Si-Si bond length is 0.236 nm) (Chan et al. 2008). Hence, the introduction of P atom does not break the symmetry of the system, and the relaxation is small, The P-Si bond length between different nanocrystals varies at most by 0.002 nm. The resulting HFS is illustrated in Fig. 3. For Si nanocrystals with radius larger than 1.5 nm, the effect of relaxation diminishes, and the difference between unrelaxed and relaxed results is within ∼10%.

We can consider the structural energetics of the doping site by varying the P position along the general [100] direction as illustrated in Fig. 4. Since the details of the nanocrystal surface are unknown, the P atom is not allowed to replace a Si on the surface of the nanocrystal that is passivated by a H atom. For the remaining sites, it is possible to relax the nanocrystals to an energy minimum. Results for five small nanocrystals are shown in Fig. 5. We see that for Si nanocrystals with a diameter smaller than ∼2 nm (the number of Si atoms is less than a hundred), P tends to substitute Si near the surface. Otherwise, the center of the nanocrystal is the energetically most favorable position. This suggests that a "critical size" exists for nanocrystals; below this size, P atoms will be energetically expelled toward the surface.

The presence of the P defect creates stress to the Si nanocrystal, and the doped nanocrystal can relieve this stress by expelling the P atom toward the surface, where surface relaxation or reconstruction may accommodate the stress. In contrast, if the P atom is not at the center of the nanocrystal, the defect wave function is distorted

Fig. 4 A perspective view of a Si nanocrystal. The lightly shaded atoms correspond to Si atoms, and the small atoms on the surface of the nanocrystal are H atoms. The heavily shaded atoms correspond to the Si atoms that will be substituted by P. The numbers measure the distance away from the origin in the unit of Si bond length
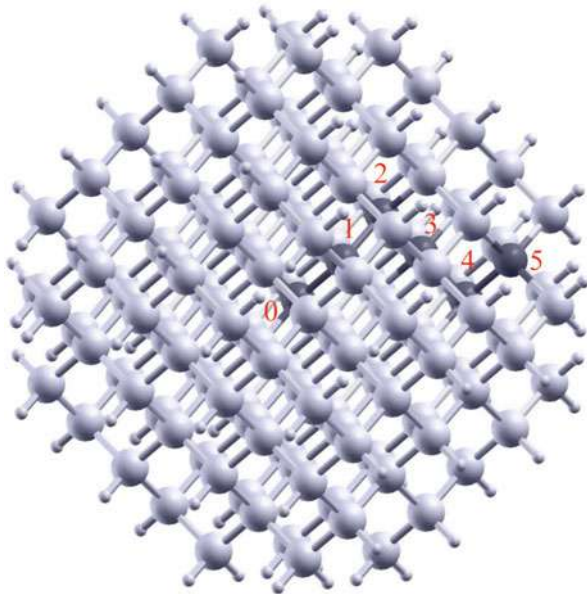
and less localized around P, leading to a loss in Coulomb energy between the P ion and the defect electron. Hence, for the slightly larger nanocrystals (the number of Si atoms more than around 150), a basin of attraction extending several Si bond lengths can be observed that favors a site for P atom near or at the center of the nanocrystal.

The HFS evaluated for different P positions inside the Si nanocrystal is shown in Fig. 5. There is a general trend for the HFS to drop notably as the P atom gets close to the surface of the nanocrystal. For a P site away from the surface, smaller variations in the HFS can be observed. However, for large Si nanocrystal, we can see that the HFS varies within ∼10% of the value with P at the center position. In comparing with the experimental results in Fig. 3, we find it unlikely that the synthesized Si nanocrystals have P located near the surface, as this will result in much smaller HFS than those calculated. P-doped Si nanocrystals are made experimentally by sputtering followed by annealing at high temperature (Fujii et al. 2003). The P atom resides in energetically favorable locations, which in general are close to the center of the Si nanocrystal, during the annealing process.

3   Self-Purification of Defects in Semiconductors

Doping of nanocrystals is often a difficult task (Stowell et al. 2003; Mikulec et al. 2000; Beaulac et al. 2010). "Self-purification" mechanisms are claimed to make this task even more difficult, i.e., a defect or impurity atom can be expelled from the nanocrystal to the surface leaving the interior of the nanocrystal defect-free (Dalpian and Chelikowsky 2006). Several reports in the literature support this thesis as there
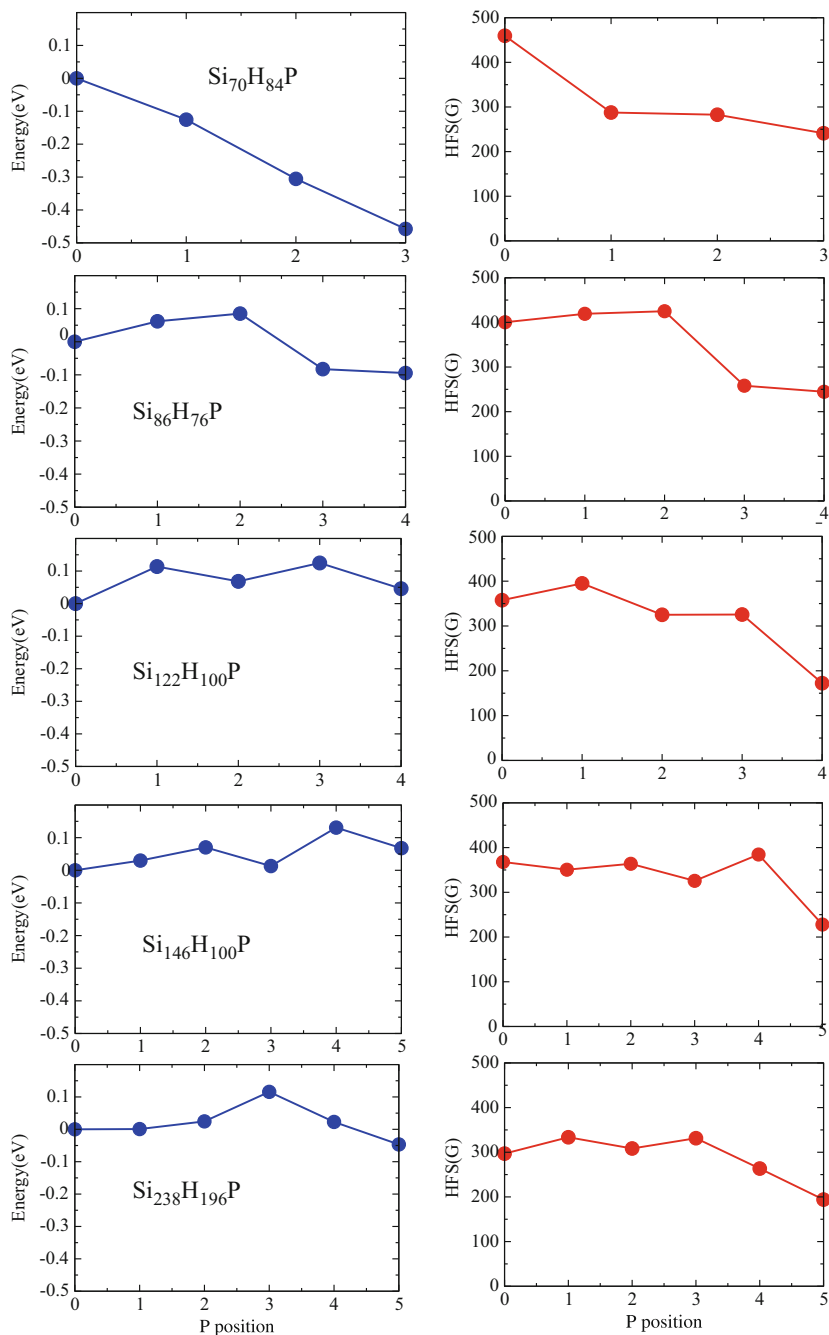
Fig. 5 Difference in energy (left) and HFS (right) as the P atom moves away from the center of the Si nanocrystal. The energies are with respect to the energy of the Si nanocrystal with P at the center. The x-axis measures the distance of P atom away from the origin as illustrated in Fig. 4

is a tendency for impurities to cluster at the surface of the nanocrystals. To avoid this situation, specialized synthesis procedures have been developed to inhibit the diffusion of a dopant to the surface (Pradhan et al. 2005; Yang et al. 2008; Nag et al. 2008).

Computational work regarding theoretical mechanisms for doping at the nanoscale remains somewhat limited owing to the many degrees of freedom involved in any simulations. One proposal suggests that the incorporation of impurities is predominantly limited by the kinetics of the adsorption on the surface of the nanocrystals (Erwin et al. 2005). Another proposal focuses on thermodynamic effects in order to understand difficulty in doping in nanocrystals (Dalpian and Chelikowsky 2006; Chelikowsky et al. 2011).

Under thermodynamical equilibrium, the population of impurities or defects in a given host material is given by

$$N = N_s \exp\{(G_f/k_bT)\} \tag{6}$$

where N_s is the total number of impurity sites in the lattice, G_f is the Gibbs free formation energy of the impurity, k_b is the Boltzmann constant, and T is the temperature. Assuming that the entropic contribution will be similar for several configurations of a similar defect, one can substitute G_f by the formation energy of the defect: ΔH_f.

Consider the case of a Mn atom replacing a Cd atom in CdSe as an example appropriate for synthesizing a dilute magnetic semiconductor. The formation energy of a substitutional Mn impurity in a host CdSe material is a function of the elemental chemical potentials of the impurity and the host constituents and also a function of the Fermi energy of the system (Zhang and Northrup 1991; Wei 2004).

When using ab initio total energy methods, this formation energy can be calculated as follows:

$$\Delta H_f = E_T^{Mn} - E_T + (E_{Cd} + \mu_{Cd}) - (E_{Mn} + \mu_{Mn}), \tag{7}$$

where E_T^{Mn} is the total energy of the host with the impurity, E_T is the total energy of the host without the impurity, E_{Cd} and E_{Mn} are the elemental energies of Cd and Mn, and μ_i is related to the range of the chemical potentials in different thermodynamic limits.

For nanoscale materials, such as nanocrystals, the number of available sites where the impurities can be localized (N_s) can be as much as 20 orders of magnitude smaller than in standard bulk materials. Consequently, as the number of available sites for the impurities to be localized is much smaller than bulk, we would expect that nanostructures should have a smaller number of impurities than bulk materials. This can be understood by considering that N_s in Eq. 6 is small. For example, if the probability of defect in a bulk crystal is 1 in 10,000, the probability of a hundred-atom nanocrystal having such a defect would be small. This was first proposed by Turnbull in 1950, many decades before the advent of nanoscience and nanotechnology. Although this is a simplistic approach, e.g., it ignores the role of

a surface altogether, it represents one of the first attempts to explain why "small" systems might have a negligible probability of containing defects.

Besides the number of available sites, the formation energy (H_f) can also play a role in the population of defects in a nanocrystal. The formation energy of a certain impurity depends on the position of the impurity inside the nanostructure, i.e., it varies from the surface of the nanostructure or at its core (as indicated for P in Si in Fig. 5). The formation energy also depends on the size of the nanostructure. Quantum confinement effects, which are larger in smaller nanocrystals, can lead to an increased formation energy, leading to smaller impurity concentrations. For example, quantum confinement effects are observed in DX$^-$ centers in GaAs quantum dots (Li et al. 2005; Besteiro et al. 2010); B- and P-doping in Si nanoclusters (Cantele et al. 2005) and nanowires (Peelaers et al. 2006); donors, acceptors, and isovalent impurities in ZnSe (Li et al. 2008); Mn impurities in InAs nanocrystals (Arantes et al. 2008); and Mn in CdSe nanocrystalsChan et al. (2009). Similar effects have also been reported for B and N impurities in SiC nanowires (Santos de Oliveira and Miwa 2009) and Mn impurities in InP nanowires (Schmidt 2008).

For the specific case of Mn impurities in CdSe nanocrystals, calculations have indicated a small tendency for impurities to be localized at the surface of the nanocrystals (Chan et al. 2009). For example, an energy gain of 0.02 eV was found for a Mn atom moving from the center of the nanocrystal toward its surface. This trend is observed in small nanocrystals ($Cd_{30}Se_{28}$) and for larger nanocrystals ($Cd_{79}Se_{68}$). For the larger nanocrystals, an structural anisotropy for the stability of the impurities was observed. The energy gain is larger when the impurity moves along the $\langle 100 \rangle$ direction than when it moves along the $\langle 111 \rangle$ direction. This anisotropy occurs owing to a more effective strain relief in the $\langle 111 \rangle$ direction. By construction, the nanocrystals used in these calculations (Chan et al. 2009) present a dipole moment in the $\langle 111 \rangle$ direction, whereas this does not happen in the $\langle 100 \rangle$ direction. This dipole moment can affect the energetics of the Mn defect inside the nanocrystal. In Fig. 6a, we show the calculated self-consistent potentials along two different directions inside the nanocrystal. As can be observed, this potential is anisotropic along the $\langle 111 \rangle$ direction, which is the cause for the difference in the energetics of the impurities along these two directions. It is possible that impurities could be manipulated by subjecting these systems to a directed electric field (Chan et al. 2009).

In Fig. 6 the energy difference for some of the studied positions for the Mn atom inside the nanocrystal is illustrated in different directions. The insets (a) and (b) show the direction in which the positions are tested. In (c) we show the variation of the total energy for these positions (Chan et al. 2009).

Although the fact that the Mn atom prefers to be localized at the surface of the nanocrystal clearly indicates that at equilibrium the impurities would all tend to be localized at these positions, i.e., would be expelled from the nanocrystal, this diffusion of the impurity into the surface might be kinetically inhibited. This might happen because the diffusion barrier for the Mn atom to move toward the surface might be large, and it would take a long time for a single jump to occur. Diffusion
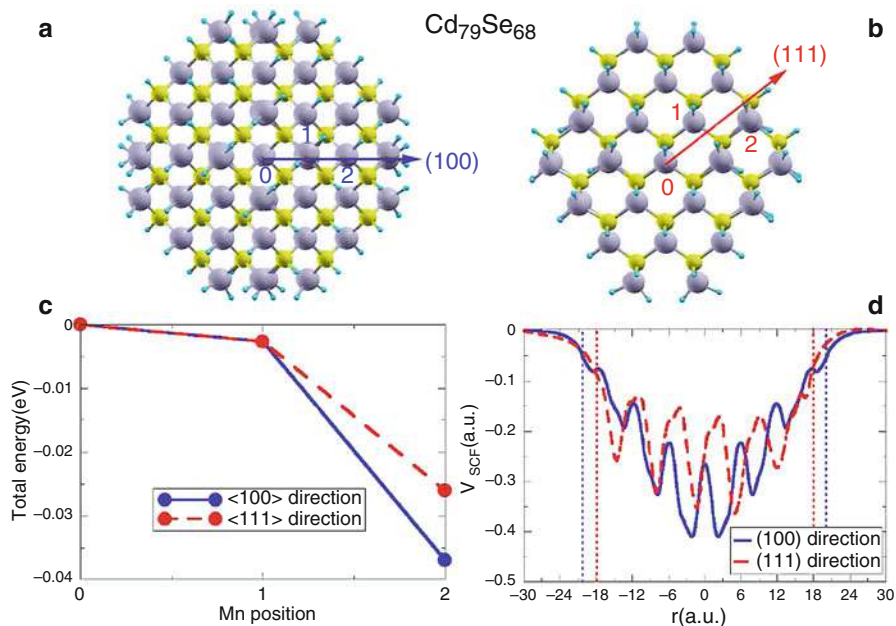
Fig. 6 (**a**) and (**b**) An atomic model of a $Cd_{79}Se_{68}$ nanocrystal in two different perspectives. The Cd atoms are colored gray, Se in yellow, and the small blue atoms at the surface correspond to the capping atoms. The number labels the Cd atom to be substituted by Mn. (**c**) The change in total energy of the $Cd_{78}Se_{68}Mn$ nanocrystal as Mn moves toward the surface. (**d**) A plot of the self-consistent potential V_{SCF} (averaged in the plane perpendicular to the plotting direction). The troughs in the potential correspond to Se atoms, while the peaks correspond to the Cd atoms for the (100) direction and the shoulders close to the peaks are the positions of the Cd atoms for the (111) direction. The dotted lines indicate the boundary of the nanocrystal

barriers for the bulk material are usually of the order of a few eV, leading to diffusion hopping times as large as 1 h (Norris et al. 2008). To assess this problem for nanoscale materials, the diffusion barrier for a Mn atom for a bulk CdSe crystal and inside a CdSe nanocrystal have been calculated (Chan et al. 2009). The mechanism involving the diffusion of a Mn can be quite complex, including *kickout* mechanisms and exchange of atoms. However, previous work suggests that the diffusion of Mn in bulk GaAs and CdSe involves interstitial defects (Edmonds et al. 2004). In Fig. 7, we illustrate the calculated barrier for interstitial diffusion in bulk CdSe and for the $(Cd_{30}Se_{28})$ nanocrystal. This diffusion barrier is related to the trajectory of a Mn atom from a tetrahedral interstitial position, where the Mn atom is surrounded by four Cd atoms (T_{Cd}) to an interstitial position where there are four Se (T_{Se}) atoms around the Mn atom. Each intermediate structure was obtained by fixing one atom in the nanocrystal and the x coordinate of the Mn atom: all other coordinates were optimized. A total of 11 geometries were used in order to obtain the diffusion barrier and the saddle point in each path. The diffusion barrier for the nanocrystals was calculated to be 0.35 eV, and for bulk we obtain a value of 0.84 eV. As the diffusion

Fig. 7 Diffusion barriers of
the Mn atom for the $Cd_{30}Se_{28}$
nanocrystal and for bulk
CdSe. The ball and stick
models at the bottom indicate
the diffusion pathway
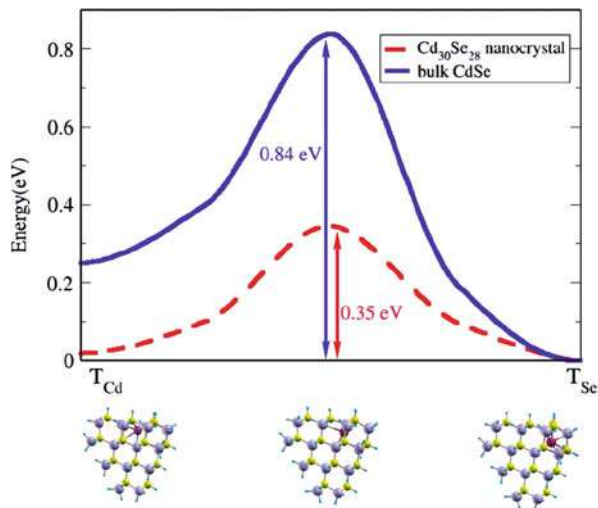
barrier in nanocrystals is dramatically smaller than in bulk, the expulsion of the
defect to the surface might easily take place in nanocrystals. Another interesting
observation from Fig. 7 is that the Mn atom prefers to be localized at the T_{Se} position
as opposed to the T_{Cd} site for the bulk limit. This preference disappears in the
nanocrystals owing to the larger localization of the Mn_d levels in quantum confined
systems.

Studies have also revealed that there might be a significant change in the stability
of an impurity inside a nanocrystal as the size of the nanocrystal changes (Dalpian
and Chelikowsky 2006). We have observed that, for smaller nanocrystals, where
quantum confinement effects are stronger, the formation energy of an impurity
is usually larger than in bulk or in larger nanocrystals. In Fig. 8, we illustrate the
variation of the formation energy, compared to bulk value, for several different types
of impurities in different host materials taken from several different sources. As can
be observed in these examples, the formation energy usually increases (becomes
less favorable) when the size of the nanocrystal decreases. For some impurities
this effect is larger than for others, such as when large valence differences exist,
although this trend is present in all defects shown. As this effect is observed in
several different kinds of impurities, its origin might be related to several different
effects. One of the most important effects is related to the increased localization of
the impurity levels in smaller nanocrystals. This would increase the energetic cost
to form these defects as the size of the nanocrystal decreases.

4   Raman Spectroscopy of Nanocrystals

Raman spectroscopy is a particularly useful technique for characterizing the prop-
erties of nanostructured Si materials (Li et al. 2000; Volodin et al. 1998; Nesheva
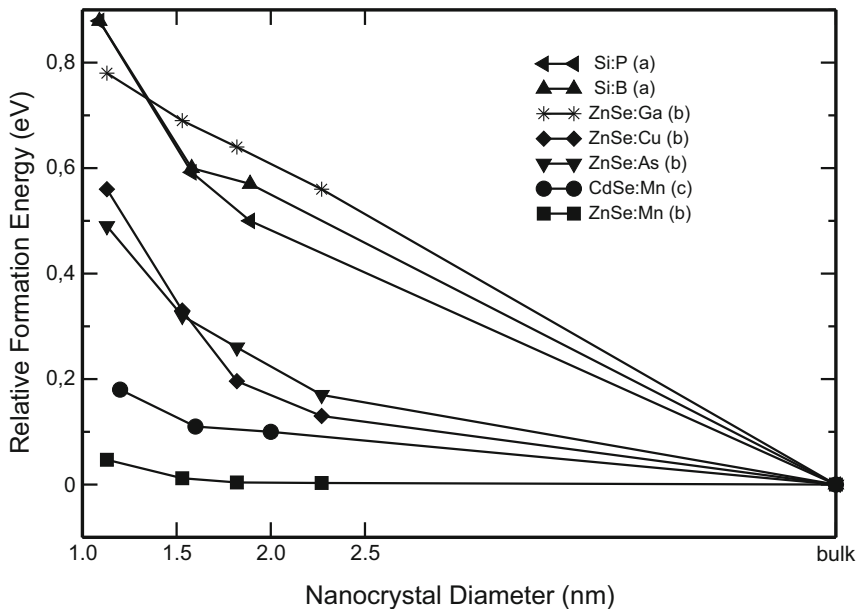
Fig. 8 Variation of the formation energy of several impurities as the size of a nanocrystal is changed. (**a**) Cantele et al. (2005), (**b**) Li et al. (2008), (**c**) Dalpian and Chelikowsky (2006)

et al. 2002) as it is a well-established experimental technique for characterizing a wide range of materials. It is highly sensitive to both the atomic and electronic structure of the substrate. Consequently, materials have unique Raman signatures, which can be used in identifying different species.

Inelastic scattering of light, or Raman scattering, is conceptually simple. When light interacts with matter, such as a molecule, most of the light is scattered elastically. This is called Rayleigh scattering. A small fraction, (10^{-6}), either gains or loses energy during the collision by moving the molecule into a different vibrational mode. The shift in wavelength corresponds to the energetic difference between the vibrational modes. Generally, at ambient temperatures, the majority of molecules are in the lowest energy mode and are excited to a higher mode by the photons. As a result, the exiting radiation is lower energy than the incident radiation. This is called Stokes Raman and is typically more prevalent than the opposite effect, anti-Stokes Raman.

Not every vibrational mode is Raman active. A mode is only Raman active if the derivative of the polarizability with respect to the normal coordinate is nonzero at the equilibrium point. As a rule of thumb, stretching modes that lengthen bond lengths and change the size of the molecule are generally Raman active. Symmetric bending modes that do not change bond length are usually not Raman active (Li et al. 2000; Volodin et al. 1998; Nesheva et al. 2002).

Raman spectroscopy has many advantages as an experimental technique. Raman analysis can be performed on samples in gas, liquid, or solid phase. It is useful for biological materials and can be performed on aqueous solutions without interference from the water solvent. One of the foremost disadvantages of Raman spectroscopy is that the signal strength is very weak.

Si nanostructures can be optically active by demonstrating a strong visible photoluminescence (Zhang et al. 1995, 1994; Wolkin et al. 1999; Furukawa and Miyasato 1988) in striking contrast to the bulk phase, making them suitable for use as optoelectronic devices (Pavesi et al. 2000). Also, their compatibility with existing Si-based technologies enhances their utility in nanoscale electronics applications (Tiwari et al. 1996; Choi et al. 1998). In addition to changes in the optical properties of bulk materials at the nanoscale, changes in the vibrational modes of a nanocrystal also exist owing to a physically constrained nanostructure. An analysis of Raman spectroscopy centers on understanding a combination of changes in the optical and vibrational properties of Si nanostructures. A striking feature of the Raman spectra of Si nanocrystals is that the main optical Raman peak red-shifts and broadens as the nanocrystal size is decreased.

To explain such observations, Richter, Wang, and Ley (RWL) proposed a phenomenological model attributing these confinement effects to a relaxation of the wave-vector selection rule (Richter et al. 1981). This model is widely accepted and used extensively in the calculation of Raman spectra and interpretation of experimental data (Wei et al. 2007; Faraci et al. 2005, 2006; Valentin et al. 2008). More sophisticated techniques such as the partial density method and bond polarizability approximation have also been used to study the Raman spectra of Si nanocrystals (Zi et al. 1996, 1997, 1998; Cheng and Ren 2002). These techniques give better results than the RWL theory; however, they still rely on parameters derived from bulk calculations, and their transferability is not guaranteed.

The two primary pieces of information needed to calculate a Raman spectrum are the vibrational modes and the dielectric susceptibility, both of which can be calculated from first principles using pseudopotentials and density functional theory. The computational methods used for the Raman work outlined here are discussed elsewhere in this handbook (Chap. 24 "Extending the Scale with Real-Space Methods for the Electronic Structure Problem" of the MTM set by Chelikowsky) and in the literature (Steele 1971; Wilson et al. 1980; Khoo et al. 2010; Bobbitt and Chelikowsky 2016a, b). The vibrational properties of nanocrystals can be computed by the force constant method (Parlinski et al. 1997), and Raman spectra were evaluated using the Placzek approximation (Porezag and Pederson 1996; Umari and Pasquarello 2005; Parlinski et al. 1997; Bobbitt and Chelikowsky 2016a, b).

We can illustrate the vibrational modes of Si nanocrystals by computing the dynamical matrix. As an example, we consider Si nanocrystals $Si_{87}H_{76}$, $Si_{147}H_{100}$, and $Si_{281}H_{163}$ having diameters 1.4, 1.7, and 2.2 nm, respectively, that were selected for our study. These nanocrystals were generated by cutting atom-centered spheres out of bulk Si and passivating dangling bonds on their surfaces with H atoms. The nanocrystals are illustrated in Fig. 9.

We plot the calculated Si vibrational density of states (VDOS) for each nanocrystal in Fig. 10. VDOS contributions from passivating H atoms have been excluded

Si$_{87}$H$_{76}$ (1.4 nm)   Si$_{147}$H$_{100}$ (1.7 nm)   Si$_{281}$H$_{172}$ (2.2 nm)
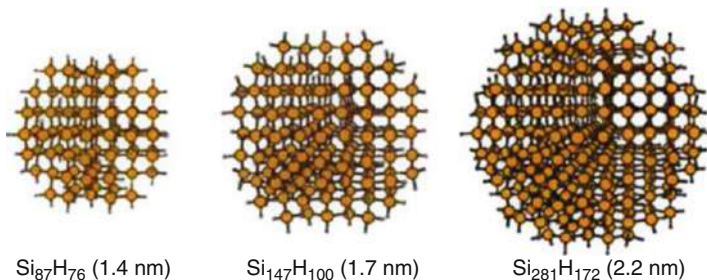
Fig. 9   Ball and stick model for relaxed structures of H-passivated Si nanocrystal

Fig. 10   Si VDOS per atom for Si nanocrystals and bulk Si
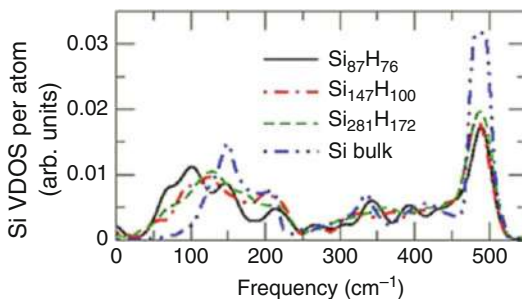
as they are irrelevant to the problem. We also computed the VDOS of bulk Si to represent the limit of infinite nanocrystal size.

There are two prominent peaks in the Si VDOS, a low-frequency acoustic peak centered in the range of 100–150 cm^{-1} and a high-frequency optical peak in the vicinity of 480 cm^{-1}. As the nanocrystal size gets smaller, the acoustic peak is red-shifted, and the amplitude of the high-frequency optical peak is decreased. Both these observations point toward a shift of VDOS to lower frequencies with decreasing nanocrystal size. The total shift in VDOS frequencies is due to the competing influences of surface atom under-coordination and bond length contraction (Khoo et al. 2010). Under-coordinated Si surface atoms have a greater impact on nanocrystal vibrational frequency than bond length contractions. It can be argued that the red shift of Si VDOS for nanocrystals with decreasing nanocrystal size (as seen in Fig. 10) is due to their larger surface-to-volume ratios, as this leads to a larger proportion of under-coordinated surface atoms that lower vibrational frequencies.

In Fig. 11, we plot the calculated Raman cross sections as a function of frequency for the Si$_{87}$H$_{76}$, Si$_{147}$H$_{100}$, and Si$_{281}$H$_{172}$ nanocrystals with a broadening of 2.5 cm^{-1}. The main Raman peak in the range of 500–520 cm^{-1} is asymmetrically broadened and red-shifted with decreasing nanocrystal diameters, consistent with experimental observations. Our calculated optical mode frequency for bulk Si is 522 cm^{-1}, as indicated by the dashed line, and it is in good agreement with experimental values (Iqbal and Veprek 1982). Among the nanocrystals we have
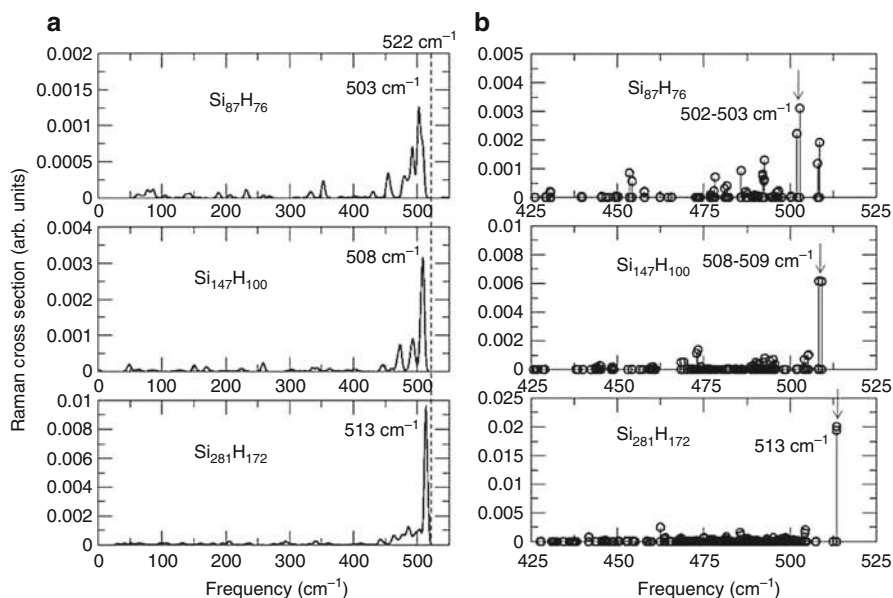
Fig. 11 Calculated Raman cross sections for Si nanocrystals with (**a**) broadening of 2.5 cm^{-1} and (**b**) no broadening. Dotted line in (**a**) represents frequency of Raman peak position for bulk Si, and arrows in (**b**) denote nanocrystal eigenmodes corresponding to the Γ-point bulk optical mode

studied, only $Si_{281}H_{172}$ (diameter 2.2 nm) is large enough to be compared directly with the experiment, and our calculated peak frequency of 513 cm^{-1} agrees well with the experimentally measured value of 514 cm^{-1} for nanocrystalline Si of diameter 2.3 nm (Ossadnik et al. 1999).

Also in Fig. 11, we plotted the Raman cross-sectional contribution from each nanocrystal mode as a function of mode frequency without broadening. It can be seen that the number of vibrational modes that contribute to the Raman spectra increases as the nanocrystal diameter is decreased and this leads to an asymmetric broadening and partial red shift of the Raman peak. These observations are consistent with the RWL theory, which predicts an increasing number of vibrational modes contributing to the Raman spectra for decreasing nanocrystal diameters. However, it is also clear from the figure that most of the Raman red shift does not originate from the activation of lower-frequency modes. Peak frequencies in the broadened Raman spectra of Fig. 11 follow closely the frequencies of three nearly degenerate vibrational modes with the largest Raman cross-sectional contributions (marked by arrows). By projecting these eigenmodes onto bulk phonon eigenvectors, we have identified them as originating from the optical modes in bulk, and it is the lowering of these mode frequencies with decreasing nanocrystal size that leads to the Raman red shift. The softening of these eigenmodes results from the presence of under-coordinated Si atoms on the nanocrystal surface, as noted before. Our explanation represents a qualitative departure from the RWL theory, which attributes

Fig. 12 The vibrational density of states for $Si_{87}H_{76}$ and $Si_{86}H_{76}P$ nanocrystals with P-dopant near the surface and center positions
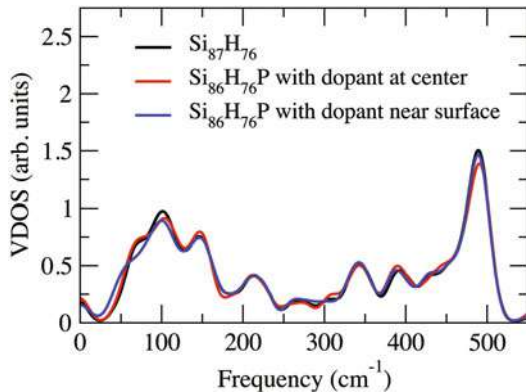
Table 1 Force constants and atomic masses of the Si and P atoms at the center of the $Si_{87}H_{76}$ and $Si_{86}H_{76}P$ nanocrystals

	Force constant (eV/Å^2)	Atomic mass (amu)
Si	14.3	28.1
P	12.9	31.0

the Raman red shift to a relaxation of the wave-vector selection rule. Instead, the size dependence of nanocrystal vibrational frequencies has a much stronger effect on the Raman peak frequencies in Si nanocrystals.

Suppose we replace a Si atom by a P in a Si nanocrystal. We can also compute the Raman spectra for different sites and assess whether the Raman spectra change. In Fig. 12, we illustrate the vibrational density of states results for P at different sites within the $Si_{87}H_{76}$ nanocrystal. The results may seem odd as nothing of note happens, i.e., most of the vibrational modes remain unchanged. We can compute the force constants for P and Si atoms in equivalent sites, and to verify this is the case. For the center atom of the nanocrystal, the force constants associated with an infinitesimal displacement of a Si atom and a P atom are listed in Table 1, along with the corresponding atomic masses.

The force constants are very similar for both atoms. The P constant is smaller and we expect that. First, the mass of the atom is slightly higher than Si. Second, P has an extra electron. There is no bonding orbital for the extra P electron; all the bonds are saturated. This weakens the bonds between the P atom and neighboring Si atoms. This can indeed be verified on close inspection of the peaks at 100 and 480 cm^{-1} in Fig. 12. If one moves the P atom from the center of the nanocrystal to the surface; we do not find significant differences. As a consequence, the vibrational modes do not offer a means to either detect or locate the P atom. This is not the situation for the Raman spectra (Khoo and Chelikowsky 2014). In Fig. 13, the Raman spectra are illustrated with and without a P-dopant. Two dopant cases are presented with the P atom at the center of the nanocrystal and near the surface.

The Raman spectra for the undoped case are dominated by a single prominent peak at 503 cm^{-1} originating from the so-called Γ-point optical mode. This

Fig. 13 Raman cross section
for $Si_{87}H_{76}$ and $Si_{86}H_{76}P$
nanoclusters with P-dopant at
the center and near the
surface
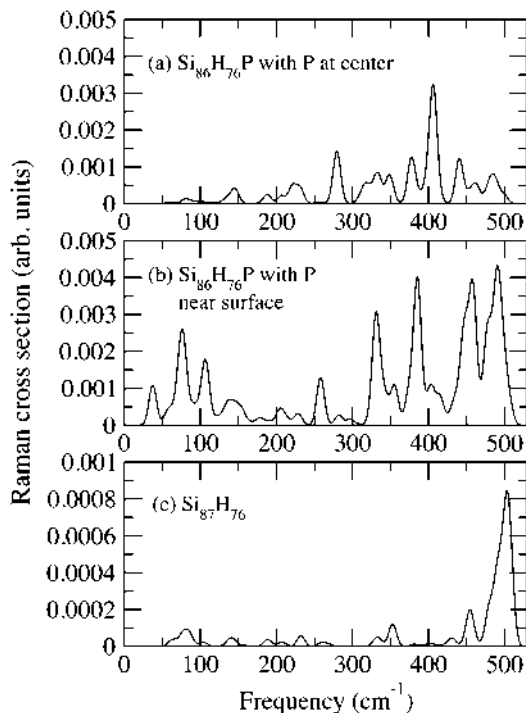
frequency is slightly red-shifted relative to the bulk-Si value of $522\,cm^{-1}$ owing
to vibration softening and selection rule effects in a finite system.

The introduction of a P-dopant leads to qualitative changes in the calculated
Raman spectra. The number and positions of the peaks are changed depending on
the position of the dopant in the nanocrystal. For the center-doped nanocrystal, the
main feature in the spectrum is a peak at about $400\,cm^{-1}$. For the surface-doped
case, there are multiple peaks appearing at the frequencies 70, 320, 380, 450, and
$480\,cm^{-1}$.

The electric susceptibility and its derivative with respect to nuclear displacements
play an important role in understanding Raman spectra. This assumption is consis-
tent with the increase in magnitude of the Raman spectra, as we expect P-doping to
increase the responsiveness of Si to external electric fields. Raman can be used as a
tool to predict where dopants may reside, whereas the vibrational modes cannot.

5   Summary

As the physical size of electronic materials decreases, quantum-related properties
emerge. One goal of simulation and computation is to consider systems of increasing
size and complexity to address how properties emerge. Advances in physical

concepts, algorithms, and computational hardware allow us to consider larger and larger systems. Conversely, experimentalists can synthesize and characterize smaller and smaller nanostructures. These trends have "crossed" for many electronic materials allowing theorists to simulate systems that are larger than the smallest ones made by experimentalists. This cross-point will have significant long-term consequences for understanding quantum mechanical phenomena that results from the physical confinement of the system. As Feynman noted such an understanding is important as the design rules that work successfully at macroscopic dimensions may not be operative at nanoscale dimensions.

References

Arantes JT, Dalpian GM, Fazzio A (2008) Quantum confinement effects on Mn-doped InAs nanocrystals: a first-principles study. Phys Rev B 78:045402

Beaulac R, Ochsenbein ST, Gamelin DR. (2010) Colloidal transition-metal-doped quantum dots, chap. 11. In: Klimov VI (ed) Nanocrystal quantum dots, 2nd edn. (CRC Press)

Besteiro LV, Tortajada L, Tiago ML, Gallego LJ, Chelikowsky JR, Alemany MMG (2010) Efficient n-type doping of zinc-blende III-V semiconductor quantum dots. Phys Rev B 81:121307

Bobbitt NS, Chelikowsky JR (2016a) First-principles study of vibrational properties and Raman spectra in Li-doped Si nanocrystals. Chem Phys Lett 646:136

Bobbitt NS, Chelikowsky JR (2016b) Real-space pseudopotential study of vibrational properties and Raman spectra in Si-Ge core-shell nanocrystals. J Chem Phys 144:124110

Canham L (2000) Gaining light from silicon. Nature 408:411

Car R, Parrinello M (1985) Unified approach for molecular dynamics and density functional theory. Phys Rev Lett 55:2471

Cantele G, Degoli E, Luppi E, Magri R, Ninno D, Iadonisi G, Ossicini S (2005) First-principles study of n- and p- doped silicon nanoclusters. Phys Rev B 72:113303

Chadi DJ (1994) Doping in ZnSe, ZnTe, MgSe, and MgTe wide-band-gap semiconductors. Phys Rev Lett 72:534

Chan T-L, Tiago ML, Kaxiras E, Chelikowsky JR (2008) Size limits on doping phosphorus into silicon nanocrystals. Nano Lett 8:596

Chan T-L, Zayak AT, Dalpian GM, Chelikowsky JR (2009) Role of confinement on diffusion barriers in semiconductor nanocrystals. Phys Rev Lett 102:025901

Chang KJ, Dacorogna MM, Cohen ML, Mignot JM, Chouteau G, Martinez G (1985) Superconductivity in high-pressure metallic phases of Si. Phys Rev Lett 54:2375

Chelikowsky JR, Alemany MMG, Chan TL, Dalpian GM (2011) Computational studies of doped nanostructures. Rep Prog Phys 74:046501

Cheng W, Ren SF (2002) Calculations on the size effects of Raman intensities of silicon quantum dots. Phys Rev B 65:205305

Choi BH, Hwang SW, Kim IG, Shin HC, Kim Y, Kim EK (1998) Fabrication and room-temperature characterization of a silicon self-assembled quantum-dot transistor. Appl Phys Lett 73:3129

Dalpian G, Chelikowsky JR (2006) Self-purification in semiconductor nanocrystals. Phys Rev Lett 96:226802

Debernardi A, Fanciulli M (2009) Confinement effect in P doped spherical Si nanocrystals. Solid State Sci 11:961

Edmonds KW, Bogusławski P, Wang KY, Campion RP, Novikov SN, Farley NRS, Gallagher BL, Foxon CT, Sawicki M, Dietl T, Buongiorno Nardelli M, Bernholc J (2004) Mn interstitial diffusion in (Ga,Mn)As. Phys Rev Lett 92:037201

Ellingson RJ, Beard MC, Johnson JC, Yu PR, Micic O, Nozik AJ, Shabaev A, Efros AL (2005) Highly efficient multiple exciton generation in colloidal PbSe and PbS quantum dots. Nano Lett 5:865

Erwin SC, Zu LJ, Haftel MI, Efros AL, Kennedy TA, Norris DJ (2005) Doping semiconductor nanocrystals. Nature 436:91

Faraci G, Gibilisco S, Russo P, Pennisi AR, Compagnini G, Battiato S, Puglisi R, La Rosa S (2005) Si/SiO_2 core shell clusters probed by Raman spectroscopy. Eur Phys J B 46:457

Faraci G, Gibilisco S, Russo P, Pennisi AR, La Rosa S (2006) Modified Raman confinement model for Si nanocrystals. Phys Rev B 73:033307

Feher G (1959) Electron spin resonance experiments on donors in silicon. I. Electronic structure of donors by the electron nuclear double resonance technique. Phys Rev 114:1219

Feynman RP (1992) There's plenty of room at the bottom (data storage). J Microelectromech Syst 1:60–66

Fujii M, Mimura A, Hayashi Sh, Yamamoto Y, Murakami K (2002) Hyperfine structure of the electron spin resonance of phosphorus-doped Si nanocrystals. Phys Rev Lett 89:206805

Fujii M, Toshikiyo K, Takase Y, Yamaguchi Y, Hayashi S (2003) Below bulk-band-gap photoluminescence at room temperature from heavily P- and B-doped Si nanocrystals. J Appl Phys 94:1990

Furukawa S, Miyasato T (1988) Quantum size effects on the optical band gap of microcrystalline Si:H. Phys Rev B 38:5726

Gur I, Fromer NA, Geier ML, Alivisatos AP. (2005) Air-stable all-inorganic nanocrystal solar cells processed from solution. Science 310:462

Iqbal Z, Veprek S (1982) Raman scattering from hydrogenated microcrystalline and amorphous silicon. J Phys C Solid State Phys 15:37

Kittel C (2005) Introduction to solid state physics, 8th edn. Wiley, Hoboken

Khoo KH, Chelikowsky JR (2014) First-principles study of vibrational modes and Raman spectra in P-doped Si nanocrystals. Phys Rev B 89:195309

Khoo KH, Zayak AT, Kwak H, Chelikowsky JR (2010) First-principles study of confinement effects on the Rahman spectra of Si nanocrystals. Phys Rev Lett 105:115504

Klimov VI, Ivanov SA, Nanda J, Achermann M, Bezel I, McGuire JA, Piryatinski A (2007) Single-exciton optical gain in semiconductor nanocrystals. Nature 447:441

Li GH, Ding K, Chen Y, Han HX, Wang ZP (2000) Photoluminescence and raman scattering of silicon nanocrystals prepared by silicon ion implantation into SiO_2 films. J App Phys 88:1439

Li J , Wei SH, Wang LW (2005) Stability of the DX^- center in GaAs quantum dots. Phys Rev Lett 94:185501

Li J, Wei SH, Li SS, Xia JB (2008) Origin of the doping bottleneck in semiconductor quantum dots: a first-principles study. Phys Rev B 77:113304

Melnikov DV, Chelikowsky JR (2004) Quantum confinement in phosphorus-doped silicon nanocrystals. Phys Rev Lett 92:46802

Michalet X, Pinaud FF, Bentolila LA, Tsay JM, Doose S, Li JJ, Sundaresan G, Wu AM, Gambhir SS, Weiss S. (2005) Quantum dots for live cells, in vivo imaging, and diagnostics. Science 307:538

Mikulec FV, Kuno M, Bennati M, Hall DA, Griffin RG, Bawendi MG (2000) Organometallic synthesis and spectroscopic characterization of manganese-doped CdSe nanocrystals. J Am Chem Soc 122:2532

Nag A, Chakraborty S, Sarma DD (2008) To dope Mn^{2+} in a semiconducting nanocrystal. J Am Chem Soc 130:10605

Nesheva D, Raptis C, Perakis A, Bineva I, Aneva Z, Levi Z, Alexandrova S, Hofmeister H (2002) Raman scattering and photoluminescence from si nanoparticles in annealed SiO_x thin films. J Appl Phys 92:4678

Neumark GF (1997) Defects in wide band gap II-VI crystals. Mater Sci Eng R 21:1–46

Norris DJ, Efros AL, Erwin SC (2008) Doped nanocrystals. Science 319:1776

Ossadnik C, Veprek S, Gregora I (1999) Applicability of Raman scattering for the characterization of nanocrystalline silicon. Thin Solid Films 337:148

Öğüt S, Chelikowsky JR, Louie SG (1997) Quantum confinement and optical gaps in Si nanocrystals. Phys Rev Lett 79:1770

Öğüt S, Burdick R, Saad Y, Chelikowsky JR (2003) Ab Initio calculations for large dielectric matrices of confined systems. Phys Rev Lett 90:127401

Parlinski K, Li ZQ, Kawazoe Y (1997) First-principles determination of the soft mode in cubic ZrO_2. Phys Rev Lett 78:4063

Pavesi L, Turan R (2010) Silicon nanocrystals, 1st edn. Wiley, Berlin

Pavesi L, Dal Negro L, Mazzoleni C, Franzo G, Priolo F (2000) Optical gain in silicon nanocrystals. Nature 408:440

Peelaers H, Partoens B, Peeters FM (2006) Formation and segregation energies of B and P doped and BP codoped silicon nanowires. Nano Lett 6:2781

Porezag D, Pederson MR (1996) Infrared intensities and Raman-scattering activities within density-functional theory. Phys Rev B 54:7830

Pradhan N, Goorskey D, Thessing J, Peng X (2005) An alternative of CdSe nanocrystal emitters: pure and tunable impurity emissions in ZnSe nanocrystals. J Am Chem Soc 127:17586

Richter H, Wang ZP, Ley L (1981) The one phonon Raman spectrum in microcrystalline silicon. Solid State Commun 39:625

Santos de Oliveira IS, Miwa RH (2009) Boron and nitrogen impurities in SiC nanowires. Phys Rev B 79:085427

Schmidt TM (2008) Surface effects on the energetic and spintronic properties of InP nanowires diluted with Mn: first-principles calculations. Phys Rev B 77:085325

Shirley E, Chelikowsky JR, Louie SG, Martinez G (eds) (2000) Optical properties of materials. In: Materials research society symposium, vol 579

Steele D (1971) Theory of vibrational spectroscopy. Saunders, Philadelphia

Stowell CA, Wiacek RJ, Saunders AE, Korgel BA (2003) Synthesis and characterization of dilute magnetic semiconductor manganese-doped indium arsenide nanocrystals. Nano Lett 3:1441

Tiwari S, Rana F, Hanafi H, Hartstein A, Crabbe EF, Chan K (1996) A silicon nanocrystals based memory. Appl Phys Lett 68:1377

Toumey C (2005) Apostolic succession. Eng Sci 68:22

Troparevsky MC, Chelikowsky JR (2001) Structural and electronic properties of CdS and CdSe clusters. J Chem Phys 114:943

Turnbull D (1950) Formation of crystal nuclei in liquid metals J App Phys 21:1022

Umari P, Pasquarello A (2005) Infrared and Raman spectra of disordered materials from first principles. Diamond Relat Mater 14:1255

Valentin A, See J, Galdin-Retailleau S, Dolfus P (2008) Study of phonon modes in silicon nanocrystals using the adiabatic bond charge model. J Phys Condens Matter 20:145213

Van de Walle CG, Blochl PE (1993) First-principles calculations of hyperfine parameters. Phys Rev B 47:4244

Volodin VA, Efremov MD, Gritsenko VA, Kochubei SA (1998) Raman study of silicon nanocrystals formed in SiN_x films by excimer laser or thermal annealing. Appl Phys Lett 73:1212

Wei SH (2004) Overcoming the doping bottleneck in semiconductors. Comput Mater Sci 30:337

Wei WS, Xu GY, Wang JL, Wang TM (2007) Raman spectra of intrinsic and doped hydrogenated nanocrystalline silicon films. Vacuum 81:656

Weil JA, Bolton JR (2007) Electron paramagnetic resonance: elementary theory and practical applications, 2nd edn. Wiley, Hoboken

Wilson EB, Cross PC, Decius JC (1980) The theory of infrared and Raman vibrational spectra. Dover, New York

Wolkin MV, Jorne J, Fauchet PM, Allan G, Delerue C (1999) Electronic states and luminescence in porous silicon quantum dots: the role of oxygen. Phys Rev Lett 82:197

Yang Y, Chen O, Angerhofer A, Cao YC (2008) On Doping CdS/ZnS Core/Shell Nanocrystals with Mn. J Am Chem Soc 130:15649

Yin MT, Cohen ML (1980) Microscopic theory of the phase transformation and lattice dynamics of Si. Phys Rev Lett 45:1004

Yoffe AD (2001) Semiconductor quantum dots and related systems: electronic, optical, luminescence and related properties of low dimensional systems. Adv Phys 50:1

Zhang SB, Northrup JE (1991) Chemical-potential dependence of defect formation energies in GaAs – application to Ga self-diffusion. Phys Rev Lett 67:2339

Zhang D, Kolbas RM, Milewski PD, Lichtenwalner DJ, Kingon AI, Zavada JM (1994) Light emission from thermally oxidized silicon nanoparticles. Appl Phys Lett 65:2684

Zhang Q, Bayliss SC, Hutt DA (1995) Blue photoluminescence and local structure of Si nanostructures embedded in SiO_2 matrices. Appl Phys Lett 66:1977

Zi J, Buscher H, Falter C, Ludwig W, Zhang KM, Xie XD (1996) Raman shifts in Si nanocrystals. Appl Phys Lett 69:200

Zi J, Zhang KM, Xie XD (1997) Comparison of models for Raman spectra of Si nanocrystals. Phys Rev B 55:9263

Zi J, Zhang KM, Xie XD (1998) Microscopic calculations of Raman scattering from acoustic phonons confined in Si nanocrystals. Phys Rev B 58:6712

Electronic Structure of Atomically Precise Graphene Nanoribbons

31

Leopold Talirz and Carlo A. Pignedoli

Contents

Abstract

Atomically precise graphene nanoribbons promise to combine the outstanding electronic properties of graphene with an electronic bandgap that is sufficiently large for digital-logic applications at room temperature. After a brief introduction

L. Talirz (✉)
Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, Sion, Switzerland
e-mail: leopold.talirz@epfl.ch

C. A. Pignedoli
Swiss Federal Laboratories for Materials Science and Technology, Empa, Dübendorf, Switzerland
e-mail: carlo.pignedoli@empa.ch

to nanoribbon nomenclature, experimental synthesis, and characterization, this chapter guides through the theoretical toolbox available for modeling their electronic structure, starting from simple empirical models and ending with state-of-the-art ab initio methods.

1   Introduction

1.1   Terminology

Graphene nanoribbons (GNRs) are strips of graphene with nanoscale widths, meaning widths below 100 nm. While the term is firmly established today, Fig. 1a illustrates that, historically, authors have been using alternative terms, such as *carbon nanoribbons*, *nanographite ribbons*, or *one-dimensional graphite*. Besides the width limit, the term GNR imposes no hard geometrical constraints in terms of length, aspect ratio, or edge shape and thus applies to a diverse set of graphene-derived nanostructures.
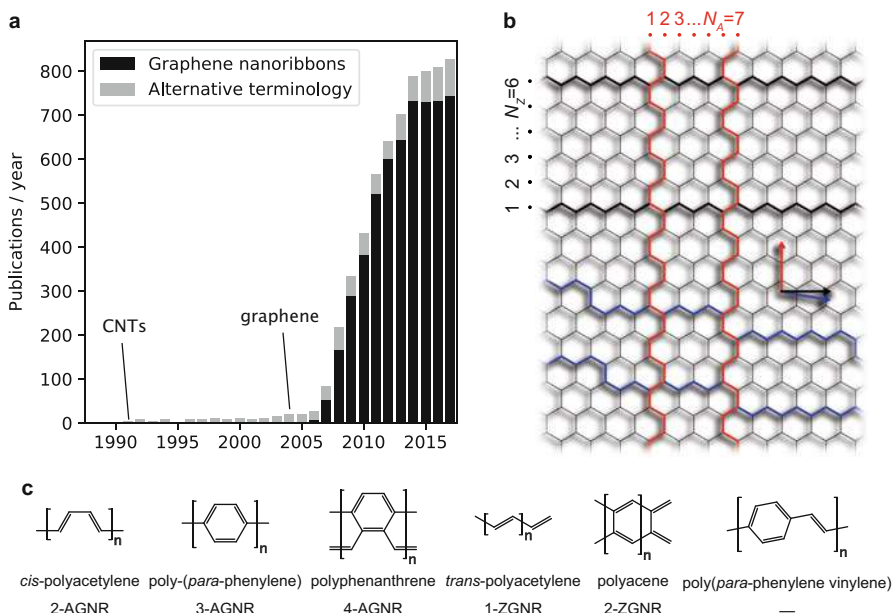
Fig. 1 Graphene nanoribbon (GNR) nomenclature. (**a**) Scientific articles related to GNRs from 1997 to 2017, indicating the discovery of carbon nanotubes (CNTs) and graphene. Counted are publications with *graphene nanoribbon(s)* in their title, abstract, or keywords (black). Alternative terms (gray) include, in descending order of frequency, *graphene ribbon(s)*, *carbon nanoribbon(s)*, *nanographite ribbon(s)*, *graphite ribbon(s)*, *graphitic ribbon(s)*, *one-dimensional graphite*, *graphene nanostrip(s)*, *nanographene ribbon(s)*, and *graphene nanowiggle(s)*. Source: Web of Science. (**b**) GNRs with armchair (red), zigzag (black), and chiral (blue) edges, shown on a graphene sheet. (**c**) GNR-type polymers with their chemical and GNR nomenclature

One defining characteristic of a GNR is the direction, along which a graphene sheet would need to be cut in order to produce it. Graphene has two high-symmetry directions, termed *armchair* and *zigzag* after the shape of the edges they produce. Fig. 1b shows examples of ribbons cut along the armchair direction (red) and the zigzag direction (black). Ribbons cut along any intermediate direction (blue) are termed *chiral*.

By convention, the width of armchair GNRs (AGNRs) is given as the number N_A of carbon dimer lines across the ribbon and the width of zigzag GNRs (ZGNRs) as the number N_Z of carbon zigzag lines. While dedicated notations exist for GNRs with specific edge shapes, there is no generally accepted notation covering GNRs with arbitrary edge shape.

The freedom in the edge shape also differentiates GNRs from the closely related single-walled carbon nanotubes (CNTs): only a subset of GNRs can be rolled up to form a defect-free CNT, and a CNT can be unzipped in multiple ways, producing different GNRs. As a consequence, the notation used to classify CNTs based on the chiral vector (n, m) cannot simply be transferred to GNRs.

The GNRs produced from the bottom-up through linking of molecular building blocks belong to the family of polymers, and some of the most narrow GNRs listed in Fig. 1c are well-known in polymer science. As polyacetylene still lacks benzenoid rings, it is debatable whether its *trans* and *cis* configurations should be considered the narrowest zigzag and armchair GNR, respectively. The wider poly-acene (2-ZGNR) and poly(*para*-phenylene) (3-AGNR), however, already satisfy this criterion.

Finally, GNR terminology often stops at defining the carbon backbone, although the decoration of the GNR edge by other chemical species can significantly impact the properties of a GNR. If no other information is provided, GNR edges are often implicitly assumed to be decorated by hydrogen such as to passivate any dangling bonds. Nevertheless, there are other relevant possibilities, and a complete specification of a GNR needs to include information on the edge passivation.

1.2   Why Graphene Nanoribbons?

Theoretical investigations of the electronic structure of GNR-type polymers can be traced back to the 1980s, when extending the carbon backbone of *trans*- and *cis*-polyacetylene was considered a promising avenue for improving the environmental stability of these π-conjugated conductive polymers (Yamabe et al. 1982; Tanaka and Yamabe 1985; Tanaka et al. 1987; Lowe et al. 1986).

In the 1990s, GNRs were reintroduced as a convenient theoretical model to study the effect of edge shape on the electronic and magnetic properties of nanoscale graphene, in view of their relevance for porous carbon and defective carbon nanotubes (Volpilhac and Hoarau 1978; Klein 1994; Fujita et al. 1996; Nakada et al. 1996).

In the 2000s, the seminal experiments on the electric field effect in graphene (Novoselov et al. 2004) renewed interest in GNRs – this time not as organic

conductors but as "semiconducting graphene." Graphene field-effect transistors have been shown to operate at frequencies beyond 400 GHz (Cheng et al. 2012) by leveraging the high mobility of charge carriers in graphene. The lack of a bandgap, however, limits the on/off ratio of graphene transistors to ≈10 at room temperature, which is far too low for many applications, both in high-frequency signal amplification and in digital electronics (Schwierz 2010, 2013). Cutting graphene into nanoribbons offers a natural way of opening a bandgap via quantum confinement.

The first key question is how narrow these ribbons need to be in order to exhibit significant electronic bandgaps. A rudimentary estimate of how quantum confinement affects the bandgap can be derived directly from the linear dispersion relation of graphene near the Fermi level:

$$E(q) \approx \pm \hbar v_F q \tag{1}$$

where the plus sign corresponds to the π^* conduction band, the minus sign corresponds to the π valence band, $v_F \approx 10^6 \frac{m}{s}$ is the Fermi velocity (Grüneis et al. 2008), and $\mathbf{q} = \mathbf{k} - \mathbf{K}$ the crystal wave vector measured relative to the Dirac point $\mathbf{K}$.

In a GNR of width w, the transverse component of $\mathbf{k} = (k^\perp, k^\parallel)$ is quantized to discrete values $k_n^\perp = n\frac{\pi}{w} = n\,\Delta k$, $n \in \mathbb{N}$. This limits how closely $\mathbf{k}$ can approach the Dirac point in reciprocal space. If the Dirac point falls midway between $k_n^\perp$ and $k_{n+1}^\perp$, the closest distance to the Dirac points is $\Delta k/2$, thus yielding a bandgap

$$\Delta = 2|E(\Delta k/2)| = \hbar v_F \Delta k \approx 2 \text{ eV nm}/w \tag{2}$$

Room-temperature digital-logic applications require on/off ratios in the $10^4 - 10^6$, which can be translated into a minimum bandgap of 0.4 eV (Kim et al. 2011). Equation (2) indicates that in order to open a bandgap of this magnitude in graphene, the graphene ribbons indeed cannot be more than a few nanometers wide.

While Eq. (2) is useful as a rule of thumb (Berger et al. 2006), it has numerous shortcomings, which are pointed out and addressed over the course of this chapter, working toward a computational model of the electronic structure of GNRs that is suitably accurate for comparison to experiments on atomically precise GNRs.

1.3   The Race for Atomic Precision

Opening a bandgap by quantum confinement offers the exciting prospect of tuning the value of the bandgap to just the right value for the intended application, simply by controlling the width of the GNRs. However, this path is not quite as straightforward as one might anticipate.

At a width of 5 nm, a GNR is just about 50 carbon atoms across. Patterning structures of such small dimensions is a major challenge for state-of-the-art

lithography (Hwang et al. 2015). On top of this, the electronic structure of these one-dimensional systems is also sensitive to the atomic structure of their edges, and even weakly irregular edges are predicted to give rise to scattering and localization of electrons (Gunlycke et al. 2007; Stampfer et al. 2009). If the high charge carrier mobilities of the parent material graphene are to be maintained, the GNR edges – meaning both the carbon skeleton and its passivation – should be made atomically precise.

Cutting graphene into atomically precise nanoribbons clearly is a daunting task, and so bottom-up approaches have been explored that aim at synthesizing a GNR from molecular precursors. Unfortunately, traditional polymerization chemistry faces a solubility problem with conjugated macromolecules containing some hundreds to thousands of atoms: as the width and length of the GNR increases, so does the strength of intermolecular π-π interactions, promoting aggregation (Yang et al. 2008; Narita et al. 2015).

One way to overcome this problem is to combine solution-based synthesis of specifically designed molecular precursors with on-surface synthesis of the GNRs, as sketched in Fig. 2a: the precursor molecules are evaporated onto a crystalline noble metal surface, which activates their docking sites. The activated molecules diffuse on the surface and undergo polymerization. At elevated temperature, additional carbon-carbon bonds form via cyclodehydrogenation, yielding the target GNR (Cai et al. 2010). By design, the precursor defines the GNR width, edge structure, and edge passivation down to the single atom, i.e., it defines its electronic structure. This ability to produce GNRs with uniform electronic properties is a crucial prerequisite for GNR-based technology (ITRS 2013).

The on-surface approach has not only been successful in providing a growing range of atomically precise GNRs (Fig. 2b) (Talirz et al. 2016a) but also has the advantage of producing the GNRs on a clean surface, where they can be studied by high-resolution surface-probe techniques, such as atomic force microscopy (AFM) and scanning tunneling microscopy (STM). As these techniques make it possible to determine both the atomic structure and the electronic structure of individual GNRs experimentally, they provide important benchmarks for the accuracy of computational modeling efforts.

1.4 Computational Challenges

Synthetic routes to a number of GNRs have been found, but their exploration is technically challenging, time-consuming, and expensive. This provides opportunities for computational modeling in the form of the following challenges:

1. Predicting which GNR targets are worth pursuing for a given application
2. Evaluating possible synthetic routes to a target GNR
3. Guiding the interpretation of complex measurements

In the context of atomically precise GNRs, the focus has only recently shifted from their synthesis to their incorporation into field-effect transistors (Bennett et al.
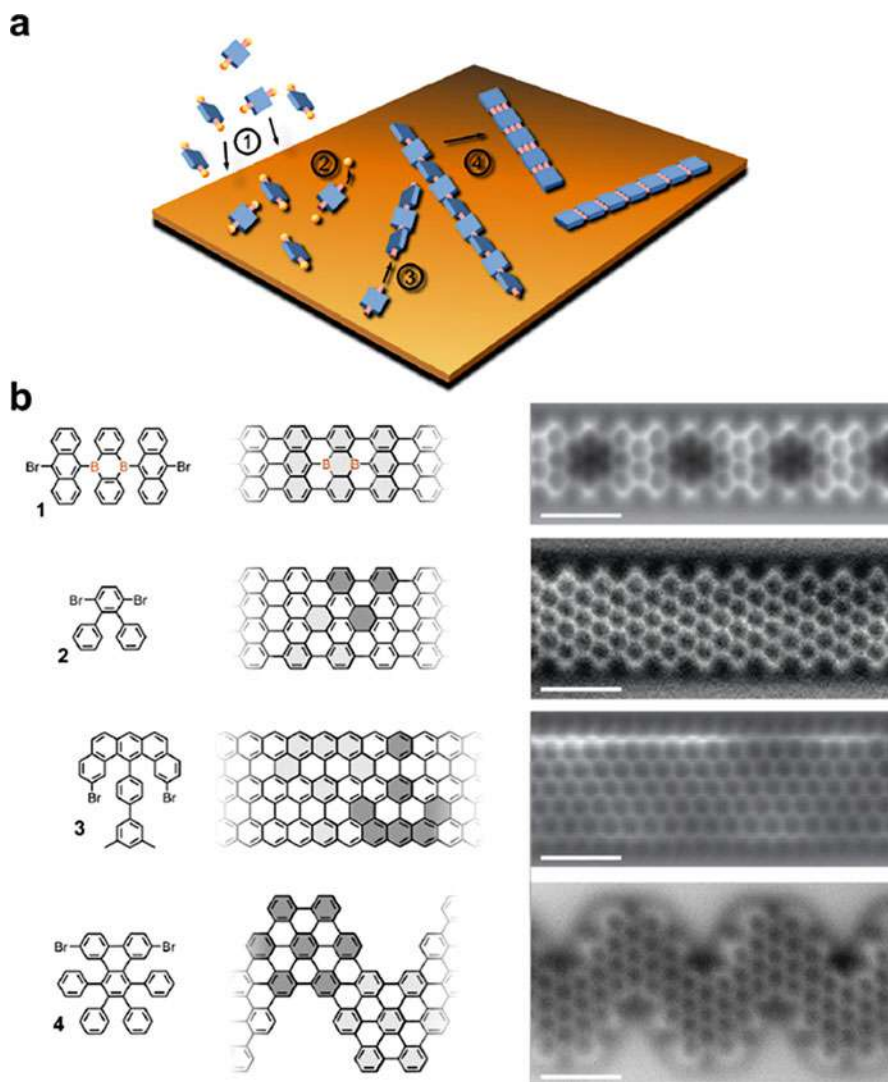
Fig. 2 On-surface synthesis of atomically precise GNRs. (**a**) Molecular Lego: deposition of the molecular building block (1) onto a crystalline noble metal surface at temperature $T = T_1$ activates their docking sites (2) by removing the halogen atoms. The activated radicals diffuse across the surface and polymerize (3). At elevated temperature $T_2 > T_1$, additional chemical bonds form (4), yielding the planar GNRs. (**b**) Selection of successful synthetic routes toward a B-doped 7-AGNR (Kawai et al. 2015), 9-AGNR (Talirz et al. 2017), 6-ZGNR (Ruffieux et al. 2016), and chevron-type GNR (Kawai 2013). Shown are sketches of precursor molecules and the GNR structure, together with atomically resolved images of the GNRs, obtained using non-contact atomic force microscopy (AFM) with CO tips. Scale bar 1 nm. AFM images are reproduced from corresponding references with permission

2013; Chen et al. 2016; Llinas et al. 2017). While electronic structure calculations do inform the decisions on which GNR targets to pursue, those calculations tend not to consider the GNRs' environment in a device, since the atomistic details of this environment in current prototype devices are still unknown. As the production of field-effect transistors using atomically precise GNRs matures, however, this first challenge will become increasingly important.

The second challenge has been recognized for some time, and efforts have been directed toward it (Talirz et al. 2016b). In practice, however, computational insight does not yet play a major role in the selection of synthetic strategies. Predicting the reaction path of a complex molecule on a noble metal surface still is a difficult problem, and it is outside the scope of this chapter.

It is the third challenge, where computational modeling has made the greatest contribution so far, in terms of rationalizing nontrivial observations in microscopy and spectroscopy. In view of this fact, it is instructive to briefly go through the experimental tools available for studying the electronic structure of atomically precise graphene nanoribbons and the observables they provide access to (Fig. 3).

On the one hand, there are "ionizing" spectroscopies, which excite the sample through the removal or the addition of electrons, thus leaving the sample in a charged excited state. This family includes scanning tunneling spectroscopy (STS, electron removal or addition), photoelectron spectroscopy (PES, electron removal), and inverse photoelectron spectroscopy (IPES, electron addition). The minimum energy required to remove an electron from the neutral GNR is termed *ionization energy I*, and the maximum energy that can be gained by adding an electron to the
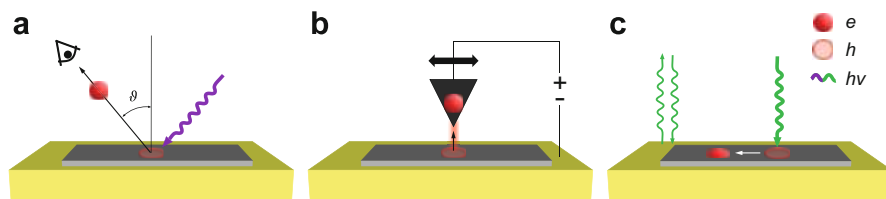
Fig. 3 Characterizing the electronic structure of GNRs from on-surface synthesis: (**a**) photoelectron spectroscopy (PES). Electrons are ejected from the sample using X-ray or ultraviolet lasers. The distribution of kinetic energies of ejected electrons provides access to the density of states of the sample. In angle-resolved PES (ARPES), detection is recorded as a function of the polar angle ϑ (and azimuthal angle φ), which allows to resolve the energy-momentum relation of the initial states. (**b**) Scanning tunneling spectroscopy (STS): electrons tunnel from the sample to a sharp tip or vice versa, depending on the applied bias voltage. By scanning the tip across the sample surface, the derivative of the tunneling current with respect to the bias voltage provides a map of the local density of states. (**c**) Reflectance difference spectroscopy (RDS): electron-hole pairs are created by excitations via a tunable laser, and reflectance is recorded as a function of photon energy. The optical absorption by the GNR is isolated by measuring the difference in reflectance between incoming light polarized along the GNRs (which must be uniformly aligned) and light polarized perpendicular to the GNRs

neutral GNR is the *electron affinity A*. Their difference defines the *electronic* (band) gap $\Delta_{el} = I - A$, also known as *transport* gap or *fundamental* gap.

On the other hand, there are "optical" spectroscopies, in which the sample is excited through absorption of a photon of energy $h\nu < I$ and thus remains charge-neutral. This family includes reflectance difference spectroscopy, a technique adapted particularly to the study of surface-adsorbed GNRs. After an optical excitation, the sample contains both an excited electron and the hole it left behind. It can therefore be viewed as a combination of two independent charged excitations: one that adds the hole and one that adds the electron. The simultaneous presence of electron and hole, however, leads to a Coulomb attraction between the two oppositely charged particles which reduces the energy required to create the charge-neutral excitation. If the electronic gap is a *direct* one (i.e. the corresponding optical transition is not forbidden by momentum-based selection rules), then the *optical* gap $\Delta_{opt} = \Delta_{el} - E_b$ is reduced by the *exciton binding energy* Eb.

The better a material can screen the Coulomb interaction between the electron and hole, the lower is the exciton binding energy. While exciton binding energies in bulk inorganic semiconductors are often negligibly small (<25 meV for Si, GaAs and GaN Dvorak et al. 2013), the weaker electronic screening in nanomaterials can give rise to substantial exciton binding energies in excess of 1 eV.

This has two important consequences for the quantitative modeling of GNRs: First, there is no *single* band structure that describes all electronic excitations in a GNR, and any viable theoretical approach needs to consider the type of excitation (in the following, we use Δ for the gap of non-interacting theories and add the appropriate subscripts only where the method takes the type of excitation into account). And second, in any real-world scenario, the material *surrounding* the GNR (which every atom of the GNR is in direct contact with) plays a crucial role in determining how charges are screened *inside* the GNR. While the study of GNRs in isolation provides important insights into their intrinsic electronic structure, for challenges 1 and 3 it is indispensable to take the influence of their surrounding into account.

2   Empirical Models

2.1   Clar's Theory: π-Electrons with Pen and Paper

Clar's theory of the aromatic sextet provides a simple formalism to predict the distribution of π-electrons in polycyclic aromatic hydrocarbons (PAHs) (Clar 1972). It is based on the insight that π-electrons in PAHs tend to *delocalize*. The π-electron distribution in PAHs is therefore not properly represented by one particular Kekulé structure consisting of single and double bonds. A more appropriate representation is provided by the superposition or *resonance* of *all* possible Kekulé structures.

The Clar sextet shown in Fig. 4 provides an intuitive visual representation of the delocalized nature of the six π-electrons in an aromatic cycle – two possible configurations of single and double bonds are replaced by one ring. This notation

Fig. 4 The Clar sextet, representing the superposition of two complementary Kekulé structures
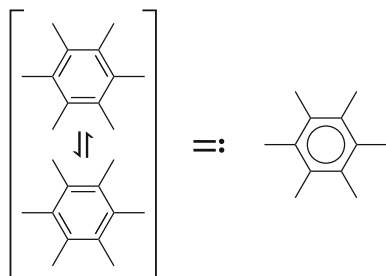
Fig. 5 Clar formulas for (**a**) triphenylene, (**b**) anthracene, and (**c**) graphene. The unit cell of graphene is shown in gray
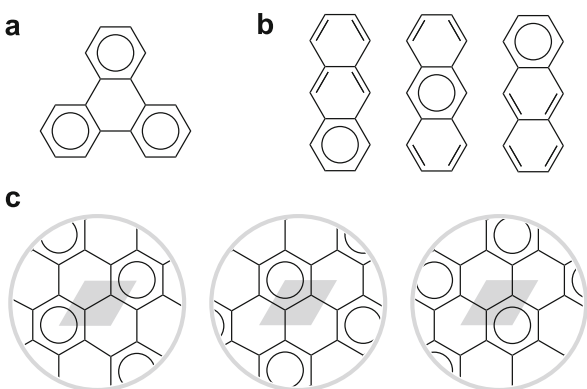
allows to cover the different possible Kekulé structures of a given molecule with significantly less effort, simply because a structure containing n Clar sextets represents 2^n Kekulé structures. It also introduces a natural hierarchy: the more Clar sextets a structure contains, the more Kekulé structures it represents and thus the higher its weight in the superposition.

The Clar structure (or Clar *cover*) with the maximum number of sextets is termed the *Clar formula*. It is shown in Fig. 5 for three different cases. Triphenylene has a unique Clar formula containing three Clar sextets. Anthracene can host only one Clar sextet at a time but has three Clar formulas that again should be considered *resonating* between each other. Note that the superposition of Clar structures can lead to certain Kekulé structures being counted multiple times. In the case of anthracene, for example, the Kekulé structures represented by the central Clar formula in Fig. 5b are already contained in the two other Clar formulas on the left and right.

Since the bonds sticking out of a Clar sextet are single bonds, a Clar structure can never contain two Clar sextets directly adjacent to each other. The maximally dense network of Clar sextets is found in graphene, where it forms a $(\sqrt{3} \times \sqrt{3})$R30° superstructure with respect to the graphene unit cell. In infinite graphene, this network of Clar sextets can be shifted between three equivalent positions, as shown in Fig. 5c.

Up to this point, the introduction of Clar sextets has merely improved the *visualization* of the π-electron distribution. The understanding gained in this process, however, allows to make qualitative predictions about actual physical observables. By distinguishing between single bonds, aromatic bonds, and double bonds with corresponding bond orders of 1, 1.5, and 2, bond lengths can be predicted on a qualitative level (Pauling et al. 1935). The relative thermodynamic stability of different molecules can be judged by analyzing the degree of π-electron delocalization via counting of Clar sextets. And finally, the π-electron distribution *itself* is an observable that is accessible, for example, in scanning tunneling microscopy.

While Clar formulas are the Clar structures that, individually, represent the largest number of Kekulé structures, they do not necessarily cover *all* possible Kekulé structures. In general, a full analysis thus needs to go beyond the Clar formula and consider Clar structures with lower number of aromatic sextets. For finite benzenoid nanostructures, Clar structures can be computed straightforwardly by a computer program (Chou and Witek 2012, 2014).

Clar's theory has been applied extensively to nanographenes (Fujii and Enoki 2012). In the case of graphene nanoribbons, Clar's theory provides insight into how the boundary conditions imposed by different edge types impact delocalization. Armchair graphene nanoribbons with monohydrogenated edges are found to group into three families, depending on their width N_A (Wassmann et al. 2010):

- $N_A = 3p$: one unique Clar formula
- $N_A = 3p + 1$: two Clar formulas
- $N_A = 3p + 2$: $3p$ Clar formulas

with $p \in \mathbb{N}$. As illustrated by the examples shown in Fig. 6a–c, the Clar formulas for $N_A = 3p$ and $N_A = 3p + 1$ exclude certain hexagons from hosting sextets, giving rise to significant variations in bond order. The Clar formulas for $N_A = 3p + 2$ on the other hand allow the sextets to propagate throughout the GNR, just like in graphene, and the simple tight-binding model discussed in Sect. 2.2 predicts these GNRs to be gapless.

This classification based on the number of Clar formulas can be extended to edges with varying hydrogenation states, and it has been demonstrated that the corresponding predictions concerning bond lengths and π-electron distribution as a function of edge geometry and passivation are in agreement with density functional theory calculations (Wassmann et al. 2010).

Clar's theory furthermore provides a simple explanation for the localized states found at monohydrogenated graphene zigzag edges. As shown in Fig. 6d, based on the ordinary Kekulé structures, only a single Clar sextet can be placed directly at a monohydrogenated zigzag edge. In comparison to infinite graphene, the π-electrons thus have significantly less opportunity to delocalize, resulting in a low thermodynamic stability. If, however, a double bond is broken in order to introduce an unpaired electron at every third "zag," the dense network of Clar sextets of graphene can be recovered (see Fig. 6e).
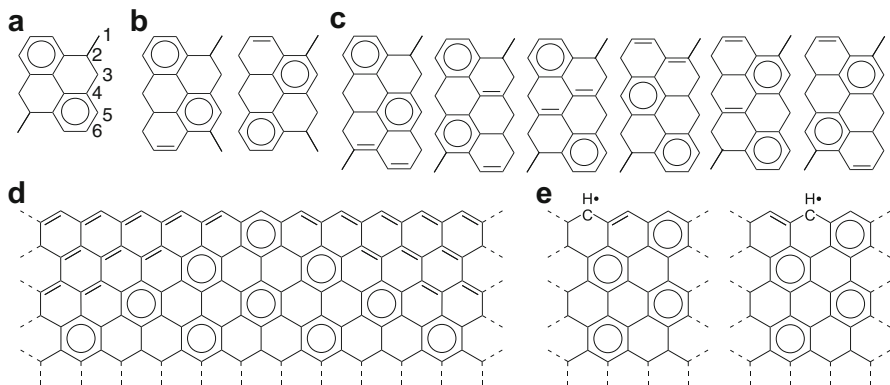
Fig. 6 Clar formulas of graphene nanoribbons. (**a–c**) Clar formulas for armchair graphene nanoribbons of the $N_A = 3p$ (**a**), $N_A = 3p + 1$ (**b**), and $N_A = 3p + 2$ (**c**) families. (**d**) Aperiodic Clar formula for infinite zigzag edge. (**e**) Periodic Clar formulas for infinite zigzag edge, enabled by introducing one unpaired electron every three "zags." (Figures (a–c) adapted from Corso et al. 2018)

The breaking of double bonds constitutes a relaxation of the usual requirement that Kekulé structures form the basis for determining the π-electron distribution. The stabilization gained by the delocalization can, however, outweigh the cost of the bond breaking. When this is the case, Clar's theory predicts zigzag edges to be characterized by highly reactive electronic states that are localized near the edge and host one electron every three "zags" – which is confirmed by higher-level theories such as tight binding (Fujita et al. 1996) or density functional theory (Son et al. 2006).

2.2   Tight Binding: Hopping on the Honeycomb Lattice

In the tight-binding approach, the problem of determining the electronic wave function in all of space is reduced to finding its optimal representation in a basis of pre-determined orbitals that are localized at atomic *sites*. For the case of graphene, the tight-binding representation can be made particularly simple, and its formulation predated the seminal experiments on graphene by many decades (Wallace 1947; Saito et al. 1998).

A general Hamiltonian matrix element between orbital m at site i and orbital n at site j is given by

$$H_{im,jn} = \langle \phi_{im} | \hat{\mathscr{H}} | \phi_{jn} \rangle = \langle \phi_{im} | - \hbar^2 \nabla^2 / 2m + V(\mathbf{r}) | \phi_{jn} \rangle \tag{3}$$

In graphene, all carbon atoms are electronically equivalent and have three nearest neighbors separated by $a \approx 0.142$ nm (Pauling et al. 1935), while the next-nearest neighbor distance is larger by a factor of $\sqrt{3} \approx 1.7$. As a first approximation, the

interaction between carbon sites can therefore be restricted to nearest neighbors. Since the main focus is on the delocalized π-electrons, which are less strongly bound than the electrons in the σ-bonds, one can further represent each carbon atom by a single $2p_z$ orbital, i.e., the indexes m, n can be dropped. Finally, the on-site term $H_{i,i}$ is identical for all sites and can be set to zero, since it results merely in a constant shift of the energy scale. The only non-zero matrix element remaining is the hopping integral (or *transfer* integral) $H_{i,j} =: -t$ between two neighboring sites i and j. In practice, t is usually treated as a parameter that is adjusted to fit the resulting electronic band structure against experiment (Bostwick et al. 2006) or ab initio calculations (Saito et al. 1998), yielding values in the range between 2.5 and 3 eV (Reich et al. 2002).

In second quantization, the corresponding tight-binding Hamiltonian reads

$$\hat{\mathscr{H}} = -t \sum_{\langle ij \rangle} \hat{c}_j^\dagger \hat{c}_i + \hat{c}_i^\dagger \hat{c}_j \tag{4}$$

where the sum extends over pairs of nearest neighbors and $\hat{c}_i^\dagger, \hat{c}_j$ are the electron creation and annihilation operators at site i and j.

Solving the tight-binding model involves solving the generalized eigenvalue problem

$$H\mathbf{v} = \varepsilon S \mathbf{v} \tag{5}$$

$$S_{ij} = \langle \phi_i | \phi_j \rangle \tag{6}$$

for the eigenenergies ε_α with corresponding eigenvectors $\psi_\alpha(\mathbf{r}) = \sum_j v_{\alpha j} \phi_j(\mathbf{r})$. In graphene, however, the overlap S_{ij} between $2p_z$ orbitals on neighboring sites is found to be < 0.1 (Reich et al. 2002) and can be neglected, leaving t as the sole parameter of the model. Note that this makes the model equivalent to the Hückel molecular orbital theory (Hückel 1933) with resonance integral $\beta = t$ and the Coulomb integral α set to zero.

For the two-dimensional hexagonal lattice of graphene, Hamiltonian (4) yields the dispersion

$$E(\mathbf{k}) = \pm t \sqrt{3 + 2\cos\left(\sqrt{3}k_x\right) + 4\cos\left(\sqrt{3}k_x/2\right)\cos\left(3k_y/2\right)} \tag{7}$$

where $\mathbf{k} = (k_x, k_y)$ is given in units of $1/a$. As shown in Fig. 7c, this formula already predicts the semimetallic nature of graphene, including the linear dispersion at the K and K' points. One simple way of fixing the parameter t is to match the experimental Fermi velocity $v_F \approx 10^6 \frac{m}{s}$ via the relation $\hbar v_F = 3ta/2$, yielding $t \approx 3.1$ eV.

Furthermore, this model allows to study the local electronic structure near graphene edges (Fujita et al. 1996; Nakada et al. 1996). When the edge carbon atoms are assumed to be passivated by one hydrogen atom each, they are sp^2-hybridized,
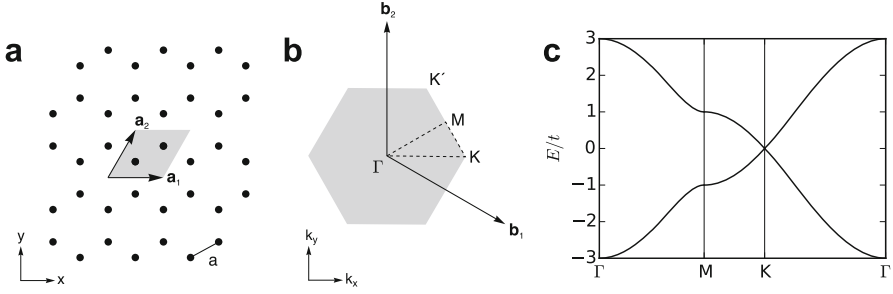
Fig. 7 (**a**) Real-space representation of graphene's primitive cell with lattice vectors $\mathbf{a}_1 = (\sqrt{3}, 0)$, $\mathbf{a}_2 = \left(\frac{\sqrt{3}}{2}, \frac{3}{2} \right)$ in units of the nearest-neighbor distance a. (**b**) First Brillouin zone of graphene with reciprocal lattice vectors $\mathbf{b}_1 = \left(\frac{2\pi}{\sqrt{3}}, -\frac{2\pi}{3} \right)$, $\mathbf{b}_2 = \left(0, \frac{4\pi}{3} \right)$ and high-symmetry points $\Gamma = (0, 0)$, $M = \left(\frac{\pi}{\sqrt{3}}, \frac{\pi}{3} \right)$, $K = \left(\frac{4}{3} \frac{\pi}{\sqrt{3}}, 0 \right)$. (**c**) Band structure corresponding to (7), following the path indicated in (**b**)
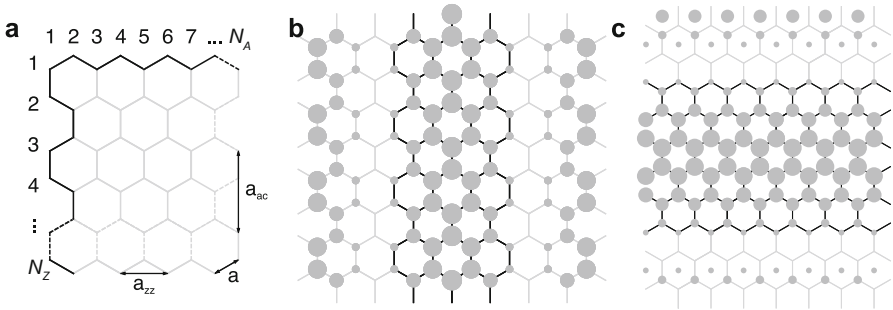
Fig. 8 (**a**) Numbering scheme for widths of ribbons with armchair and zigzag edges. The lattice constants are related by $a_{ac} = \sqrt{3} a_{zz} = 3a$. (**b**) Lowest π-electronic state of 7-AGNR, repeated periodically. (**c**) Lowest π-electronic state of 6-ZGNR, repeated periodically. The wave function at site j is represented by a circle with area $\propto |\psi(\mathbf{r}_j)|$

and thus each contributes one π-electron, just like the carbon sites in the interior. In nearest-neighbor tight binding, the edge boundary condition simply requires the wave function to vanish on the nearest-neighbor sites *outside* the edge.

Figure 8b shows the lowest π-electronic state of an AGNR. Considering the AGNR embedded into a virtual graphene lattice, one realizes that its eigenstates can be obtained by restricting an appropriate linear combination of two eigenstates of graphene to the lattice sites of the AGNR. The allowed crystal wave vectors have the form $\mathbf{k}^{\pm} = (\pm r \frac{2\pi}{2w}, k_y)$, where $r \in \{1, \dots, N_A\}$ and $w = \frac{\sqrt{3}}{2}(N_A + 1)$ is the width of the AGNR. The electronic band structure of AGNRs can thus be obtained simply by inserting $\mathbf{k}^{\pm}$ into (7), which corresponds to cutting the dispersion of graphene along N_A straight lines of constant k_x (see Fig. 9).
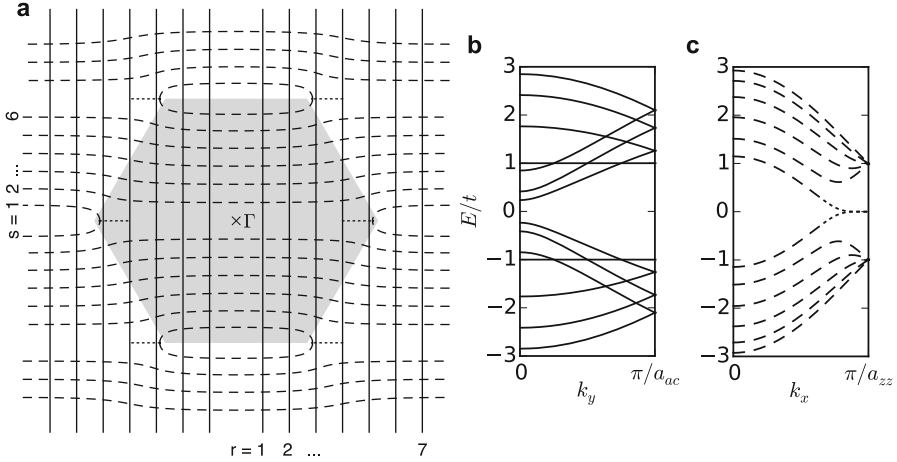
Fig. 9 (**a**) Brillouin zone of graphene with cutting lines for $N_A = 7$-AGNR (continuous) and 6-ZGNR (dashed). Band structures of 7-AGNR (**b**) and 6-ZGNR (**c**), plotted in their respective first Brillouin zones

For ZGNRs this is not the case, as illustrated by Fig. 8c: the periodic repetition of the eigenstates of the GNR would not even lie on the virtual graphene lattice. Nevertheless, they can still be obtained analytically (Wakabayashi et al. 2010), and their dispersion can still be represented as N_Z *curved* cuts of the graphene dispersion, except for the localized edge states that are indicated as dotted lines in Fig. 9a, c.

The formula for the bandgap of AGNRs differs between the three families identified previously via Clar's theory and yields $\Delta(3p+1) > \Delta(3p) > \Delta(3p+2)$:

$$\Delta(N_A) = \begin{cases} -2t\left[1 + 2\cos\left(\frac{r}{N_A+1}\pi\right)\right], & N_A = 3p, \quad r = 2p+1 \\ +2t\left[1 + 2\cos\left(\frac{r}{N_A+1}\pi\right)\right], & N_A = 3p+1, \ r = 2p+1 \\ 0, & N_A = 3p+2, \ r = 2p+2 \end{cases} \tag{8}$$

where $p \in \mathbb{N}$, and r denotes the index of the band closest to the Fermi level.

In the limit of large widths ($N_A \to \infty$), this yields a first refinement of the rudimentary formula (2):

$$\Delta(N_A) = \begin{cases} ta\frac{\pi}{w} + \mathcal{O}\left(\frac{1}{w^2}\right) \approx 1.4\,\text{eV nm}/w, & N_A \in \{3p, 3p+1\} \\ 0, & N_A = 3p+2 \end{cases} \tag{9}$$

By taking the boundary conditions at the armchair edges properly into account, this model predicts a smaller bandgap for the $N_A = 3p$ and $N_A = 3p+1$ families than formula (2), while for $N_A = 3p+2$, the bandgap vanishes entirely, since the cutting lines cross the K point.

In agreement with Clar's theory, the model predicts the existence of nonbonding electronic states that are localized near the zigzag edge and host one electron every three "zags." The corresponding energy band, shown in Fig. 9 for the 6-ZGNR, is metallic for ZGNRs of all width. For general GNR shapes, where no analytic solutions are available, there are open source tools that simplify the task of constructing and solving the tight-binding model for both finite and periodic systems (Coh and Vanderbilt 2017).

The formal equivalence with the Hückel molecular orbital theory allows to transfer the graph theoretical results obtained in this field to the tight-binding description of graphene nanostructures. A graphene nanostructure can be regarded as a hexagonal (or *benzenoid*) graph, the carbon atoms being its *vertexes* $\{v_i\}$ and the bonds between carbon atoms its *edges*. The nearest-neighbor tight-binding Hamiltonian (4) is then given simply by $(-t)$ times the *adjacency matrix* (Günthard and Primas 1956; Li et al. 2012)

$$A_{ij} = \begin{cases} 1 & v_i, v_j \text{ adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Since the trace of A vanishes, so does the sum of its eigenvalues, i.e., $\sum_i \varepsilon_i = 0$, where the sum goes over both occupied and empty states.

For hexagonal graphs, the number of negative and positive eigenvalues, corresponding to the number of bonding and antibonding orbitals, is equal (Fajtlowicz et al. 2005). Furthermore, their number can be found by counting the maximum number of pairwise nonadjacent edges β, i.e., the maximum number of double bonds in the corresponding Kekulé structure (Fajtlowicz et al. 2005). Since the total number of orbitals is given by the total number of vertexes N_v, the number of zero-energy *nonbonding* orbitals is obtained as $\eta = N_v - 2\beta$. These considerations can be used to classify the electronic structure of nanographenes based on topological arguments only (Wang et al. 2009). Note, however, that η is *not* equal to the number of edge-localized states, since their energies may differ from zero.

The nearest-neighbor single-orbital tight-binding model can be extended in numerous ways. Respecting the non-zero overlap integrals results in breaking particle-hole symmetry (Saito et al. 1998). Introducing site-dependent potentials in order to model symmetry breaking due to a supporting substrate or to capture the different chemical environment at the edge leaves the bandgaps unchanged to first order. Adapting hopping integrals to the different carbon-carbon bond lengths near the GNR edges, however, opens a bandgap for AGNRs of the $N_A = 3p + 2$ family (Son et al. 2006). The range of hopping may be extended. Allowing hopping between second-nearest neighbors affects the electronic structure near the Fermi level only by a rigid shift to first order in the hopping integrals (White et al. 2007). Hopping between third-nearest neighbors, on the other hand, plays a significant role (Gunlycke and White 2008). Increased flexibility may also be obtained by enlarging the basis set, e.g., to three orbitals per carbon site (Boykin et al. 2011). The additional complexity introduced by these extensions, however, comes at the cost of

additional parameters, which are typically fitted to band structures calculated by ab initio methods.

The extensions described above have in common that the electrons are still treated as noninteracting Fermions. In GNRs with widths in the low nm regime, dielectric screening is reduced significantly, giving rise to enhanced Coulomb interactions between quasiparticles. It is therefore unlikely that any of these models can be fitted to accurately describe the quasiparticle band structure of graphene nanoribbons over the whole range of widths from narrow GNRs up to graphene (Boykin et al. 2011).

2.3   The Hubbard Model: Many-Body Physics On-Site

One extension of the tight-binding model that *does* introduce explicit Coulomb interactions is the Hubbard model (Hubbard 1963):

$$\hat{\mathscr{H}} = -t \sum_{\langle ij \rangle, \sigma} \left(\hat{c}_{j\sigma}^{\dagger} \hat{c}_{i\sigma} + \hat{c}_{i\sigma}^{\dagger} \hat{c}_{j\sigma} \right) + U \sum_{i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \tag{10}$$

where the electron spin $\sigma \in \{\uparrow, \downarrow\}$ is now made explicit, $\hat{n}_{i\sigma} = \hat{c}_{i\sigma}^{\dagger} \hat{c}_{i\sigma}$ is the spin-resolved electron density at site i, and $U > 0$ describes the energy penalty associated with the on-site Coulomb repulsion between two electrons of different spins (in a single-orbital model, two electrons of the *same* spin can never occupy the same lattice site due to the Pauli exclusion principle).

The physics described by the Hubbard model is that of a competition between the kinetic energy of electrons, which is minimized by *delocalization*, and the Coulomb energy, which favors complete *localization* of each electron on one lattice site. The apparent simplicity of this lattice model for a short-range Coulomb interaction is deceptive. The Hubbard term is diagonal in real space but non-diagonal in momentum space, while exactly the opposite is true for the tight-binding term. Today, more than half a century after Hubbard's original description of the model, exact solutions for ground-state energy, excitation spectrum, etc. are available only for one-dimensional chains, where two electrons passing each other must actually hop directly "through" each other (Essler et al. 2005) (GNRs are *not* one-dimensional in this sense).

Some exact properties of the Hubbard model relevant to graphene nanostructures are known. Lieb proved for bipartite lattices at half-filling (and a repulsive value of U) that the ground state is nondegenerate and its total spin S is given by $S = \frac{1}{2}|N_A - N_B|$, where N_A and N_B are the number of atoms of the two sublattices (Lieb 1989). The theorem holds equally for periodic and finite structures (Lieb 1989). In particular, the total spin of nanographenes with equal numbers of atoms of both carbon sublattices, such as zigzag graphene nanoribbons, is therefore always equal to zero. Furthermore, Sorella and Tosatti performed quantum Monte Carlo calculations for the honeycomb lattice at zero temperature (Sorella and Tosatti

1992), showing that a transition from the semimetallic to an insulating state occurs, when the ratio U/t grows beyond a value of 4.5 ± 0.5.

For computing spectral properties and electron distributions of graphene nanoribbons, however, exact solutions of the Hubbard model are typically not feasible, and approximation is required. In this context it is helpful to know the relative magnitude of the two model parameters t and U. Like t, U is not a physical observable and is obtained by fitting models to experiment. Following (Yazyev 2010), the data concerning U seems to be rather scarce. From studying the localized spin distribution associated with solitons in thin films of polyacetylene (Ito et al. 1974), an effective $U \approx 3$ eV was extracted (Thomann et al. 1985). Assuming that the value of U in graphene is similar, one obtains $U/t \approx 1$. This places graphene neither in the extreme of $U \ll t$, where Gutzwiller's variational approach would yield a semimetal with a renormalized hopping integral (Gutzwiller 1965; Brinkman and Rice 1970), nor in the extreme $U \gg t$ of the Mott insulator, which can be described in terms of a Heisenberg model with antiferromagnetic coupling depending on t. For graphene, a different route for approximation is therefore required.

In the mean-field approximation, $\hat{n}_{i\sigma}$ is replaced by $\langle\hat{n}_{i\sigma}\rangle + \delta\hat{n}_{i\sigma}$, and terms of $\mathscr{O}(\delta\hat{n}_{i\sigma}^2)$ are neglected, resulting in

$$\hat{\mathscr{H}} = -t \sum_{\langle ij\rangle,\sigma} \left(\hat{c}_{j\sigma}^\dagger \hat{c}_{i\sigma} + \hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} \right) + U \sum_i \langle\hat{n}_{i\uparrow}\rangle\hat{n}_{i\downarrow} + \langle\hat{n}_{i\downarrow}\rangle\hat{n}_{i\uparrow} - \langle\hat{n}_{i\uparrow}\rangle\langle\hat{n}_{i\downarrow}\rangle \quad (11)$$

The Hubbard term now involves only one-particle operators, and the model (11) can be solved by determining the optimal distribution of spins in a self-consistent manner. Another way to arrive at the same approximation is to make the Ansatz of a single Slater determinant of spin orbitals. In this sense, the mean-field approximation of the Hubbard model is nothing but the unrestricted Hartree-Fock method on a lattice.

For armchair GNRs, the mean-field Hubbard model predicts no breaking of spin symmetry and thus produces the same band structure as the tight-binding model. For zigzag GNRs, however, magnetic moments are predicted to form at the zigzag edges (Fujita et al. 1996), which give rise to a staggered sublattice potential for the electrons of each spin channel and opens a bandgap. As illustrated in Fig. 10a, the spins align in parallel along the edge and antiparallel between opposite edges.

While it is common practice to draw spins pointing out of the GNR plane, it is important to keep in mind that there is no preferred absolute alignment of spins at room temperature due to the low spin-orbit splitting in freestanding graphene nanostructures. Even the *relative* alignment of spins must be limited in range (see Section "Density Functional Theory: Exchange and Correlation in Real Space"), since the Mermin-Wagner theorem rules out long-range magnetic order at finite temperature in one and two dimensions (Mermin and Wagner 1966).

Mean-field approximations replace a many-body problem by a one-body problem with an averaged *effective* interaction. In general, this neglect of correlation may be justified, if many interactions are involved in the average – the accuracy of the mean-
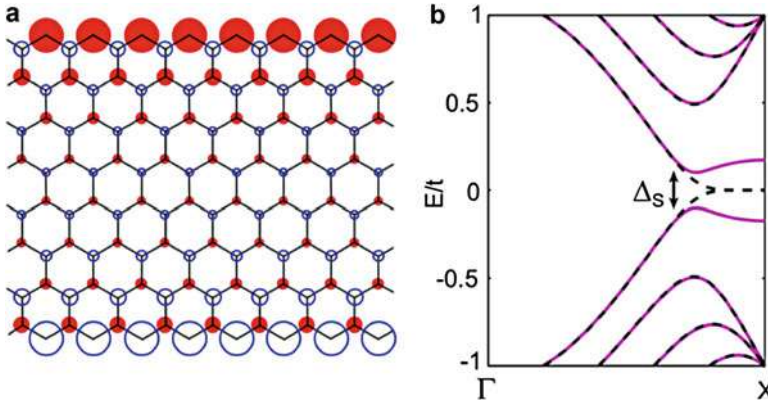
Fig. 10 ZGNRs in the Hubbard model. (**a**) Spin-up (filled circles) and spin-down (empty circles) densities of the 8-ZGNR for $U/t = 1.2$. Circle area is proportional to local magnetic moment. (**b**) Comparison between tight-binding bands (dashed lines) and Hubbard bands (solid lines) for 8-ZGNR. Band structures for spin-up and spin-down are identical. (Adapted with permission from Yazyev (2010), Copyright (2010) by IOP Publishing Ltd)

field approximation of the Hubbard model should increase with the coordination number Z of each lattice site. A priori, the two-dimensional honeycomb lattice with $Z = 3$ would not seem like a particularly suitable candidate, and, indeed, the neglect of quantum fluctuations leads to a substantial shift of the Mott-Hubbard transition from $U/t \approx 4.5$ to $U/t \approx 2.23$ (Sorella and Tosatti 1992). Still, the ratio of $U/t \approx 1$ in graphene puts it safely below the transition, and case studies for finite nanographenes show that the mean-field description agrees reasonably well with exact diagonalization and Quantum Monte Carlo simulations (Feldner et al. 2010). This makes the mean-field Hubbard model a computationally inexpensive alternative to ab initio methods when studying the spin distribution in nanographenes and its effect on the π-electronic structure (Yazyev 2010).

3   Ab Initio Methods

3.1   From First Principles

All ab initio electronic structure methods are ultimately rooted in the quantum mechanical Schrödinger equation:

$$\hat{\mathscr{H}} \, \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{R}_1, \ldots, \mathbf{R}_M, t) = i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{R}_1, \ldots, \mathbf{R}_M, t) \, .$$

(12)

The Schrödinger equation describes the time evolution of a system of N electrons and M nuclei in terms of the wave function Ψ, whose absolute square

$|\Psi(\{\mathbf{r}_j\}, \{\mathbf{R}_k\}, t)|^2$ equals the probability to find the system in the configuration specified by the electronic coordinates $\{\mathbf{r}_j\}$ and nuclear coordinates $\{\mathbf{R}_k\}$ at time t. Further coordinates, such as the spin of electrons or nuclei, are suppressed here for clarity and can be thought of as being included in $\mathbf{r}_j$ and $\mathbf{R}_k$.

The Hamiltonian operator $\hat{\mathscr{H}}$ contains the physical model appropriate for the particular system and the quantities of interest – here, graphene nanoribbons and their π-electronic structure. This information can be used to justify simplifications of the most general form of $\hat{\mathscr{H}}$, three of which are outlined in the following.

First, in the absence of time-dependent external potentials, $\hat{\mathscr{H}}$ does not explicitly depend on time. Without loss of generality, the search for solutions to Eq. (12) can thus be restricted to stationary states, whose trivial time-dependence $\Psi(t) = \exp(-iEt/\hbar)\Psi(0)$ can be factored out. This yields the time-independent Schrödinger equation

$$\hat{\mathscr{H}} \, \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{R}_1, \ldots, \mathbf{R}_M) = E \, \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{R}_1, \ldots, \mathbf{R}_M) \,. \qquad (13)$$

Second, graphene nanoribbons are made up of light elements, and the electrons in both π- and σ-bands move at velocities far below the speed of light. For a basic estimate, one connects the maximum binding energy $E_b \approx 20\,\mathrm{eV}$ of σ-electrons in graphene to the kinetic energy T via the virial theorem $T = \langle \hat{T} \rangle = -E = E_b$ (Engel and Dreizler 2011) and finds a Lorentz factor $\gamma = 1 + E_b/m_e c^2 \approx 1 + 4 \cdot 10^{-5}$, corresponding to a velocity $v = \sqrt{1 - \gamma^{-2}}c \approx 0.01c$.

The low atomic number of carbon also leads to weak spin-orbit coupling, which is estimated to be of the order of $1\,\mu\mathrm{eV}$ for flat graphene and $\approx 10\,\mu\mathrm{eV}$ in the presence of the slight corrugations that can occur in substrate-supported graphene (Huertas-Hernando et al. 2006). It is therefore safe to consider the nonrelativistic limit for the valence electronic structure of graphene nanoribbons.

And third, carbon nuclei are $M_n/m_e = 12u/m_e \approx 2.2 \cdot 10^4$ times heavier than electrons and thus generally move slower. Born and Oppenheimer showed that, when expanding the Hamiltonian $\hat{\mathscr{H}}$ up to fourth order in $\kappa = \sqrt[4]{m_e/M_n}$, the electrons follow the nuclear motion *adiabatically*, i.e., always remaining in their instantaneous ground state (Born and Oppenheimer 1927; Born and Huang 1954). In the adiabatic Born-Oppenheimer approximation, the wave function can be written as a product

$$\Psi(\{\mathbf{r}_j\}, \{\mathbf{R}_k\}) = \Psi_{\{\mathbf{R}_k\}}(\{\mathbf{r}_j\}) \, \chi(\{\mathbf{R}_k\})$$

of an electronic wave function $\Psi_{\{\mathbf{R}_k\}}$ and a nuclear wave function χ, where the equation for $\Psi_{\{\mathbf{R}_k\}}$ involves only the instantaneous *positions* of the nuclei, not their momenta. In the following, the subscript $\{\mathbf{R}_k\}$ is dropped for convenience.

The adiabatic approximation can break down, when electronic excitations can occur at the energy scales corresponding to nuclear vibrations. While the vanishing bandgap in graphene makes the adiabatic approximation problematic (Pisana et al. 2007), the narrow graphene nanoribbons discussed here have substantial electronic

gaps, and the adiabatic approximation should hold. Note, however, that this does not imply that nuclear vibrations can be disregarded altogether. When computing the electronic structure at finite temperatures, the effect of nuclear vibrations should be taken into account. Even at zero temperature, the quantum-mechanical zero-point motion of light nuclei can give rise to an appreciable renormalization of the electronic gap, such as $\approx 6\%$ for the optical bandgap of diamond (Cardona and Thewalt 2005). While these effects can be captured *within* the adiabatic approximation, they require the computation of the electron-phonon coupling matrix elements and have historically often been neglected. For recent ab initio work in this direction, see (Poncé et al. 2014; Cannuccia and Marini 2012).

Combining these three simplifications yields the time-independent Schödinger equation for nonrelativistic electrons in the Born-Oppenheimer approximation

$$\hat{\mathscr{H}}\,\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) = E\,\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) \tag{14}$$

where the Hamiltonian is now given explicitly by

$$\hat{\mathscr{H}} = -\sum_j \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_j} + \frac{1}{2}\sum_{j\neq k}\frac{1}{4\pi\varepsilon_0}\frac{e^2}{|\mathbf{r}_j - \mathbf{r}_k|^2} \tag{15}$$

$$-\sum_{jJ}\frac{1}{4\pi\varepsilon_0}\frac{Z_J e}{|\mathbf{r}_j - \mathbf{R}_J|} + \frac{1}{2}\sum_{J\neq K}\frac{1}{4\pi\varepsilon_0}\frac{(Z_J e)(Z_K e)}{|\mathbf{R}_J - \mathbf{R}_K|} \quad . \tag{16}$$

It contains the kinetic energy of electrons, the Coulomb repulsion between electrons, the Coulomb attraction between electrons and nuclei, as well as the Coulomb repulsion between nuclei. Since the motion of electrons and nuclei has been decoupled, the only term involving more than one spatial coordinate of the wave function is the Coulomb interaction between electrons. The Hamiltonian can therefore be written as

$$\hat{\mathscr{H}} := \sum_j \hat{h}(\mathbf{r}_j) + \frac{1}{2}\sum_{j\neq k}\frac{1}{4\pi\varepsilon_0}\frac{e^2}{|\mathbf{r}_j - \mathbf{r}_k|^2} \tag{17}$$

where $\hat{h}(\mathbf{r}_j)$ is a one-particle operator.

Finding accurate and efficient approximations to Eq. (14) is a formidable task, and generations of theoretical chemists and physicists have developed a whole arsenal of ab initio methods to tackle it. The following sections mention three methods that have proven useful in the study of atomically precise graphene nanoribbons.

While tight binding and the mean-field Hubbard model are usually introduced as empirical models with parameters to be fitted either against experiment or accurate calculations, it is worth noting that, these models (and their parameters) can be derived rigorously as approximations of ab initio methods such as density functional

theory (Foulkes and Haydock 1989; Foulkes 2016). Their extension provides a smooth transition – via semiempirical methods – toward ab initio methods free of empirical parameters.

3.2 Density Functional Theory: Exchange and Correlation in Real Space

Density functional theory (DFT) has evolved to become the most frequently used electronic structure method to study condensed matter systems. Since great introductory texts (Martin 2004) as well as reference works (Engel and Dreizler 2011) are readily available, this section assumes the reader is familiar with the basic theory and focuses on aspects relevant to the practical application of DFT to GNRs.

The electronic structure of GNRs predicted by Kohn-Sham density functional theory with standard (semi-)local exchange-correlation functionals is in qualitative agreement with the mean-field Hubbard model, when the various extensions discussed in Sect. 2.2 are taken into account. As shown in Fig. 11, AGNRs are predicted to be semiconducting with no spin polarization. Within each family, the bandgap is again approximately inversely proportional to the GNR width w and can be parametrized for $w \to \infty$ as (Yang et al. 2007):

$$\Delta(N_A) \approx \begin{cases} 0.8\,\text{eV}\,\text{nm}/w, \ N_A = 3p \\ 1.6\,\text{eV}\,\text{nm}/w, \ N_A = 3p + 1 \\ 0.3\,\text{eV}\,\text{nm}/w, \ N_A = 3p + 2 \end{cases} \tag{18}$$

thus providing a further refinement of Eq. (2).

In ZGNRs, the local (spin-)density approximation (LDA) confirms the familiar breaking of spin and sublattice symmetry, which opens a bandgap that is roughly inversely proportional to the GNR width w (Son et al. 2006). DFT has also been used to map out the dispersion of rotational spin-wave excitations along the zigzag edge, yielding a spin-spin correlation $\langle \hat{\mathbf{s}}_i \, \hat{\mathbf{s}}_{i+l} \rangle \propto \exp(-l/\xi(T))$ with correlation length $\xi(T) \approx 300\,\text{K}\,\text{nm}/T$ for $T > 10\,\text{K}$ (Yazyev and Katsnelson 2008). At room temperature, spins along the zigzag edge are thus expected to align only over length scales of $\approx 1\,\text{nm}$. Possible avenues for increasing ξ to values suitable for room-temperature spintronics applications include enhancing the magnetic anisotropy through substrate-induced spin-orbit coupling. For example, the spin-orbit splitting in graphene can be increased more than hundredfold to $\approx 10\,\text{meV}$ via adsorption on transition metal dichalcogenides (Wang et al. 2016b).

While the qualitative agreement with the (extended) mean-field Hubbard model also holds for the shape of the single-particle electronic states on the carbon lattice, DFT seamlessly incorporates changes in atomic geometry due to the presence of edges and their chemical passivation. This makes DFT well-suited for comparison against experimental methods that probe and resolve the local density of states, such as scanning tunneling microscopy/spectroscopy (Talirz et al. 2013; Kimouche et al.
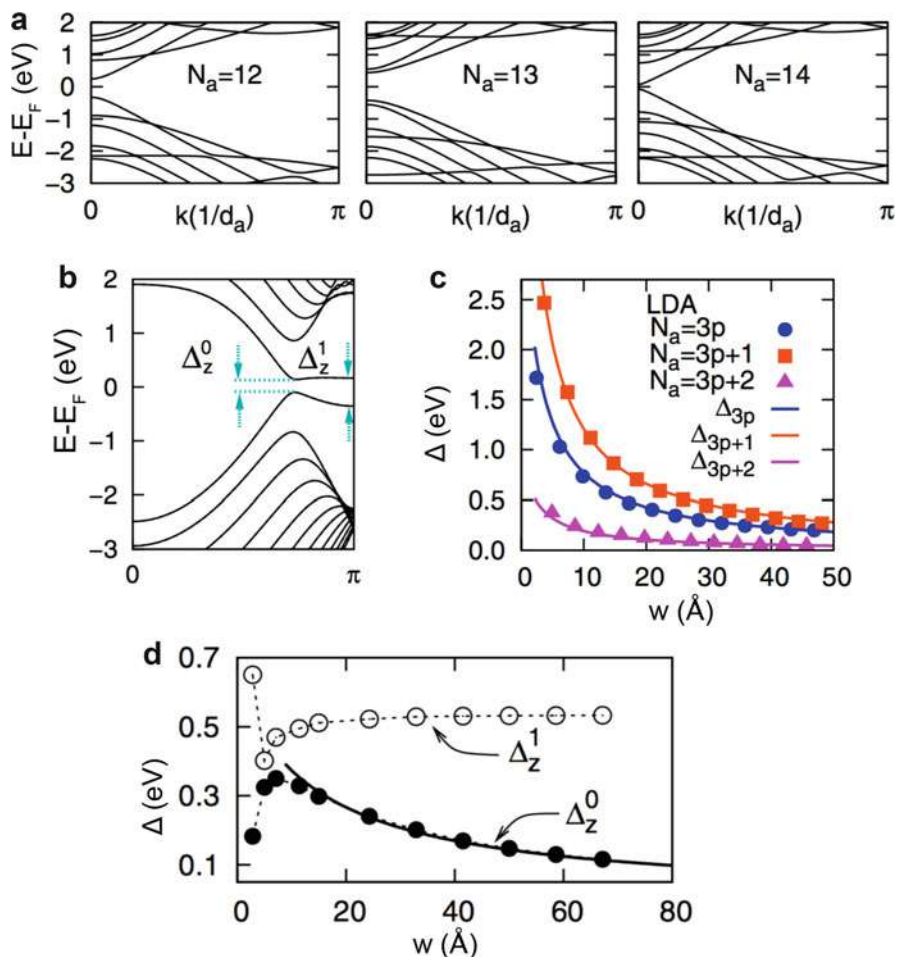
Fig. 11 Electronic structure of GNRs in the local density approximation (LDA). (**a**) Band structure of armchair GNRs from groups $N_A = 3p, 3p + 1, 3p + 2$. (**b**) Band structure of 12-ZGNR. (**c**) Kohn-Sham gap Δ of AGNRs as a function of width w. (**d**) Kohn-Sham gap Δ_Z^0 of ZGNRs as a function of width w. (Adapted with permission from Son et al. (2006), copyright (2006) American Physical Society)

2015). One peculiarity of STM is that the tip scans over the atoms at a distance of several Å, i.e., it proves the exponential tail of the electronic states, and care needs to be taken in order to ensure an accurate representation of these tails in STM simulations (Tersoff 1989).

Gas-phase DFT calculations of GNRs involving hundreds of atoms can be routinely performed with a wide range of well-tested codes (Lejaeghere et al. 2016). Conformational changes due to adsorption on a substrate can be included by considering a combination of quantum mechanics and molecular mechanics (QM/MM),
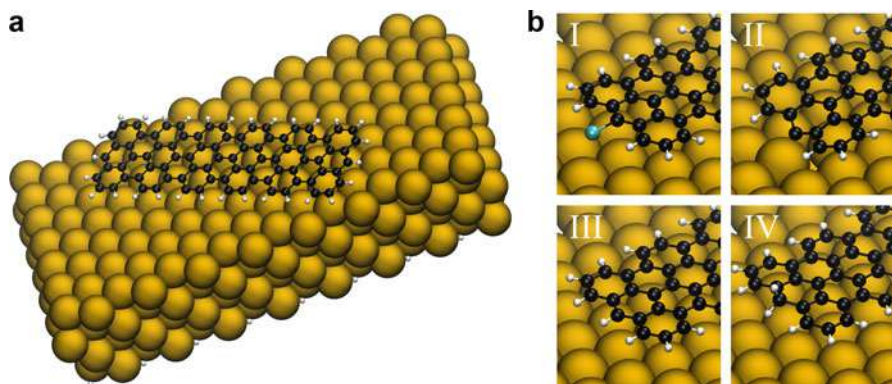
Fig. 12 Example of slab geometry: (**a**) short 7-AGNR adsorbed on four-layer Au(111) slab, terminated by H to remove the lower surface state. 2×4 nm unit cell containing 674 atoms. The unit cell is repeated periodically in three dimensions and relaxed with PBE, plus DFT-D3 dispersion correction. (**b**) Close-up of relaxed geometry for terminations with Br (I), radical (II), H (III), and 2H (IV) at the central carbon atom of the terminus. Termination (II) interacts strongly with the Au(111) surface, causing the carbon atom to bend down toward the surface by $\approx$1Å. (Adapted with permission from Talirz et al. 2013, copyright (2013) American Chemical Society)

where the substrate is treated using cheap empirical force fields. If the GNR is expected to form chemical bonds with the substrate, as shown in panel II of Fig. 12b, the substrate needs to be treated at the level of DFT. In this case, the number of atoms can easily grow beyond a thousand, as both the number of layers of the substrate slab and the distance between surface-adsorbed GNRs need to be converged.

Another challenge associated with modeling interactions with a substrate is dispersion forces: by construction, the usual (semi-)local approximations to correlation lead to an exponential decay of the interaction between atomic fragments for large separations R (Rydberg et al. 2003) and are thus unable to accurately capture dispersion forces with their asymptotic form $\propto 1/R^6$. Yet, in the absence of chemical bonds, dispersion forces dominate the interaction of GNRs with a substrate and play a key role in determining the adsorption height. This deficiency can be corrected by adding interatomic potentials of appropriate asymptotic form, termed "DFT-D," or double integrals over the electron density and gradient, termed "vdW-DF," to the density functional (Grimme 2011; Berland et al. 2015). While these corrections improve the description of dispersion forces substantially and at negligible computational overhead, both face the problem of how to avoid double counting of correlation. In DFT-D, this is handled by damping the correction at short interatomic distances, and it has been pointed out that molecular binding energies in some cases are more sensitive to the empirical parametrization of this damping function than to the physically meaningful dispersion coefficients (Hanke 2011). In vdW-DF, the correction is undamped, thus handing the task back to the density functional. These issues become particularly relevant for systems that fall in between the extremes of dispersion-dominated physisorption and covalent bonding,

where the DF and the correction need to work together. One example from this chemisorption regime is ZGNRs adsorbed on Au(111), where DFT-D-corrected calculations predict the preservation of magnetic edge states (Li et al. 2013), but low-temperature STS experiments seem to suggest otherwise (Ruffieux et al. 2016; Wang et al. 2016a).

Finally, Kohn-Sham DFT is a theory of the electronic ground state, and while knowledge of the exact density functional would enable finding the exact electronic ground-state density, the energy spectrum of the fictitious noninteracting Kohn-Sham system does not correspond to excitation energies of the many-body system (except for the energy of the highest occupied level, which corresponds to the ionization energy) (Perdew and Levy 1983). In particular, the Kohn-Sham gap differs from both the electronic gap Δ_{el} and the optical gap Δ_{opt}, even in exact DFT.

3.3 The *GW* Approximation: Screening of Charged Excitations

The GW approximation of many-body perturbation theory replaces the electronic self-energy Σ by $i\hbar GW$ (the first-order expansion in W), where G denotes the single-particle Green function and W the screened Coulomb interaction (Hedin 1965). The GW approximation has a long history of successful application in predicting electronic bandgaps in solids (Hybertsen and Louie 1985; Godby et al. 1986; van Schilfgaarde et al. 2006) and is described in detail in the literature (Hedin and Lundqvist 1969; Martin et al. 2016) and in the Chap. 15, "Modeling Excited States of Confined Systems" by Hung and Öğüt in the MTM set of this handbook. More recently, it has been shown to provide reasonably accurate ionization potentials and electron affinities for small molecules (Knight et al. 2016). In GNRs, the accuracy of the GW approximation can be expected to improve due to stronger screening of the interaction W and, importantly, provides a framework for computing electron addition and removal energies that captures the nonlocal screening from the GNRs' environment out of the box.

Early GW calculations of GNRs in isolation were performed starting from the orbitals and Kohn-Sham energies in the LDA (G_0W_0@LDA) (Yang et al. 2007). While the GW quasiparticle band structure remains qualitatively similar to the DFT band structure, the electronic gap of AGNRs opens substantially to

$$\Delta_{el}(N_A) \approx \begin{cases} 3.1\,\mathrm{eV\,nm}/w, & N_A = 3p \\ 4.4\,\mathrm{eV\,nm}/w, & N_A = 3p + 1 \\ 1.5\,\mathrm{eV\,nm}/w, & N_A = 3p + 2 \end{cases} \qquad (19)$$

as shown in Fig. 13a. Importantly, the corrections depend on the character of the electronic states (σ-states are more strongly corrected than π-states at the same energy) and result in an increase of dispersion near the Fermi level by 10–20%. This increase in bandwidth is consistent with GW calculations for graphene, where

Fig. 13 Electronic gap Δ_{el}
of armchair GNRs in GW.
(**a**) LDA gap and
quasiparticle corrections at
the G_0W_0@LDA level.
(Reproduced with permison
from Yang et al. (2007),
copyright (2007) American
Physical Society). (**b**)
Reduction of G_0W_0@LDA
gaps due to screening by
metallic substrate, computed
via image charge model
(image plane is located 1.7 Å
below the GNR). (Adapted
with permission from
Kharche and Meunier (2016),
copyright (2016) American
Chemical Society)
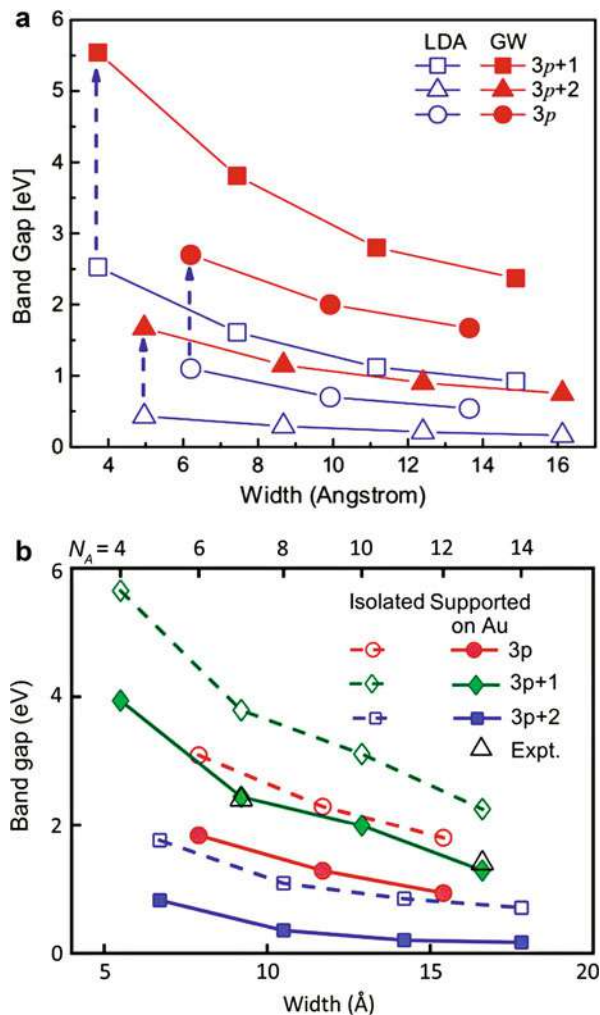
the resulting increase of the Fermi velocity improves agreement with experiment
compared to the LDA (Grüneis et al. 2008; Trevisanutto et al. 2008).

Unfortunately, there is currently no meaningful reference to validate the accu-
racy of GW calculations for GNRs in the gas phase: the large size of GNRs
prevents experimental studies of the gas phase and makes it challenging to
obtain complementary theoretical values from accurate but unfavorably scaling
quantum chemistry methods. For a meaningful comparison to experiments, be it
photoelectron spectroscopy or scanning tunneling spectroscopy, GW calculations
of GNRs therefore need to consider the substrate, on which the GNRs are adsorbed.
For "simple" substrates like NaCl or graphene, this is feasible with significant
computational effort (Jiang et al. 2013). In the on-surface synthesis approach,

however, GNRs are usually synthesized on the noble metal surfaces Au(111) and Ag(111), whose many electrons pose a major challenge for the convergence of GW calculations (Tamblyn et al. 2011). For AGNRs on metallic surfaces, a simple remedy is to introduce the screening by the substrate into the gas-phase GW calculation via the method of image charges from classical electrostatics: a charge q in the GNR is attracted toward the metal surface as if interacting with an image charge $-q$, mirrored across the image plane. The position of the image plane (Lam and Needs 1993) as well as the adsorption height of the GNR can be computed at the level of DFT, thus making the model free of adjustable parameters (Kharche and Meunier 2016). Given the approximations involved, the excellent agreement with experiment shown in Fig. 13b may however be fortuitous.

Figure 13b illustrates that screening by a metallic substrate reduces the large gas-phase GW bandgaps almost back to the original DFT values, which raises the question whether the same cannot be achieved with less effort by using adequate DFT exchange-correlation functionals. While generalized Kohn-Sham schemes hold some promise in this respect (Stein et al. 2010), the answer is likely negative, since the nonlocal screening depends sensitively on the details of the GNRs' environment (Biller et al. 2011). For example, moving a 7-AGNR from the Au(111) surface onto a NaCl monolayer island already increases Δ_{el} by 0.5 eV (Wang et al. 2016a).

Turning to ZGNRs, GW predicts a substantial gap of the order of $\Delta_{el} \approx 1$eV for widths in the range of $1-2$ nm (Yang et al. 2007) (not shown). Recent measurements of Δ_{el} for zigzag edges on a monolayer of NaCl on Au(111) are compatible with these predictions (Ruffieux et al. 2016; Wang et al. 2016a). It is important, however, to keep in mind that these experiments were carried out at 4 K, i.e., significantly below the temperature at which spin-wave excitations are predicted to start playing a dominant role (Yazyev and Katsnelson 2008). The size of Δ_{el} at room temperature remains to be determined and may be significantly smaller or vanish entirely.

Finally, it is worth noting that most of the GW calculations for GNRs to date involve additional approximations, such as plasmon-pole approximations for the energy dependence of the self-energy (Hybertsen and Louie 1986), a lack of self-consistency (which introduces a dependence on the starting point), and potentially unconverged basis sets. Recent validation efforts rightfully point out the necessity of eliminating these technical approximations (van Setten et al. 2015; Maggio and Kresse 2017), and future calculations will be able to take advantage of the significant progress made in modern GW implementations, which reduce both the computational complexity and the prefactor (Govoni and Galli 2015; Liu et al. 2016; Wilhelm et al. 2018).

3.4 The Bethe-Salpeter Equation: Electron-Hole Interaction

The two-particle Bethe-Salpeter equation (BSE) provides a framework for explicitly describing the interaction between a photoexcited electron and hole. Building on the GW quasiparticle spectrum as well as the screened Coulomb interaction W, the BSE
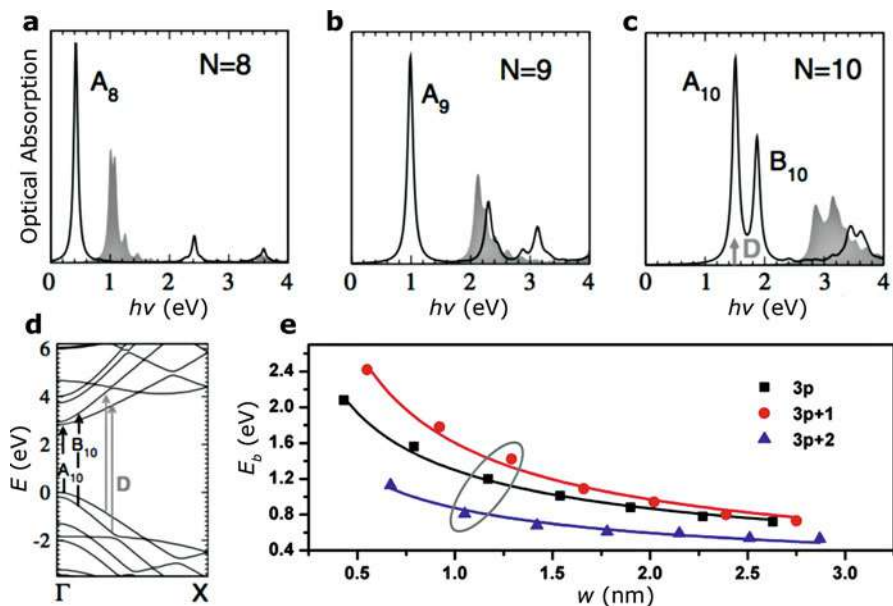
Fig. 14 Optical absorption spectra of armchair GNRs from GW+BSE. (**a–c**) Spectra for armchair GNRs of width $N_A = 8, 9, 10$ with (black lines) and without (gray) inclusion of electron-hole interaction, computed for linear polarization along the GNR axis. D indicates dark transitions with vanishing oscillator strength. (**d**) GW quasiparticle band structure for $N_A = 10$, indicating the character of optical transitions. (**e**) Exciton binding energies E_b for lowest-energy optical transition. Values for $N_A = 8, 9, 10$ are encircled in gray. ((**a–d**) adapted with permission from (Prezzi et al. 2008), copyright (2008) American Physical Society (**e**) adapted with permission from (Zhu and Su 2011), copyright (2011) American Chemical Society)

is one of the most accurate methods available for computing optical spectra and exciton binding energies in extended systems; see Onida et al. (2001) and Martin et al. (2016).

Figures 14a–c show the optical absorption spectra of selected AGNRs, as computed by the GW+BSE approach. AGNRs have *direct* bandgaps, and the lowest-energy optical transition involves the valence and conduction band. For $N_A = 8, 9, 10$, substantial exciton binding energies E_b of the order of 1 eV lead to a corresponding reduction of the optical Δ_{opt} with respect to the electronic gap Δ_{el}, thus highlighting the importance of differentiating between ionizing and optical excitations in AGNRs.

Similar to the case of carbon nanotubes (Marinopoulos et al. 2003), optical absorption in GNRs is strongly quenched for light polarized orthogonal to the GNR axis. This is leveraged in reflectance difference spectroscopy (RDS), which measures the difference in complex reflectance for incident light polarized along and across the axis of GNRs aligned on an optically isotropic substrate (see Fig. 3c). For the 7-AGNR, the theoretical value of $\Delta_{opt} \approx 1.9$ eV has been found to agree

reasonably well with the experimental value of $\Delta_{opt} \approx 2.1\,\text{eV}$ from RDS (Denk et al. 2014).

Compared to charged excitations, the effect of substrate screening on optical excitations in AGNRs is significantly reduced. This is a direct consequence of the fact that the corresponding excitons are charge-neutral and do not display a static dipole moment (Garcia-Lastra and Thygesen 2011). Minor renormalization is still expected due to inhomogeneity in the charge distribution, as demonstrated for the case of CNTs adsorbed on metals (Spataru 2013). The increased robustness of the optical transition energies against screening from the environment, however, allows to make contact between the characterization of atomically precise GNRs synthesized on metal surfaces and those synthesized in solution (Vo et al. 2014; Narita et al. 2014).

The $GW + \text{BSE}$ approach is computationally challenging – not only does it require the quasiparticle spectrum as input, an adequate description of the excitonic states requires very dense sampling of reciprocal space, and the computational complexity of solving the BSE is $\mathcal{O}(N^4)$ in the number of atoms N (Martin et al. 2016). In practice, additional approximations are often employed, such as using a static interaction W or neglecting nonresonant coupling (Tamm-Dancoff approximation). In this context, the validation of existing BSE implementations, in analogy to the recent efforts for GW approximation (van Setten et al. 2015), is highly anticipated.

4   Conclusions

Theoretical descriptions of the electronic structure of GNRs have come a long way since the empirical models employed in the twentieth century. Ab initio methods, such as density functional theory and many-body perturbation theory, have become valuable tools that assist in the structural and electronic characterization of atomically precise GNRs in highly controlled experimental settings.

As these GNRs start being incorporated into prototype devices, however, efforts both from the experimental and theory side are needed in order to arrive at a predictive accuracy that allows to point out the best GNR for a given application. While experimentalists will need to find ways of increasing control over the atomic-level environment of the GNR in a device, there are a number of challenges on the theory side waiting to be addressed. This chapter concludes with a "wish list" of these outstanding challenges, the solution of each of which would constitute a significant step forward:

- Vibrational effects on the bandgap of AGNRs at the GW level, including zero-point motion renormalization
- Dispersion-aware DFT that is accurate across interaction regimes, from physisorption over chemisorption to covalent bonds

- Refinement of embedding procedures, for capturing substrate effects beyond classical image charge models and modeling the diverse dielectric environment in devices
- Benchmarking GW calculations of AGNRs in the gas phase against accurate quantum chemistry references
- Elimination of starting-point dependence in GW calculations of GNRs, perhaps via quasiparticle self-consistent GW (QSGW)
- Quantifying the effects of spin-wave excitations on the electronic structure of ZGNRs at room temperature, potentially including substrate-induced enhancement of spin-orbit coupling

References

Bennett PB, Pedramrazi Z, Madani A, Chen YC, de Oteyza DG, Chen C, Fischer FR, Crommie MF, Bokor J (2013) Bottom-up graphene nanoribbon field-effect transistors. Appl Phys Lett 103(25):253114. https://doi.org/10.1063/1.4855116

Berger C, Song Z, Li X, Wu X, Brown N, Naud C, Mayou D, Li T, Hass J, Marchenkov AN, Conrad EH, First PN, de Heer WA (2006) Electronic confinement and coherence in patterned epitaxial graphene. Science 312(5777):1191–1196. https://doi.org/10.1126/science.1125925

Berland K, Cooper VR, Lee K, Schröder E, Thonhauser T, Hyldgaard P, Lundqvist BI (2015) Van der Waals forces in density functional theory: a review of the vdW-DF method. Rep Prog Phys 78(6):066501. https://doi.org/10.1088/0034-4885/78/6/066501

Biller A, Tamblyn I, Neaton JB, Kronik L (2011) Electronic level alignment at a metal-molecule interface from a short-range hybrid functional. J Chem Phys 135(16):164706. https://doi.org/10.1063/1.3655357

Born M, Huang K (1954) Dynamical theory of crystal lattices. Clarendon press, Oxford

Born M, Oppenheimer JR (1927) Zur Quantentheorie der Molekeln. Ann Phys 389(20):457–84. https://doi.org/10.1002/andp.19273892002

Bostwick A, Ohta T, Seyller T, Horn K, Rotenberg E (2006) Experimental determination of the spectral function of graphene, arXiv:cond-mat/0609660. https://arxiv.org/abs/cond-mat/0609660

Boykin TB, Luisier M, Klimeck G, Jiang X, Kharche N, Zhou Y, Nayak SK (2011) Accurate six-band nearest-neighbor tight-binding model for the π-bands of bulk graphene and graphene nanoribbons. J Appl Phys 109(10):104304. https://doi.org/10.1063/1.3582136

Brinkman WF, Rice TM (1970) Application of Gutzwiller's variational method to the metal-insulator transition. Phys Rev B 2(10):4302–4304. https://doi.org/10.1103/PhysRevB.2.4302

Cai J, Ruffieux P, Jaafar R, Bieri M, Braun T, Blankenburg S, Muoth M, Seitsonen AP, Saleh M, Feng X, Müllen K, Fasel R (2010) Atomically precise bottom-up fabrication of graphene nanoribbons. Nature 466(7305):470–3. https://doi.org/10.1038/nature09211

Cannuccia E, Marini A (2012) Zero point motion effect on the electronic properties of diamond, trans-polyacetylene and polyethylene. Eur Phys J B 85(9):320

Cardona M, Thewalt MLW (2005) Isotope effects on the optical spectra of semiconductors. Rev Mod Phys 77(4):1173–1224. https://doi.org/10.1103/RevModPhys.77.1173

Chen Z, Zhang W, Palma CA, Lodi Rizzini A, Liu B, Abbas A, Richter N, Martini L, Wang XY, Cavani N, Lu H, Mishra N, Coletti C, Berger R, Klappenberger F, Kläui M, Candini A, Affronte

M, Zhou C, De Renzi V, del Pennino U, Barth JV, Räder HJ, Narita A, Feng X, Müllen K (2016) Synthesis of graphene nanoribbons by ambient-pressure chemical vapor deposition and device integration. J Am Chem Soc 138(47):15488–15496. https://doi.org/10.1021/jacs.6b10374

Cheng R, Bai J, Liao L, Zhou H, Chen Y, Liu L, Lin YC, Jiang S, Huang Y, Duan X (2012) High-frequency self-aligned graphene transistors with transferred gate stacks. Proc Nat Acad Sci 109(29):11588–11592. https://doi.org/10.1073/pnas.1205696109

Chou CP, Witek HA (2012) An algorithm and FORTRAN program for automatic computation of the zhang – zhang polynomial of benzenoids. MATCH Commun Math Comput Chem 68:3–30

Chou CP, Witek HA (2014) ZZDecomposer: a graphical toolkit for analyzing the zhang-zhang polynomials of benzenoid structures. MATCH Commun Math Comput Chem 71:741–764

Clar E (1972) The aromatic sextet. Wiley, London

Coh S, Vanderbilt D (2017) Python Tight Binding (PythTB). http://www.physics.rutgers.edu/pythtb

Corso M, Carbonell-Sanromà E, de Oteyza DG (2018) Bottom-up fabrication of atomically precise graphene nanoribbons. In: de Oteyza D, Rogero C (eds) On-Surface Synthesis II. Advances in Atom and Single Molecule Machines, Springer, Cham

Denk R, Hohage M, Zeppenfeld P, Cai J, Pignedoli CA, Söde H, Fasel R, Feng X, Müllen K, Wang S, Prezzi D, Ferretti A, Ruini A, Molinari E, Ruffieux P (2014) Exciton-dominated optical response of ultra-narrow graphene nanoribbons. Nat Commun 5(May):4253. https://doi.org/10.1038/ncomms5253

Dvorak M, Wei SH, Wu Z (2013) Origin of the variation of exciton binding energy in semiconductors. Phys Rev Lett 110(1):016402. https://doi.org/10.1103/PhysRevLett.110.016402

Engel E, Dreizler RM (2011) Density functional theory. Springer, Berlin/Heidelberg

Essler FHL, Frahm H, Göhmann F, Klümper A, Korepin VE (2005) The one-dimensional Hubbard model. Cambridge University Press, Cambridge

Fajtlowicz S, John PE, Sachs H (2005) On maximum matchings and eigenvalues of benzenoid graphs. Croat Chem Acta 78(2):195–201

Feldner H, Meng ZY, Honecker A, Cabra D, Wessel S, Assaad FF (2010) Magnetism of finite graphene samples: mean-field theory compared with exact diagonalization and quantum Monte Carlo simulations. Phys Rev B 81(11):115416. https://doi.org/10.1103/PhysRevB.81.115416

Foulkes WMC (2016) Tight-binding models and coulomb interactions for s, p, and d electrons. In: Quantum materials: experiments and theory: lecture notes of the autumn school on correlated electrons, Forschungszentrum Jülich, oCLC: 974567924

Foulkes WMC, Haydock R (1989) Tight-binding models and density-functional theory. Phys Rev B 39(17):12520–12536. https://doi.org/10.1103/PhysRevB.39.12520

Fujii S, Enoki T (2012) Clar's aromatic sextet and π-electron distribution in nanographene. Angew Chem 124(29):7348–7353. https://doi.org/10.1002/ange.201202560

Fujita M, Wakabayashi K, Nakada K, Kusakabe K (1996) Peculiar localized state at zigzag graphite edge. J Phys Soc Japan 65(7):1920–1923. https://doi.org/10.1143/JPSJ.65.1920

Garcia-Lastra JM, Thygesen KS (2011) Renormalization of optical excitations in molecules near a metal surface. Phys Rev Lett 106(18):187402. https://doi.org/10.1103/PhysRevLett.106.187402

Godby RW, Schlüter M, Sham LJ (1986) Accurate exchange-correlation potential for silicon and its discontinuity on addition of an electron. Phys Rev Lett 56(22):2415–2418. https://doi.org/10.1103/PhysRevLett.56.2415

Govoni M, Galli G (2015) Large scale GW calculations. J Chem Theory Comput 11(6):2680–2696. https://doi.org/10.1021/ct500958p

Grimme S (2011) Density functional theory with London dispersion corrections. Wiley Interdiscip Rev Comput Mol Sci 1(2):211–228. https://doi.org/10.1002/wcms.30

Grüneis A, Attaccalite C, Pichler T, Zabolotnyy V, Shiozawa H, Molodtsov SL, Inosov D, Koitzsch A, Knupfer M, Schiessling J, Follath R, Weber R, Rudolf P, Wirtz L, Rubio A (2008) Electron-electron correlation in graphite: a combined angle-resolved photoemission and first-principles study. Phys Rev Lett 100(3):037601. https://doi.org/10.1103/PhysRevLett.100.037601

Gunlycke D, White CT (2008) Tight-binding energy dispersions of armchair-edge graphene nanostrips. Phys Rev B 77(11):115116. https://doi.org/10.1103/PhysRevB.77.115116

Gunlycke D, Areshkin DA, White CT (2007) Semiconducting graphene nanostrips with edge disorder. Appl Phys Lett 90(14):142104. https://doi.org/10.1063/1.2718515

Günthard HH, Primas H (1956) Zusammenhang von Graphentheorie und MO-Theorie von Molekeln mit Systemen konjugierter Bindungen. Helv Chim Acta 39(6):1645–1653. https://doi.org/10.1002/hlca.19560390623

Gutzwiller MC (1965) Correlation of electrons in a narrow s band. Phys Rev 137(6A):A1726–A1735. https://doi.org/10.1103/PhysRev.137.A1726

Hanke F (2011) Sensitivity analysis and uncertainty calculation for dispersion corrected density functional theory. J Comput Chem 32:1424–30. https://doi.org/10.1002/jcc.21724

Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139(3A):A796–A823. https://doi.org/10.1103/PhysRev.139.A796

Hedin L, Lundqvist S (1969) Effects of electron-electron and electron-phonon interactions on the one-electron states of solids. In: Frederick Seitz DT, Ehrenreich H (eds) Solid state physics, vol 23. Academic Press, New York and London, pp 1–181

Hung L, Öğüt S (2018) Modeling excited states of confined systems. In: Handbook of materials modeling. Springer, Cham.

Hubbard J (1963) Electron correlations in narrow energy bands. Proc R Soc London A Math Phys Eng Sci 276(1365):238–257. https://doi.org/10.1098/rspa.1963.0204

Hückel E (1933) Die freien Radikale der organischen Chemie. Z Phys 83(9–10):632–668. https://doi.org/10.1007/BF01330865

Huertas-Hernando D, Guinea F, Brataas A (2006) Spin-orbit coupling in curved graphene, fullerenes, nanotubes, and nanotube caps. Phys Rev B 74(15):155426. https://doi.org/10.1103/PhysRevB.74.155426

Hwang WS, Zhao P, Tahy K, Nyakiti LO, Wheeler VD, Myers-Ward RL, Jr CRE, Gaskill DK, Robinson JA, Haensch W, Xing HG, Seabaugh A, Jena D (2015) Graphene nanoribbon field-effect transistors on wafer-scale epitaxial graphene on SiC substrates a. APL Mater 3(1):011101. https://doi.org/10.1063/1.4905155

Hybertsen MS, Louie SG (1985) First-principle theory of quasiparticles: calculation of band gaps in semiconductors ans insulators. Phys Rev Lett 55(13):1418

Hybertsen MS, Louie SG (1986) Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. Phys Rev B 34(8):5390–5413. https://doi.org/10.1103/PhysRevB.34.5390

Ito T, Shirakawa H, Ikeda S (1974) Simultaneous polymerization and formation of polyacetylene film on the surface of concentrated soluble Ziegler-type catalyst solution. J Polym Sci Polym Chem Ed 12(1):11–20. https://doi.org/10.1002/pol.1974.170120102

ITRS (2013) Emergent research materials. Technical report ITRS

Jiang X, Kharche N, Kohl P, Boykin TB, Klimeck G, Luisier M, Ajayan PM, Nayak SK (2013) Giant quasiparticle bandgap modulation in graphene nanoribbons supported on weakly interacting surfaces. Appl Phys Lett 103(13):133107. https://doi.org/10.1063/1.4822427

Kawai S (2013) Non-contact AFM image of chevron-type GNR (unpublished)

Kawai S, Saito S, Osumi S, Yamaguchi S, Foster AS, Spijker P, Meyer E (2015) Atomically controlled substitutional boron-doping of graphene nanoribbons. Nat Commun 6:8098. https://doi.org/10.1038/ncomms9098

Kharche N, Meunier V (2016) Width and crystal orientation dependent band gap renormalization in substrate-supported graphene nanoribbons. J Phys Chem Lett 7(8):1526–1533. https://doi.org/10.1021/acs.jpclett.6b00422

Kim K, Choi JY, Kim T, Cho SH, Chung HJ (2011) A role for graphene in silicon-based semiconductor devices. Nature 479(7373):338. https://doi.org/10.1038/nature10680

Kimouche A, Ervasti MM, Drost R, Halonen S, Harju A, Joensuu PM, Sainio J, Liljeroth P (2015) Ultra-narrow metallic armchair graphene nanoribbons. Nat Commun 6:10177. https://doi.org/10.1038/ncomms10177

Klein DJ (1994) Graphitic polymer strips with edge states. Chem Phys Lett 217(3):261–265. https://doi.org/10.1016/0009-2614(93)E1378-T

Knight JW, Wang X, Gallandi L, Dolgounitcheva O, Ren X, Ortiz JV, Rinke P, Körzdörfer T, Marom N (2016) Accurate ionization potentials and electron affinities of acceptor molecules III: a benchmark of GW methods. J Chem Theory Comput 12(2):615–626. https://doi.org/10.1021/acs.jctc.5b00871

Lam SC, Needs RJ (1993) First-principles calculations of the screening of electric fields at the aluminium (111) and (110) surfaces. J Phys Condens Matter 5(14):2101. https://doi.org/10.1088/0953-8984/5/14/009

Lejaeghere K, Bihlmayer G, Björkman T, Blaha P, Blügel S, Blum V, Caliste D, Castelli IE, Clark SJ, Corso AD, de Gironcoli S, Deutsch T, Dewhurst JK, Marco ID, Draxl C, Dułak M, Eriksson O, Flores-Livas JA, Garrity KF, Genovese L, Giannozzi P, Giantomassi M, Goedecker S, Gonze X, Grånäs O, Gross EKU, Gulans A, Gygi F, Hamann DR, Hasnip PJ, Holzwarth NaW, Iuşan D, Jochym DB, Jollet F, Jones D, Kresse G, Koepernik K, Küçükbenli E, Kvashnin YO, Locht ILM, Lubeck S, Marsman M, Marzari N, Nitzsche U, Nordström L, Ozaki T, Paulatto L, Pickard CJ, Poelmans W, Probert MIJ, Refson K, Richter M, Rignanese GM, Saha S, Scheffler M, Schlipf M, Schwarz K, Sharma S, Tavazza F, Thunström P, Tkatchenko A, Torrent M, Vanderbilt D, van Setten MJ, Speybroeck VV, Wills JM, Yates JR, Zhang GX, Cottenier S (2016) Reproducibility in density functional theory calculations of solids. Science 351(6280):aad3000. https://doi.org/10.1126/science.aad3000

Li X, Shi Y, Gutman I (2012) The chemical connection. In: Li X, Shi Y, Gutman I (eds) Graph energy. Springer, New York, pp 11–17

Li Y, Zhang W, Morgenstern M, Mazzarello R (2013) Electronic and magnetic properties of zigzag graphene nanoribbons on the (111) surface of Cu, Ag, and Au. Phys Rev Lett 110(21):216804. https://doi.org/10.1103/PhysRevLett.110.216804

Lieb EH (1989) Two theorems on the hubbard model. Phys Rev Lett 62(16):1927–1927. https://doi.org/10.1103/PhysRevLett.62.1927.5

Liu P, Kaltak M, Klimeš J, Kresse G (2016) Cubic scaling GW: towards fast quasiparticle calculations. Phys Rev B 94(16):165109. https://doi.org/10.1103/PhysRevB.94.165109

Llinas JP, Fairbrother A, Barin GB, Shi W, Lee K, Wu S, Choi BY, Braganza R, Lear J, Kau N, Choi W, Chen C, Pedramrazi Z, Dumslaff T, Narita A, Feng X, Müllen K, Fischer F, Zettl A, Ruffieux P, Yablonovitch E, Crommie M, Fasel R, Bokor J (2017) Short-channel field-effect transistors with 9-atom and 13-atom wide graphene nanoribbons. Nat Commun 8(1):633. https://doi.org/10.1038/s41467-017-00734-x

Lowe JP, Kafafi SA, LaFemina JP (1986) Qualitative MO theory of some ring and ladder polymers. J Phys Chem 90(25):6602–6610. https://doi.org/10.1021/j100283a004

Maggio E, Kresse G (2017) GW vertex corrected calculations for molecular systems. J Chem Theory Comput 13:4765–4778

Marinopoulos AG, Reining L, Rubio A, Vast N (2003) Optical and loss spectra of carbon nanotubes: depolarization effects and intertube interactions. Phys Rev Lett 91(4):046402. https://doi.org/10.1103/PhysRevLett.91.046402

Martin RM (2004) Electronic structure: basic theory and practical methods. Cambridge University Press, Cambridge

Martin RM, Reining L, Ceperley DM (2016) Interacting electrons: theory and computational approaches. Cambridge University Press, New York

Mermin N, Wagner H (1966) Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic heisenberg models. Phys Rev Lett 17(22):1133–1136

Metzner W, Vollhardt D (1989) Correlated lattice fermions in d = ∞ dimensions. Phys Rev Lett 62(3):324–327. https://doi.org/10.1103/PhysRevLett.62.324

Nakada K, Fujita M, Dresselhaus G, Dresselhaus MS (1996) Edge state in graphene ribbons: nanometer size effect and edge shape dependence. Phys Rev B 54(24):17954–17961. https://doi.org/10.1103/PhysRevB.54.17954

Narita A, Feng X, Hernandez Y, Jensen SA, Bonn M, Yang H, Verzhbitskiy IA, Casiraghi C, Hansen MR, Koch AHR, Fytas G, Ivasenko O, Li B, Mali KS, Balandina T, Mahesh S, De Feyter S, Müllen K (2014) Synthesis of structurally well-defined and liquid-phase-processable graphene nanoribbons. Nat Chem 6(2):126–132. https://doi.org/10.1038/nchem.1819

Narita A, Feng X, Müllen K (2015) Bottom-up synthesis of chemically precise graphene nanoribbons. Chem Rec 15(1):295–309. https://doi.org/10.1002/tcr.201402082

Novoselov KS, Geim AK, Morozov SV, Dubonos SV, Zhang Y, Jiang D (2004) Room-temperature electric field effect and carrier-type inversion in graphene films. https://arxiv.org/abs/cond-mat/0410631

Onida G, Schmidt W, Pulci O, Palummo M, Marini a, Hogan C, Del Sole R (2001) Theory for modeling the optical properties of surfaces. Phys Status Solidi (a) 188(4):1233–1242. https://doi.org/10.1002/1521-396X(200112)188:4<1233::AID-PSSA1233>3.0.CO;2-W

Pauling L, Brockway LO, Beach JY (1935) The dependence of interatomic distance on single bond-double bond resonance1. J Am Chem Soc 57(12):2705–9. https://doi.org/10.1021/ja01315a105

Perdew JP, Levy M (1983) Physical content of the exact Kohn-Sham orbital energies: band gaps and derivative discontinuities. Phys Rev Lett 51(6):1884–1887. https://doi.org/10.1103/PhysRevLett.51.1884

Pisana S, Lazzeri M, Casiraghi C, Novoselov KS, Geim AK, Ferrari AC, Mauri F (2007) Breakdown of the adiabatic born–oppenheimer approximation in graphene. Nat Mater 6(3):198–201. https://doi.org/10.1038/nmat1846

Poncé S, Antonius G, Boulanger P, Cannuccia E, Marini a, Côté M, Gonze X (2014) Verification of first-principles codes: comparison of total energies, phonon frequencies, electron–phonon coupling and zero-point motion correction to the gap between ABINIT and QE/Yambo. Comput Mater Sci 83:341–348. https://doi.org/10.1016/j.commatsci.2013.11.031

Prezzi D, Varsano D, Ruini A, Marini A, Molinari E (2008) Optical properties of graphene nanoribbons: the role of many-body effects. Phys Rev B 77(4):041404. https://doi.org/10.1103/PhysRevB.77.041404

Reich S, Maultzsch J, Thomsen C, Ordejón P (2002) Tight-binding description of graphene. Phys Rev B 66(3):035412. https://doi.org/10.1103/PhysRevB.66.035412

Ruffieux P, Wang S, Yang B, Sánchez-Sánchez C, Liu J, Dienel T, Talirz L, Shinde P, Pignedoli CA, Passerone D, Dumslaff T, Feng X, Müllen K, Fasel R (2016) On-surface synthesis of graphene nanoribbons with zigzag edge topology. Nature 531(7595):489–492. https://doi.org/10.1038/nature17151

Rydberg H, Dion M, Jacobson N, Schröder E, Hyldgaard P, Simak S, Langreth D, Lundqvist BI (2003) Van der waals density functional for layered structures. Phys Rev Lett 91(12):126402. https://doi.org/10.1103/PhysRevLett.91.126402

Saito R, Dresselhaus G, Dresselhaus MS (1998) Tight binding calculation of molecules and solids. In: Physical properties of carbon nanotubes. Imperial College Press, London, pp 17–33

Schwierz F (2010) Graphene transistors. Nat Nanotechnol 5(7):487–96. https://doi.org/10.1038/nnano.2010.89

Schwierz F (2013) Graphene transistors: status, prospects, and problems. Proc IEEE 101(7):1567–1584. https://doi.org/10.1109/JPROC.2013.2257633

Son YW, Cohen ML, Louie SG (2006) Energy gaps in graphene nanoribbons. Phys Rev Lett 97(21):216803. https://doi.org/10.1103/PhysRevLett.97.216803

Sorella S, Tosatti E (1992) Semi-metal-insulator transition of the hubbard model in the honeycomb lattice. EPL (Europhys Lett) 19(8):699. https://doi.org/10.1209/0295-5075/19/8/007

Spataru CD (2013) Electronic and optical gap renormalization in carbon nanotubes near a metallic surface. Phys Rev B 88(12):125412. https://doi.org/10.1103/PhysRevB.88.125412

Stampfer C, Güttinger J, Hellmüller S, Molitor F, Ensslin K, Ihn T (2009) Energy gaps in etched graphene nanoribbons. Phys Rev Lett 102(5):056403. https://doi.org/10.1103/PhysRevLett.102.056403

Stein T, Eisenberg H, Kronik L, Baer R (2010) Fundamental gaps in finite systems from eigenvalues of a generalized Kohn-Sham method. Phys Rev Lett 105(26):266802. https://doi.org/10.1103/PhysRevLett.105.266802

Talirz L, Söde H, Cai J, Ruffieux P, Blankenburg S, Jafaar R, Berger R, Feng X, Müllen K, Passerone D, Fasel R, Pignedoli CA (2013) Termini of bottom-up fabricated graphene nanoribbons. J Am Chem Soc 135(6):2060–2063. https://doi.org/10.1021/ja311099k

Talirz L, Ruffieux P, Fasel R (2016a) On-surface synthesis of atomically precise graphene nanoribbons. Adv Mater 28(29):6222–6231. https://doi.org/10.1002/adma.201505738

Talirz L, Shinde P, Passerone D, Pignedoli CA (2016b) Synthesis of atomically precise graphene-based nanostructures: a simulation point of view. In: Gourdon A (ed) On-surface synthesis, advances in atom and single molecule machines. Springer International Publishing, Cham

Talirz L, Söde H, Dumslaff T, Wang S, Sanchez-Valencia JR, Liu J, Shinde P, Pignedoli CA, Liang L, Meunier V, Plumb NC, Shi M, Feng X, Narita A, Müllen K, Fasel R, Ruffieux P (2017) On-surface synthesis and characterization of 9-atom wide armchair graphene nanoribbons. ACS Nano. https://doi.org/10.1021/acsnano.6b06405

Tamblyn I, Darancet P, Quek SY, Bonev SA, Neaton JB (2011) Electronic energy level alignment at metal-molecule interfaces with a GW approach. Phys Rev B 84(20):201402. https://doi.org/10.1103/PhysRevB.84.201402

Tanaka K, Yamabe T (1985) Electronic structure of conductive conjugated systems and their physicochemical properties. In: Löwdin PO (ed) Advances in quantum chemistry, vol 17. Academic Press, pp 251–284. https://doi.org/10.1016/S0065-3276(08)60304-4

Tanaka K, Yamashita S, Yamabe H, Yamabe T (1987) Electronic properties of one-dimensional graphite family. Synth Met 17(1–3):143–148. https://doi.org/10.1016/0379-6779(87)90729-6

Tersoff JD (1989) Method for the calculation of scanning tunneling microscope images and spectra. Phys Rev B 40(17):11990–11993. https://doi.org/10.1103/PhysRevB.40.11990

Thomann H, Dalton LK, Grabowski M, Clarke TC (1985) Direct observation of Coulomb correlation effects in polyacetylene. Phys Rev B 31(5):3141–3143. https://doi.org/10.1103/PhysRevB.31.3141

Trevisanutto PE, Giorgetti C, Reining L, Ladisa M, Olevano V (2008) Ab initio *GW* many-body effects in graphene. Phys Rev Lett 101(22):226405. https://doi.org/10.1103/PhysRevLett.101.226405

van Schilfgaarde M, Kotani T, Faleev S (2006) Quasiparticle self-consistent *GW* theory. Phys Rev Lett 96(22):226402. https://doi.org/10.1103/PhysRevLett.96.226402

van Setten MJ, Caruso F, Sharifzadeh S, Ren X, Scheffler M, Liu F, Lischner J, Lin L, Deslippe JR, Louie SG, Yang C, Weigend F, Neaton JB, Evers F, Rinke P (2015) GW100: benchmarking G0W0 for molecular systems. J Chem Theory Comput 11(12):5665–5687. https://doi.org/10.1021/acs.jctc.5b00453

Vo TH, Shekhirev M, Kunkel DA, Morton MD, Berglund E, Kong L, Wilson PM, Dowben PA, Enders A, Sinitskii A (2014) Large-scale solution synthesis of narrow graphene nanoribbons. Nat Commun 5:3189. https://doi.org/10.1038/ncomms4189

Volpilhac G, Hoarau J (1978) Effect of the interlayer interactions on the diamagnetism of graphitic ribbons. Phys Rev B 17(3):1445–1449. https://doi.org/10.1103/PhysRevB.17.1445

Wakabayashi K, Sasaki Ki, Nakanishi T, Enoki T (2010) Electronic states of graphene nanoribbons and analytical solutions. Sci Technol Adv Mater 11(5):054504. https://doi.org/10.1088/1468-6996/11/5/054504

Wallace PR (1947) The band theory of graphite. Phys Rev 71(9):622–634. https://doi.org/10.1103/PhysRev.71.622

Wang S, Talirz L, Pignedoli CA, Feng X, Müllen K, Fasel R, Ruffieux P (2016a) Giant edge state splitting at atomically precise graphene zigzag edges. Nat Commun 7:11507. https://doi.org/10.1038/ncomms11507

Wang W, Yazyev O, Meng S, Kaxiras E (2009) Topological frustration in graphene nanoflakes: magnetic order and spin logic devices. Phys Rev Lett 102(15):157201. https://doi.org/10.1103/PhysRevLett.102.157201

Wang Z, Ki DK, Khoo JY, Mauro D, Berger H, Levitov LS, Morpurgo AF (2016b) Origin and magnitude of 'designer' spin-orbit interaction in graphene on semiconducting transition metal dichalcogenides. Phys Rev X 6(4):041020. https://doi.org/10.1103/PhysRevX.6.041020

Wassmann T, Seitsonen AP, Saitta AM, Lazzeri M, Mauri F (2010) Clar's theory, π-electron distribution, and geometry of graphene nanoribbons. J Am Chem Soc 132(10):3440–3451. https://doi.org/10.1021/ja909234y

White CT, Li J, Gunlycke D, Mintmire JW (2007) Hidden one-electron interactions in carbon nanotubes revealed in graphene nanostrips. Nano Lett 7(3):825–830. https://doi.org/10.1021/nl0627745

Wilhelm J, Golze D, Talirz L, Hutter J, Pignedoli CA (2018) Toward GW calculations on thousands of atoms. J Phys Chem Lett 9(2):306–312. https://doi.org/10.1021/acs.jpclett.7b02740

Yamabe T, Tanaka K, Ohzeki K, Yata S (1982) Electronic structure of polyacenacene. A one-dimensional graphite. Solid State Commun 44(6):823–825. https://doi.org/10.1016/0038-1098(82)90282-4

Yang L, Park CH, Son YW, Cohen ML, Louie SG (2007) Quasiparticle energies and band gaps in graphene nanoribbons. Phys Rev Lett 99(18):186801. https://doi.org/10.1103/PhysRevLett.99.186801

Yang X, Dou X, Rouhanipour A, Zhi L, Räder HJ, Müllen K (2008) Two-dimensional graphene nanoribbons. J Am Chem Soc 130(13):4216–4217. https://doi.org/10.1021/ja710234t

Yazyev OV (2010) Emergence of magnetism in graphene materials and nanostructures. Rep Prog Phys 73(5):056501. https://doi.org/10.1088/0034-4885/73/5/056501

Yazyev OV, Katsnelson M (2008) Magnetic correlations at graphene edges: basis for novel spintronics devices. Phys Rev Lett 100(4):047209. https://doi.org/10.1103/PhysRevLett.100.047209

Zhu X, Su H (2011) Scaling of excitons in graphene nanoribbons with armchair shaped edges. J Phys Chem A 115(43):11998–12003. https://doi.org/10.1021/jp202787h

Part VI
Thermal Transport

Thermal Transport: Overview

32

Davide Donadio

Contents

Abstract

This chapter focuses on recent progress in the modeling of conductive thermal transport in a variety of systems and materials, from crystalline solids and nanostructures to molecular junctions, composites, and biological systems. Novel emergent technologies in energy harvesting and conversion, microelectronics, phase-change memories, heat-assisted magnetic recording, sensing, and medical therapy and diagnostic have prompted the extension of thermal transport modeling beyond traditional bulk crystalline materials and heavily parameterized models. Major advances have been made possible by the development of fundamental theories and computational approaches, which are discussed in the following chapters.

D. Donadio (✉)
Department of Chemistry, University of California Davis, Davis, USA

IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
e-mail: ddonadio@ucdavis.edu

1   Introduction

The first macroscopic theory of heat transport was put forward by J.-B. Joseph Fourier in his essay *The Analytical Theory of Heat*, published in 1822. The partial differential equation that controls heat propagation in a medium remains the cornerstone of nowadays studies of heat transport. In nature, thermal energy can be transferred through three different mechanisms: conduction, convection, and radiation. Since in condensed phases and at the molecular scale convection and radiation are by far less relevant, chapters in this section focus on conductive heat transfer. According to Fourier's early observations, in stationary conditions conductive heat transport is controlled by a proportionality between the heat current $\vec{J}$ and the temperature gradient in the material ∇T:

$$\vec{J} = -\kappa \nabla T. \tag{1}$$

This simple relation is the well-known Fourier's law. The proportionality factor κ is the thermal conductivity, which is an intrinsic property of the material. To compute κ one needs a microscopic theory that describes the conduction of heat carriers, which are electrons and lattice vibrations, i.e., phonons.

In 1929 Rudolf Peierls formulated a microscopic theory of lattice heat transport, based on the assumptions that phonons obey the Boltzmann transport equation (BTE) (Peierls 1929). Later on lattice thermal conductivity, as well as several other transport coefficients, was expressed within linear response (LR) theory in terms of correlation functions by Green and Kubo (Green 1960; Kubo et al. 1957; Zwanzig 1965). BTE and LR laid the foundations of thermal transport modeling for the years to come. In spite of substantial theoretical advances in lattice dynamics and the rise of molecular dynamics, since the 1960s and for several decades, the calculation of the thermal conductivity of materials was undertaken using kinetic models that heavily rely on empirical parameters (Callaway 1959; Holland 1963).

In the past two decades, following the rise of nanotechnology and its potential application to electronics and energy materials, thermal transport has become the subject of intensive efforts in the communities of condensed matter physicists and materials scientists, with impressive progress on the experimental side (Cahill et al. 2003, 2014). New or significantly improved experimental techniques, such as time-domain thermoreflectance (Cahill 2004), Raman thermometry (Reparaz et al. 2014), scanning thermal microscopy (Grosse et al. 2011), and thermal spectroscopy (Siemens et al. 2009; Minnich et al. 2011), provide novel insight into the physics of heat transport, thus enabling fundamental discoveries and paving the way to the design of materials with tailored thermal conductivity.

Among the most remarkable discoveries in the field of thermal transport in materials in the past two decades, it is worth mentioning the measurement of the thermal conductivity of low-dimensional nanostructures (Li et al. 2003; Pop et al. 2006; Ghosh et al. 2010; Bae et al. 2013); the finding of ultralow thermal conductivity, yet crystalline, materials (Chiritescu et al. 2007); the engineering of

nanostructured materials with low κ for thermoelectric energy conversion (Yu et al. 2010; Biswas et al. 2012); and the design of high thermal conductivity polymers (Shen et al. 2010). Such achievements, which were often the result of combined theoretical and experimental efforts, boosted the development of *phononics* as a research field on its own, which addresses the control of acoustic and thermal vibrations in solids or liquids (Balandin 2005; Maldovan 2013; Volz et al. 2016).

The predictive bottom-up calculation of the thermal conductivity remains a very difficult problem in materials modeling. Modeling efforts were also crowned with success, for example, with the development of first-principles BTE, which enables the calculation of κ for simple bulk crystalline materials (Broido et al. 2007; Ward et al. 2009) and nanostructures, e.g., graphene and other 2D materials (Fugallo et al. 2014). Remarkable progress in molecular dynamics (MD) simulations has allowed researchers to tackle heat transport in complex systems, e.g., nanostructured materials (Selli et al. 2016), biomolecules (Leitner 2008), and polymer composites (Desai et al. 2005). Furthermore, today's high-performance computers and MD codes on the theory side, combined with the refinement of nanoscale fabrication on the experimental side, enable the direct comparison of atomistic simulations to experimental measurements over the same length scales. Yet, modeling heat transport in materials still presents major challenges that need to be overcome, both at a very fundamental level and at the level of addressing the actual complexity of realistic systems for technological applications.

The chapters of this section assess the state of the art of modeling thermal transport in materials. They not only provide an overview of applications to emerging materials and complex systems but also report on recent theoretical and methodological advances, since the former would not have been possible without the latter. The first three chapters address recent advances in theory and methods that enable the calculation of the thermal conductivity and of the detailed phonon properties of relatively simple materials and nanostructures by first principles, utilizing either lattice dynamics or molecular dynamics. The following three chapters address the theory and calculation of the thermal boundary resistance of various types of interfaces. The last three chapters address the application of molecular and mesoscale modeling to complex systems, such as alloys, metamaterials, biomolecules, and polymer composites.

2 Ab Initio Theory of Thermal Transport

Fourier's law clearly indicates that, in order to unravel thermal transport in materials, one has to be able to calculate the thermal conductivity. The microscopic ingredients that determine the thermal conductivity in electronic insulators and semiconductors are outlined by the kinetic theory (Ziman 1960), in which κ is expressed as a sum over the phonon branches and the momenta in the first Brillouin zone of the phonon heat capacity c_μ, group velocity v_μ, and mean free path λ_μ:

$$\kappa = \sum_{\mu}^{N_{phonons}} c_{\mu} v_{\mu} \lambda_{\mu} \tag{2}$$

where μ contracts the phonon branch i index and the momentum $\mathbf{q}$. Yet, a predictive calculation of these properties is far from straightforward.

In bulk crystals at low temperature, it is now possible to rely on lattice dynamics and compute phonon dispersion relations, which give c_{μ} and v_{μ}, by density functional perturbation theory (Baroni et al. 2001). Perturbation theory also allows the calculation of anharmonic force constants, and by projecting on the normal modes, one can get three-phonon scattering rates, which can be used to compute relaxation times by solving the linearized BTE. Using this approach and building on the pioneering work by Omini and Sparavigna (Omini and Sparavigna 1995), Broido and co-workers showcased the first first-principles calculations of the thermal conductivity of silicon and germanium (Broido et al. 2007). This approach has become the state of the art in thermal conductivity calculations of solids, proving efficient and predictive. For example it leads to the prediction of new high thermal conductivity materials (Lindsay et al. 2013), and it can be used effectively for high-throughput screening and materials design (Curtarolo et al. 2013). First-principles BTE also provides a detailed insight into the heat transport, as it breaks down the contribution to κ in single carrier properties, and it aids the interpretation of recent thermal transport spectroscopy experiments: for example, it unraveled the contribution of phonons with very long mean free paths, beyond $1\,\mu m$, to the thermal conductivity of bulk silicon even at room temperature (Regner et al. 2013). The linearized BTE approach has been used to calculate the thermal conductivity of low-dimensional systems. In this context it is essential to consider collective hydrodynamic effects, which lead to non-diffusive behavior over a broad temperature range (Lee et al. 2015; Cepellotti et al. 2015; Cepellotti and Marzari 2016)). These studies ascertained the thermal transport regime and the length dependence of κ in two-dimensional materials, thus resolving discrepancies in experimental measurements (Balandin 2011), indicating a finite theoretical bulk limit for κ in 2D materials, and providing a rationale for the length dependence of κ in graphene (Fugallo et al. 2014). Detailed accounts of first-principles BTE in different contexts are provided in ▶ Chaps. 33, "Thermal Transport by First-Principles Anharmonic Lattice Dynamics" and ▶ 34, "On the Kinetic Theory of Thermal Transport in Crystals" in this section.

Originally the applicability of the BTE approach was limited to crystalline materials at low temperatures, where the lattice dynamics is represented in terms of harmonic normal modes and anharmonic effects can be truncated to three-phonon scattering. However, efforts have been deployed to extend the capability of BTE to nanostructured systems with large unit cells, such as silicon-germanium superlattices and crystalline porous carbon (Savic et al. 2013; Pereira et al. 2013), and promising results have been recently achieved for materials at high temperature (van Roekeghem et al. 2016). Yet, BTE cannot be straightforwardly applied to disordered systems, such as glasses and liquids, for which MD is a more natural choice. Using MD, κ can be computed either from equilibrium (Ladd et al. 1986)

or non-equilibrium simulations (Müller-Plathe 1997). While the two approaches have been proven to be nearly equivalent (Schelling et al. 2002), the former is more suitable to study heat transport in extended systems, while the latter applies straightforwardly to finite systems and interfaces.

Thermal conductivity, as other transport coefficients, can be computed in equilibrium MD simulation through Green-Kubo theory. κ is expressed as the limit to infinite size and infinite time of the integral of the heat current autocorrelation function:

$$\kappa_{\alpha\beta} = \lim_{V \to \infty} \lim_{t \to \infty} \frac{1}{k_B T^2 V} \int_0^t dt' \langle J_\alpha(0) J_\beta(t') \rangle, \tag{3}$$

where κ is 3×3 tensor and $J_{\alpha,\beta}$ are the components of the heat current. Besides carefully taking care of size convergence and sufficient sampling, to calculate κ one needs an expression for the heat current, which is nontrivial for many-body empirical potentials (Fan et al. 2015). While ab initio MD simulations were long deemed unsuitable to evaluate (3), an expression for J was recently derived in the framework of density functional theory (Marcolongo et al. 2015). This breakthrough paves the way to the ab initio calculation of thermal transport in previously intractable materials such as geological fluids, glasses, and dynamically stabilized crystals, as well as condensed matter close to phase transitions. ▶ Chapter 35, "Heat Transport in Insulators from Ab Initio Green-Kubo Theory" reviews these recent advances in ab initio MD calculations of heat transport.

3   Interfaces

Regardless whether they occur spontaneously, e.g., in polycrystalline solids, or they are specifically engineered, interfaces are a distinguishing feature of complex materials. Thermal transport in real materials is often dictated by the thermal boundary resistance (TBR) of interfaces. Moreover, TBR has a major impact on several technologies, such as heat-assisted magnetic recording, heat management in microelectronic devices, and reading and writing processes of phase-change materials (PCM) memories. Heat transport at interfaces between hard and soft matter systems has become a subject of numerous studies in the last decades for its technological and medical implications.

As observed by Kapitza in 1941 (Kapitza 1941), given a thermal current (J) at the ideally abrupt interface between two different materials, a discontinuity ΔT occurs in the temperature profile. The so-called Kapitza resistance (R) is given by the ratio between the temperature drop and the flux:

$$R = -\frac{\Delta T}{J}. \tag{4}$$

Knowing the bulk phonon properties of the two materials forming the interface, it is possible to make an approximate estimate of the thermal boundary resistance via

acoustic mismatch or diffusive mismatch models (Swartz and Pohl 1989). However, these approaches imply severe approximations, and predictions are usually not quantitatively accurate. Furthermore, the definition of Kapitza resistance can be practically extended to more broadly define interfacial regions with finite thickness and even further to open systems in general. These can be generically described as systems in which two semi-infinite periodic thermal reservoirs are connected by a nonperiodic "device," as, for example, in molecular junctions (Galperin et al. 2007).

Hence, concurrently with the development of approaches to calculate the bulk thermal conductivity of extended systems, bottom-up atomistic methods and theories were developed to compute thermal boundary resistance and heat transport in open systems. LD and MD simulations can be exploited to calculate R. LD methods rely either on scattering theory (Young and Maris 1989) or on Green's functions (Fagas et al. 1999). In analogy with electronic transport, these calculations provide frequency-resolved phonon transmission function, which sheds light into the details of heat transport and can be exploited to design the thermal conductance of junctions or interfaces. The total thermal conductance is computed integrating Landauer's expression over the whole phonon spectrum (Imry and Landauer 1999). LD approaches allow one to treat open systems at any temperature, including the correct quantum populations, and can be implemented for large. However, while it is possible to take into account anharmonicity, the involved computational complexity makes LD approaches inconvenient to treat strongly anharmonic systems (Mingo 2006). MD, in turn, can be applied straightforwardly in its non-equilibrium form, and TBR can be computed from Eq. 4 by directly evaluating the heat current and the temperature drop at the interface. It is necessary to point out that classical MD cannot take into account quantum effects, and it is therefore unsuitable to treat systems at low temperatures that are below their Debye temperature. A general overview on heat transport at interfaces, including lattice dynamics models and non-equilibrium MD, is given in ▶ Chaps. 36, "Lattice Thermal Boundary Resistance." and ▶ 37, "Energy Relaxation and Thermal Transport in Molecules," gives a perspective on the use of Landauer's approach in modeling molecular junctions and interfaces between inorganic substrates and self-assembled monolayers. These are systems that have stimulated growing interest among materials scientists and engineers in the past decade (Reddy et al. 2007; Losego et al. 2012; Ong et al. 2013).

Equilibrium MD and the Green-Kubo formalism turns out to be a very versatile tool, and it can be casted in such a way as to calculate interfacial thermal conductance from temperature fluctuations (Chalopin and Volz 2013). Furthermore, linear response theory enables the calculation of the phonon dispersion relations and of the spectral transmission function across interfaces. Equilibrium MD is particularly suitable to simulate not only solid-solid junctions but also interfaces between solids and liquids or soft matter, as it avoids a number of algorithmic issues occurring while establishing a stationary heat flux in non-equilibrium MD, such as thermostatting diffusive regions of the simulation system. ▶ Chapter 38, "A Statistical Approach of Thermal Transport at Nanoscales: From Solid-State to Biological Applications," provides a detailed account of Green-Kubo theory and

calculations of thermal boundary resistance in a variety of complex systems, paving the way to simulations of heat transfer at bio-inorganic interfaces, relevant for medical applications.

4   Toward Advanced Materials and Complex Systems

The methods described in the previous sections allow for accurate and predictive heat transport calculations of the thermal conductivity of systems, for which heat transport had remained poorly understood for several decades. Alloys are a good example of such systems: SiGe alloys have been used since the 1960s by NASA for thermoelectric power generation, but only recent calculations started to unravel heat transport in such systems (Garg et al. 2011). Following the original ideas by Hicks and Dresselhaus (Hicks and Dresselhaus 1993; Dresselhaus et al. 2007), combining alloying with nanostructuring was proposed to lead to ultralow thermal conductivity (He et al. 2011). The thermal conductivity and the thermoelectric performance of silicon-based alloys can be further optimized by engineering their nanoscale structure and their chemical composition. Lattice dynamics calculations can be applied effectively to alloys, nanocomposites, and superlattices as a tool to guide such systematic optimization (Aksamija and Knezevic 2010; Savic et al. 2013; Khatami and Aksamija 2016). ▶ Chapter 39, "Thermal Conductivity of Nanostructured Semiconductor Alloys" reviews progresses in the simulation of binary and ternary Si/Ge/Sn alloys using LD and BTE.

Further refinements in designing materials with tailored phononic properties can be achieved by relying upon concepts like phononic crystals (Gorishnyy et al. 2005) and metamaterials (Pendry et al. 1996). Phononic crystals consist of materials in which regular features, e.g., pores, tailor the dispersion relations of acoustic or supersonic waves. Metamaterials are composite materials that exhibit properties that do not pertain to their single components, or, in general, to natural materials. The concept was originally developed to achieve exotic optical properties, such as negative refractive index (Pendry 2006), and was eventually adopted for acoustic properties (Liu et al. 2000). Both phononic crystals and nanophononic metamaterials were recently proposed as a way to reduce the thermal conductivity of silicon thin films or membranes and to improve their thermoelectric performance (Yu et al. 2010; Davis and Hussein 2014). In the case of thermal transport, metamaterials can be obtained by introducing surface nanostructures (Davis and Hussein 2014; Xiong et al. 2017), so as to induce localized resonances that hybridize with phonons, thus reducing the capability of transferring energy. This approach is discussed in ▶ Chap. 40, "Resonant Thermal Transport in Nanophononic Metamaterials." The concept of nanophononic metamaterials is fascinating, as it may be applied not only to optimize thermoelectric energy conversion but also for phonon cloaking, filtering, and ultimately advanced optomechanical systems.

Thanks to the ever-growing computational power and the development of efficient parallel algorithms, MD simulations with either empirical potentials or classical force fields enable the study of heat transport in systems of up to millions

of atoms, thus allowing one to tackle complex materials such as polymers and their nanocomposites. These simulations allowed researchers to address a long-standing challenge in materials science that is unraveling the design principles to create structural plastics with high thermal conductivity. While polymeric materials normally feature very low thermal conductivity, MD simulations showed that single polyethylene molecular chains have very high κ (Henry and Chen 2008). This result was eventually used to fabricate high thermal conductivity polyethylene nanofibers, in which molecular chains are aligned by drawing (Shen et al. 2010). While these early results are encouraging, simulating heat transport in polymers and their composites still faces several challenges. For example, it is still difficult to obtain representative molecular models of polymer melts, due to the multiple size scales involved in such systems and very long relaxation times. In addition, in the case of composites, the issue of thermal boundary resistance between the solid micro- or nanoparticles and the polymer matrix intertwines with the collective modes of the system. This complex phenomenology, which is so far just partially understood, has hindered the optimization of high thermal conductivity composites. ▶ Chapter 41, "Modeling of Heat Transport in Polymers and their Nanocomposites" provides a detailed overview of successes and open challenges in the simulation of polymers and their composites.

5   Conclusion

In conclusion, this section is a journey through recent achievements in the simulation of heat transport in emergent materials, which have been enabled by major theoretical and methodological developments. Some of the methods and the ideas discussed in the following chapters are very recent, and, so far, we are just scratching the surface of the possibility that they offer. In spite of such advances in both simulation and measurement techniques, several recent experiments raised new questions about the very nature of thermal transport both in nanoscale systems and in complex materials. Similarly, numerical simulations of apparently simple nonlinear models, such as the Fermi-Pasta-Ulam model, suggest universal behaviors of low-dimensional systems, which are difficult to transpose to real materials (Lepri et al. 2003). More fundamental work is needed to unravel issues such as the transport regime at the nanoscale, phonon coherence, energy propagation in amorphous materials, anharmonic effect on thermal boundary resistance, and thermal rectification. In this respect it is still reasonable to say that, as Peierls himself put it in the early 1960s, "it seems there is no problem in modern physics, for which there are on record as many false starts, and as many theories which overlook some essential feature, as in the problem of the thermal conductivity of nonconducting materials."

References

Aksamija Z, Knezevic I (2010) Anisotropy and boundary scattering in the lattice thermal conductivity of silicon nanomembranes. Phys Rev B 82(4):045319

Bae MH, Li Z, Aksamija Z, Martin PN, Xiong F, Ong ZY, Knezevic I, Pop E (2013) Ballistic to diffusive crossover of heat flow in graphene ribbons. Nat Commun 4:1734

Balandin AA (2005) Nanophononics: phonon engineering in nanostructures and nanodevices. J Nanosci Nanotechnol 5(7):1015–1022

Balandin AA (2011) Thermal properties of graphene and nanostructured carbon materials. Nat Mater 10(8):569–581

Baroni S, de Gironcoli S, Dal Corso A, Giannozzi P (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73(2):515–562

Biswas K, He J, Blum ID, Wu CI, Hogan TP, Seidman DN, Dravid VP, Kanatzidis MG (2012) High-performance bulk thermoelectrics with all-scale hierarchical architectures. Nature 489(7416):414–418

Broido DA, Malorny M, Birner G, Mingo N, Stewart DA (2007) Intrinsic lattice thermal conductivity of semiconductors from first principles. Appl Phys Lett 91(23):231922

Cahill DG (2004) Analysis of heat flow in layered structures for time-domain thermoreflectance. Rev Sci Instrum 75(12):5119–5122

Cahill DG, Ford WK, Goodson KE, Mahan GD, Majumdar A, Maris HJ, Merlin R, Phillpot SR (2003) Nanoscale thermal transport. J Appl Phys 93(2):793–818

Cahill DG, Braun PV, Chen G, Clarke DR, Fan S, Goodson KE, Keblinski P, King WP, Mahan GD, Majumdar A (2014) Nanoscale thermal transport. II. 2003–2012. Appl Phys Rev 1(1):011305

Callaway J (1959) Model for lattice thermal conductivity at low temperatures. Phys Rev Lett 113(4):1046–1051

Cepellotti A, Marzari N (2016) Thermal transport in crystals as a kinetic theory of relaxons. Phys Rev X 6(4):041013–041014

Cepellotti A, Fugallo G, Paulatto L, Lazzeri M, Mauri F, Marzari N (2015) Phonon hydrodynamics in two-dimensional materials. Nat Commun 6:6400

Chalopin Y, Volz S (2013) A microscopic formulation of the phonon transmission at the nanoscale. Appl Phys Lett 103(5):051602

Chiritescu C, Cahill DG, Nguyen N, Johnson D, Bodapati A, Keblinski P, Zschack P (2007) Ultralow thermal conductivity in disordered, layered WSe_2 crystals. Science 315(5810): 351–353

Curtarolo S, Hart GLW, Nardelli MB, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12(3):191–201

Davis BL, Hussein MI (2014) Nanophononic metamaterial: thermal conductivity reduction by local resonance. Phys Rev Lett 112(5):055505

Desai T, Keblinski P, Kumar SK (2005) Molecular dynamics simulations of polymer transport in nanocomposites. J Chem Phys 122(13):134910

Dresselhaus MS, Chen G, Tang MY, Yang R, Lee H, Wang D, Ren Z, Fleurial JP, Gogna P (2007) New directions for low-dimensional thermoelectric materials. Adv Mater 19(8):1043–1053

Fagas G, Kozorezov AG, Lambert CJ, Wigmore JK, Peacock A, Poelaert A, den Hartog R (1999) Lattice dynamics of a disordered solid-solid interface. Phys Rev B 60(9):6459–6464

Fan Z, Pereira LFC, Wang HQ, Zheng JC, Donadio D, Harju A (2015) Force and heat current formulas for many-body potentials in molecular dynamics simulations with applications to thermal conductivity calculations. Phys Rev B 92(9):3689

Fugallo G, Cepellotti A, Paulatto L, Lazzeri M, Marzari N, Mauri F (2014) Thermal conductivity of graphene and graphite: collective excitations and mean free paths. Nano Lett 14(11): 6109–6114

Galperin M, Nitzan A, Ratner MA (2007) Heat conduction in molecular transport junctions. Phys Rev B 75(15):1073

Garg J, Bonini N, Kozinsky B, Marzari N (2011) Role of disorder and anharmonicity in the thermal conductivity of silicon-germanium alloys: a first-principles study. Phys Rev Lett 106(4):045901

Ghosh S, Bao W, Nika DL, Subrina S, Pokatilov EP, Lau CN, Balandin AA (2010) Dimensional crossover of thermal transport in few-layer graphene. Nat Mater 9(7):555–558

Gorishnyy T, Ullal CK, Maldovan M, Fytas G, Thomas EL (2005) Hypersonic phononic crystals. Phys Rev Lett 94(11):2555

Green MS (1960) Comment on a paper of mori on time-correlation expressions for transport properties. Phys Rev Lett 119(3):829–830

Grosse KL, Bae MH, Lian F, Pop E, King WP (2011) Nanoscale joule heating, Peltier cooling and current crowding at graphene–metal contacts. Nat Nanotechnol 6(5):287–290

He Y, Donadio D, Galli G (2011) Morphology and temperature dependence of the thermal conductivity of nanoporous SiGe. Nano Lett 11(9):3608–3611

Henry A, Chen G (2008) High thermal conductivity of single polyethylene chains using molecular dynamics simulations. Phys Rev Lett 101(23):141

Hicks LD, Dresselhaus MS (1993) Effect of quantum-well structures on the thermoelectric figure of merit. Phys Rev B 47(19):12727–12731

Holland MG (1963) Analysis of lattice thermal conductivity. Phys Rev Lett 132(6):2461–2471

Imry Y, Landauer R (1999) Conductance viewed as transmission. Rev Mod Phys 71(2):S306–S312

Kapitza PL (1941) Heat transfer and superfluidity of helium II. Phys Rev Lett 60(4):354–355

Khatami SN, Aksamija Z (2016) Lattice thermal conductivity of the binary and ternary group-IV alloys Si-Sn, Ge-Sn, and Si-Ge-Sn. Phys Rev Appl 6:014015

Kubo R, Yokota M, Nakajima S (1957) Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance. J Phys Soc Jpn 12(11):1203–1211

Ladd AJC, Moran B, Hoover WG (1986) Lattice thermal conductivity: a comparison of molecular dynamics and anharmonic lattice dynamics. Phys Rev B 34(8):5058–5064

Lee S, Broido D, Esfarjani K, Chen G (2015) Hydrodynamic phonon transport in suspended graphene. Nat Commun 6:6290

Leitner DM (2008) Energy flow in proteins. Annu Rev Phys Chem 59:233–259

Lepri S, Livi R, Politi A (2003) Thermal conduction in classical low-dimensional lattices. Phys Rep 377(1):1–80

Li D, Wu Y, Fan R, Yang P, Majumdar A (2003) Thermal conductivity of Si/SiGe superlattice nanowires. Appl Phys Lett 83(15):3186

Lindsay L, Broido DA, Reinecke TL (2013) First-principles determination of ultrahigh thermal conductivity of boron arsenide: a competitor for diamond? Phys Rev Lett 111(2):025901–025905

Liu Z, Zhang X, Mao Y, Zhu YY, Yang Z, Chan CT, Sheng P (2000) Locally resonant sonic materials. Science 289(5485):1734–1736

Losego MD, Grady ME, Sottos NR, Cahill DG, Braun PV (2012) Effects of chemical bonding on heat transport across interfaces. Nat Mater 11(6):502–506

Maldovan M (2013) Sound and heat revolutions in phononics. Nature 503(7475):209–217

Marcolongo A, Umari P, Baroni S (2015) Microscopic theory and quantum simulation of atomic heat transport. Nat Phys 12(1):80–84

Mingo N (2006) Anharmonic phonon flow through molecular-sized junctions. Phys Rev B 74(12):1055

Minnich AJ, Johnson JA, Schmidt AJ, Esfarjani K, Dresselhaus MS, Nelson KA, Chen G (2011) Thermal conductivity spectroscopy technique to measure phonon mean free paths. Phys Rev Lett 107(9):095901

Müller-Plathe F (1997) A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. J Chem Phys 106(14):6082–6085

Omini M, Sparavigna A (1995) An iterative approach to the phonon Boltzmann equation in the theory of thermal conductivity. Phys B Condens Matter 212(2):101–112

Ong WL, Rupich SM, Talapin DV, Mcgaughey AJH, Malen JA (2013) Surface chemistry mediates thermal transport in three-dimensional nanocrystal arrays. Nat Mater 12(5):410–415

Peierls R (1929) Zur kinetischen Theorie der Wärmeleitung in Kristallen. Ann Phys 395(8):1055–1101

Pendry JB (2006) Controlling electromagnetic fields. Science 312(5781):1780–1782

Pendry JB, Holden AJ, Stewart WJ, Youngs I (1996) Extremely low frequency plasmons in metallic mesostructures. Phys Rev Lett 76(25):4773–4776

Pereira LFC, Savic I, Donadio D (2013) Thermal conductivity of one-, two- and three-dimensional sp 2carbon. New J Phys 15(10):105019

Pop E, Mann D, Wang Q, Goodson K, Dai H (2006) Thermal conductance of an individual single-wall carbon nanotube above room temperature. Nano Lett 6(1):96–100

Reddy P, Jang SY, Segalman RA, Majumdar A (2007) Thermoelectricity in molecular junctions. Science 315(5818):1568–1571

Regner KT, Sellan DP, Su Z, Amon CH, Mcgaughey AJH, Malen JA (2013) Broadband phonon mean free path contributions to thermal conductivity measured using frequency domain thermoreflectance. Nat Commun 4:1640–1647

Reparaz JS, Chávez-Ángel E, Wagner MR, Graczykowski B, Gomis-Bresco J, Alzina F, Sotomayor Torres CM (2014) A novel contactless technique for thermal field mapping and thermal conductivity determination: two-laser Raman thermometry. Rev Sci Instrum 85(3):034901–034906

Savic I, Donadio D, Gygi F, Galli G (2013) Dimensionality and heat transport in Si-Ge superlattices. Appl Phys Lett 102(7):073113

Schelling PK, Phillpot SR, Keblinski P (2002) Comparison of atomic-level simulation methods for computing thermal conductivity. Phys Rev B 65(14):517

Selli D, Boulfelfel SE, Schapotschnikow P, Donadio D, Leoni S (2016) Hierarchical thermo-electrics: crystal grain boundaries as scalable phonon scatterers. Nanoscale 8(6):3729–3738

Shen S, Henry A, Tong J, Zheng R, Chen G (2010) Polyethylene nanofibres with very high thermal conductivities. Nat Nanotechnol 5(4):251–255

Siemens ME, Li Q, Yang R, Nelson KA, Anderson EH, Murnane MM, Kapteyn HC (2009) Quasi-ballistic thermal transport from nanoscale interfaces observed using ultrafast coherent soft X-ray beams. Nat Mater 9(1):26–30

Swartz E, Pohl R (1989) Thermal boundary resistance. Rev Mod Phys 61(3):605

van Roekeghem A, Carrete J, Oses C, Curtarolo S, Mingo N (2016) High-throughput computation of thermal conductivity of high-temperature solid phases: the case of oxide and fluoride perovskites. Phys Rev X 6(4):210–216

Volz S, Ordonez-Miranda J, Shchepetov A, Prunnila M, Ahopelto J, Pezeril T, Vaudel G, Gusev V, Ruello P, Weig EM, Schubert M, Hettich M, Grossman M, Dekorsy T, Alzina F, Graczykowski B, Chavez-Angel E, Sebastian Reparaz J, Wagner MR, Sotomayor Torres CM, Xiong S, Neogi S, Donadio D (2016) Nanophononics: state of the art and perspectives. Eur Phys J B 89(1):15

Ward A, Broido DA, Stewart DA, Deinzer G (2009) Ab initiotheory of the lattice thermal conductivity in diamond. Phys Rev B 80(12):1537

Xiong S, Selli D, Neogi S, Donadio D (2017) Native surface oxide turns alloyed silicon membranes into nanophononic metamaterials with ultralow thermal conductivity. Phys Rev B 95(18):180301

Young DA, Maris HJ (1989) Lattice-dynamical calculation of the Kapitza resistance between fcc lattices. Phys Rev B 40(6):3685–3693

Yu JK, Mitrovic S, Tham D, Varghese J, Heath JR (2010) Reduction of thermal conductivity in phononic nanomesh structures. Nat Nanotechnol 5(10):718–721

Ziman JM (1960) Electrons and phonons. Oxford University Press, Oxford

Zwanzig R (1965) Time-correlation functions and transport coefficients in statistical mechanics. Annu Rev Phys Chem 16:67

Thermal Transport by First-Principles Anharmonic Lattice Dynamics

33

Lucas Lindsay and Carlos A. Polanco

Contents

Abstract

Predictive methodologies for calculating lattice thermal transport and related properties have been recently developed and benchmarked successfully against measurements in a wide range of materials, from bulk to the nanoscale. This chapter presents an overview of the theoretical underpinnings of typical numerical methods for solving the Peierls-Boltzmann equation (PBE) for phonon transport when coupled with density functional theory (DFT). We discuss how these numerical methods may be employed to advance fundamental understanding of

L. Lindsay (✉) · C. A. Polanco
Materials Science and Technology Division, Oak Ridge National Laboratory,
Oak Ridge, TN, USA
e-mail: lindsaylr@ornl.gov; polancogarca@ornl.gov

thermal transport, particularly as applied to measurements of anharmonic properties (e.g., linewidths, lattice expansion) in materials. In particular, calculations of interatomic force constants, integration methods, and construction of transport properties will be explained in detail with reference to relevant research in the literature. Furthermore, this chapter will provide an overview of applications and methods for calculations of thermal conductivities beyond perfect infinite crystals (including phonon-defect scattering) and of other measured observables, including mean free path accumulation, linewidths, and densities of states from scattering experiments.

1   Introduction

Lattice vibrational properties of crystals as governed by phonons (normal modes of vibration) play a critical role in the conversion, storage, and transmission of thermal energy, which is of significant importance to a variety of technologies, including thermoelectrics, thermal barrier coatings, electronics packaging, claddings for nuclear technologies, solid-state lighting, and batteries. Of particular relevance here is the characterization of a material's ability to conduct heat via phonons as determined by the lattice thermal conductivity κ. The study of phonons and κ has a long and storied past, yet recent advances in computational power and numerical methods are expanding our ability to predict and develop deeper insights into fundamental physics and still provide novel surprises and new understandings of these behaviors. This chapter aims to provide a basic foundation of the theoretical and computational formalism from which recent κ calculations are built, without adjustable empirical parameters, and to relate these with other experimental observables. Figure 1 gives an example calculation of κ for the rock salt structure KCl as given by the formalism described in this chapter and compared with available measured data.

2   Theoretical Underpinnings

Much of the theory presented here can be pieced together from textbook discussions (Ashcroft and Mermin 1976; Kittel 2005) of lattice dynamical behavior – two foundational resources being *Electrons and Phonons* by Ziman (2001) and *The Physics of Phonons* by G. P. Srivastava (1990). Attempting to build anharmonic lattice dynamical features from a top to bottom approach, this chapter begins with the concept of thermal energy being carried by correlated vibrations of atoms in a crystal, phonons.

2.1   Lattice Thermal Conductivity

To understand the lattice dynamics of crystalline dielectrics, Einstein proposed a theory of atoms as a set of oscillators each independently vibrating (Einstein
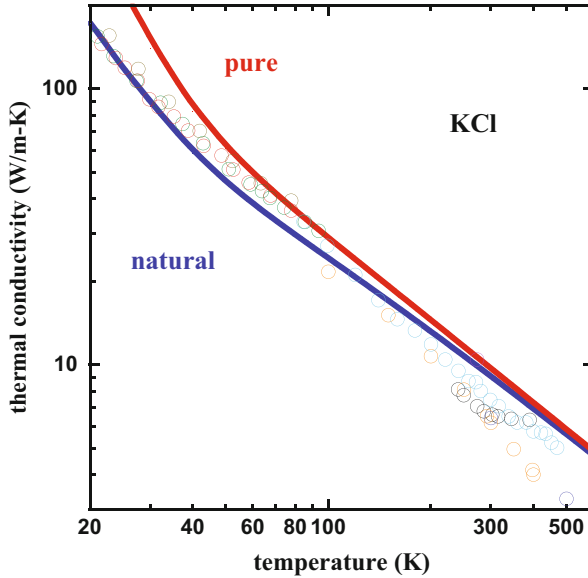
Fig. 1 Calculated κ of KCl as a function of temperature for isotopically pure material (red curve) and including naturally occurring phonon-isotope scattering (blue curve). Details of the κ calculations are described in this chapter. Colored circles give various measured κ data for KCl: black (McCarthy and Ballard 1960), green (Walker 1963), blue (Petrov et al. 1976), red (Rosenbaum et al. 1969), orange (Andersson 1985), purple (Slack and Ross 1985), brown (Peech et al. 1967), and light blue (Pettersson 1987)

1911), from which heat could flow via a random walk of thermal energy among them. This theory failed to describe specific heat and thermal conductivity, though a modification of it is often used to describe the minimum thermal conductivity of disordered or amorphous materials (Cahill and Pohl 1989; Cahill et al. 1992). The concept of atoms in the lattice collectively vibrating as coupled harmonic oscillators was introduced by Debye, where the idea of the phonon (normal modes of vibration) began (Debye 1912, 1914). This description has been very successful in describing all sorts of physical phenomena in crystals. One very important success is an accurate description and subsequent understanding of κ for crystalline solids (Srivastava 1990; Ziman 2001).

In general κ can vary spatially and temporally in a material; however, here we will assume infinite crystalline material at a particular average temperature T under the influence of a small temperature gradient $\partial T/\partial \alpha$ in the Cartesian direction α. Under these conditions, a thermal energy current J_α flows from hot to cold and is linearly proportional to the temperature gradient (Fourier's law):

$$J_\alpha = -\kappa_\alpha \frac{\partial T}{\partial \alpha} \qquad (1)$$

with the constant of proportionality being thermal conductivity, an intrinsic material property. In general, κ is a tensor quantity, though is typically described by one, two, or three components depending on crystal symmetry. Note that when discussing the thermal conductivity in general the subscript dependence will be dropped.

Taking a microscopic view, each phonon mode (labeled by wavevector $\vec{q}$ and branch index j) carries a part of the overall J_α, and κ is given as a sum over mode contributions $\kappa_{\vec{q}j\alpha}$:

$$\kappa_\alpha = \sum_{\vec{q}j} \kappa_{\vec{q}j\alpha} = \sum_{\vec{q}j} C_{\vec{q}j} v_{\vec{q}j\alpha}^2 \tau_{\vec{q}j} \tag{2}$$

where each $\kappa_{\vec{q}j\alpha}$ is a product of the phonon's mode-specific heat $C_{\vec{q}j} = n_{\vec{q}j}^0 \left(n_{\vec{q}j}^0 + 1 \right) \left(\hbar\omega_{\vec{q}j} \right)^2 / V_0 k_B T^2$ (here normalized by crystal volume V_0), velocity component in the transport direction $v_{\vec{q}j\alpha} = d\omega_{\vec{q}j}/dq_\alpha$, and lifetime $\tau_{\vec{q}j}$. Depending on crystal symmetry, this lifetime can vary with the direction of the temperature gradient and thus will sometimes be given a Cartesian subscript and called a transport lifetime $\tau_{\vec{q}j\alpha}$. This directional dependence will be disregarded here, but discussed later when relevant. $n_{\vec{q}j}^0 = 1/\left(e^{\hbar\omega_{\vec{q}j}/k_B T} - 1 \right)$ is the Bose-Einstein distribution that represents the phonon population in equilibrium, $\omega_{\vec{q}j}$ is the frequency, and k_B is Boltzmann's constant. $C_{\vec{q}j}$ and $\vec{v}_{\vec{q}j}$ are functions of the phonon frequencies alone and are thus described as harmonic properties as they arise from noninteracting harmonic phonons (Sect. 2.2). The intrinsic phonon lifetimes (Sect. 2.4) require a description of interacting phonons, thus including anharmonic terms in the interatomic potential (Sect. 2.3).

2.2 Phonons: Harmonic

A crystal is built from a base unit cell consisting of a small subset of atoms that is mapped out in space by a set of primitive lattice vectors $\vec{R}_l$ in a periodic pattern (one for each dimension of the system). Here l designates the lth unit cell and k will designate a particular atom in the unit cell. It is assumed that the atoms do not deviate significantly from their equilibrium positions (justification for the harmonic approximation) and the equations of motion are given by:

$$-\frac{\partial V_{\text{harm}}}{\partial u_{lk\alpha}} = m_k \frac{d^2 u_{lk\alpha}}{dt^2} \tag{3}$$

$u_{lk\alpha}$ is the displacement from equilibrium in the αth direction of the kth atom in the lth unit cell, m_k is the mass of this atom, and V_{harm} is the harmonic contribution to the interatomic potential energy $V = V_0 + V_1 + V_{\text{harm}} + V_{\text{anharm}}$ (lowest-order nontrivial term in a Taylor expansion of V) given by:

$$V_{\mathrm{harm}} = \frac{1}{2} \sum_{lk} \sum_{l'k'} \sum_{\alpha\beta} \Phi_{\alpha\beta}^{lkl'k'} u_{lk\alpha} u_{l'k'_\beta} \tag{4}$$

$\Phi_{\alpha\beta}^{lkl'k'} = \partial^2 V / \partial u_{lk\alpha} \partial u_{l'k'_\beta}$ are called the harmonic interatomic force constants (IFCs), and the derivatives are to be evaluated at equilibrium. Details of the calculation of these will be given in Sect. 3.1. Assuming a plane wave form for the atomic displacements $u_{lk\alpha} = \sum_{\vec{q}j} \varepsilon_{\vec{q}jk\alpha} e^{i\left(\vec{q}\cdot\vec{R}_l + \omega_{\vec{q}j}t\right)}$ (Srivastava 1990), substituting in Eqs. 3 and 4, and invoking translational invariance to eliminate the sum over unit cells, l gives an eigenvalue equation that is solved for $\omega_{\vec{q}j}$ and the eigenvectors $\varepsilon_{\vec{q}jk\alpha}$. In practice, these are determined by diagonalizing the so-called dynamical matrix for each $\vec{q}$:

$$\overline{D} = D_{\vec{q}\alpha\beta}^{kk'} = \frac{1}{\sqrt{m_k m_{k'}}} \sum_{l'} \Phi_{\alpha\beta}^{0kl'k'} e^{i\vec{q}\cdot\vec{R}_{l'}} \tag{5}$$

For polar materials, nonanalytic long-range Coulomb corrections to this dynamical matrix can play an important role in deriving optic mode frequencies, particularly longitudinal/transverse (LO/TO) splittings. An example of this is given in Fig. 2 for KCl where the LO and TO branches are highly separated at the Γ point. These long-range Coulomb corrections can be constructed from Ewald sums involving Born effective charges and high-frequency dielectric tensors from DFT calculations (Baroni et al. 2001). Not correcting for the LO/TO splitting can give artificially reduced scattering of heat-carrying acoustic modes interacting with these phonons and thus overestimation of the thermal conductivity.

Crystal periodicity reduces the calculations of all vibrational properties to the first Brillouin zone of the reciprocal lattice, in momentum space. Unit cells and symmetry are important considerations when taking on a transport calculation. The number of atoms in a typical unit cell can range from $n = 1$ (e.g., aluminum or lead) to n hundreds of atoms. This gives $3n$ phonon branches corresponding to three vibrational degrees of freedom for each unit cell atom and can be a limiting factor depending on computational resources. Figure 2 gives the phonon dispersion, relationship of $\omega_{\vec{q}j}$ and $\vec{q}$, within the first Brillouin zone and along high-symmetry directions for rock salt structure KCl. KCl has two atoms in its unit cell and thus six phonon branches – three acoustic branches with $\omega_{\vec{q}j} \to 0$ as $|\vec{q}| \to 0$ due to translational invariance of the crystal and three optic branches at higher frequencies.

Even at the simple harmonic level, various observables to compare with measurements can be determined: phonon frequencies (neutrons (Dorner 1982; Willis et al. 2009), x-rays (Burkel 2000; Rueff 2010), and Raman (Cardona 1975; Merlin et al. 2000) resonances), vibrational density of states (DOS), the volumetric-specific heat at low temperatures, elastic constants, and sound velocities. The phonon DOS is given by:
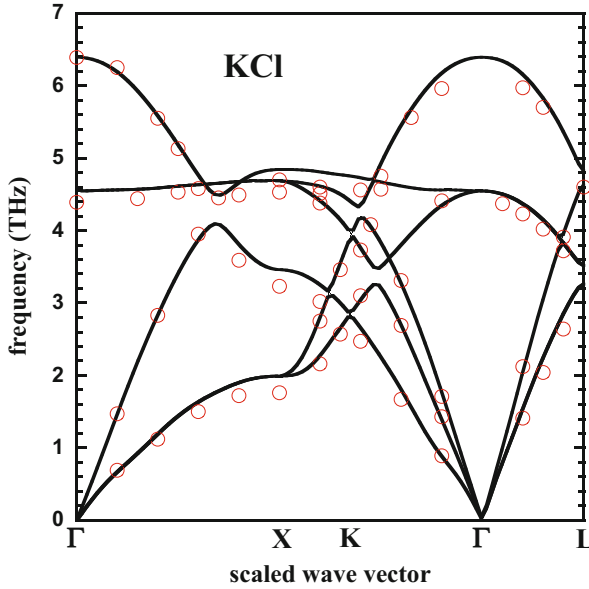
Fig. 2 Phonon dispersion along high-symmetry directions of rock salt structure KCl, black curves are calculations that employed density functional perturbation theory (Baroni et al. 2001) and red circles are from neutron scattering measurements (Copley et al. 1969)

$$\mathrm{DOS} = \sum_k \mathrm{DOS}_k = \sum_{\vec{q}jk} \left| \hat{\varepsilon}_{\vec{q}jk} \right|^2 \delta \left(E - \hbar \omega_{\vec{q}j} \right) = \sum_{\vec{q}jk} \left| \hat{\varepsilon}_{\vec{q}jk} \right|^2 \frac{\sigma}{\pi \left(\sigma^2 + \left(E - \hbar \omega_{\vec{q}j} \right)^2 \right)} \tag{6}$$

where E is the probe energy and the Dirac delta is represented by a Lorentzian distribution here (Byron and Fuller 1992). σ is an adjustable parameter, small enough to give a good representation of the Dirac delta, but large enough to smooth out unphysical variations in the DOS. The total DOS and that projected onto the separate vibrations of the constituent atoms DOS_k are given for KCl in Fig. 3.

Comparison of calculated specific heat with measurements is actually quite tricky. $C_V = \sum_{\vec{q}j} C_{\vec{q}j}$ is an easy calculation constructed from $\omega_{\vec{q}j}$ alone; however, it is an extremely hard property to measure as materials change volume with temperature. Measurements of C_P are more practical (Gaskell 2008):

$$C_P = C_V + \alpha_V^2 B V_{\mathrm{cell}} T \tag{7}$$

where α_V is the volume thermal expansion arising from anharmonicity and defined in the next section, V_{cell} is the equilibrium unit cell volume, and B is the isothermal bulk modulus. $C_V \approx C_P$ only at low temperatures, as shown in Fig. 4 for KCl.
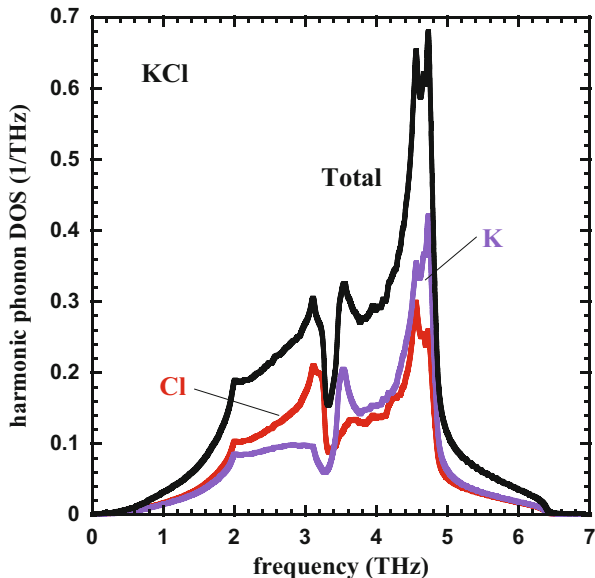
Fig. 3 Projected and total phonon density of states (DOS) for rock salt structure KCl. The red curve gives the vibrational DOS contributions from the Cl atoms, while purple gives that from the K atoms. The K and Cl atoms have similar mass and thus give similar DOS features. When masses differ appreciably, the heavier atoms govern the low-frequency DOS, and the lighter atoms govern the higher-frequency DOS

Thus, more rigorous comparison with measurements at higher temperatures requires including the anharmonic contributions of Eq. 7, as also demonstrated in Fig. 4.

2.3   Phonons: Anharmonic

While non-interacting harmonic phonons can describe (for the most part) the low-temperature-specific heat of crystalline solids and resonances from neutron, x-ray, and Raman scattering experiments, going beyond this simple picture is required to describe and understand lattice thermal expansion and conduction (Debye 1914; Peierls 1929, 2001), linewidths, and lineshifts (Maradudin and Fein 1962; Cowley 1968) – all temperature-dependent behaviors. These require knowledge of the crystal anharmonicity or how phonons interact with one another via perturbations that arise from higher-order terms in the interatomic potential, V_{anharm}.

The lowest-order anharmonic term, and presumably the most relevant within quantum perturbation theory, is V_3 (Srivastava 1990; Ziman 2001):

$$V_3 = \frac{1}{3!} \sum_{lk} \sum_{l'k'} \sum_{l''k''} \sum_{\alpha\beta\gamma} \Phi_{\alpha\beta\gamma}^{lkl'k'l''k''} u_{lk\alpha} u_{l'k'\beta} u_{l''k''\gamma} \tag{8}$$
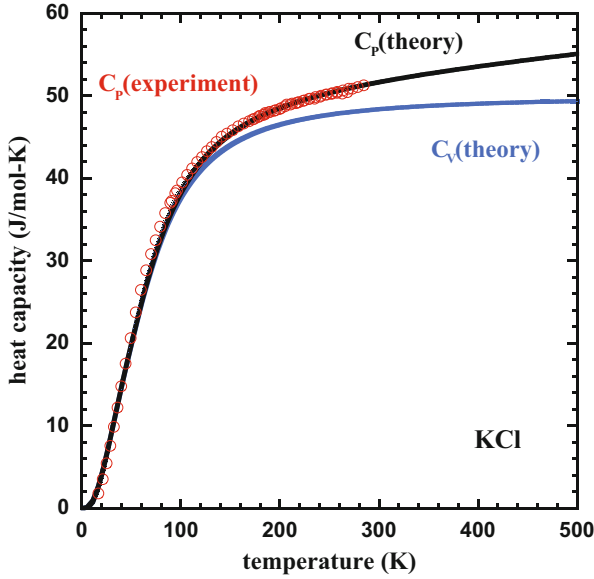
Fig. 4 Heat capacity of rock salt structure KCl. The blue curve gives C_V, while the black curve gives C_P from calculations. The red circles correspond to measured data of C_P (Southard and Nelson 1933). The measured $B = 19.7$ GPa for KCl (Norwood and Briscoe 1958) was used here

which gives rise to three-phonon interactions. $\Phi_{\alpha\beta\gamma}^{lkl'k'l''k''} = \partial^3 V/\partial u_{lk\alpha} \partial u_{l'k'\beta} \partial u_{l''k''\gamma}$ are third derivatives of the interatomic potential calculated at equilibrium, calculation of which is described in Sect. 3.1. From these, various anharmonic material properties can be constructed including mode Grüniesen parameters, which describe the frequency dependence with volume of each phonon mode (Fabian and Allen 1997):

$$\gamma_{\vec{q}j} = -\frac{V_0}{\omega_{\vec{q}j}} \frac{d\omega_{\vec{q}j}}{dV} = -\frac{1}{6\omega_{\vec{q}j}^2} \sum_k \sum_{l'k'} \sum_{l''k''} \sum_{\alpha\beta\gamma} \Phi_{\alpha\beta\gamma}^{0kl'k'l''k''} \frac{\varepsilon_{\vec{q}jk\alpha}^* \varepsilon_{\vec{q}jk'\beta}}{\sqrt{m_k m_{k'}}} r_{l''k''\gamma} \quad (9)$$

and the linear thermal expansion coefficient α (Fabian and Allen 1997; Ziman 2001):

$$\alpha = \frac{1}{3B} \sum_{\vec{q}j} C_{\vec{q}j} \gamma_{\vec{q}j} \quad (10)$$

$r_{lk\alpha}$ is the αth coordinate of the kth atom in the lth unit cell. For cubic systems like rock salt structure KCl $\alpha_V = 3\alpha$, which was used in the calculation of C_P for KCl in Fig. 4. Consideration of lattice thermal expansion is critical for effective device design as unlike materials expand and contract with temperature at different
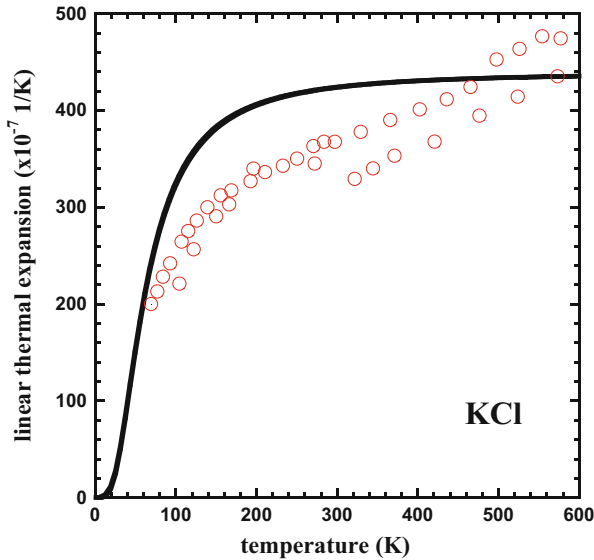
Fig. 5 Linear thermal expansion coefficient α for KCl: black curve from DFT calculations and Eqs. 8, 9, and 10; red circles sampled from measured data in (Rapp and Merchant 1973)

rates leading to structural failure at their interface. Furthermore, it is important for understanding the fundamentals of optic phonon frequency shifts and other lattice dynamical nuances (Cardona 1975; Burke and Herman 1993). Figure 5 gives α for KCl as calculated from Eqs. 8, 9, and 10 and compared with measured data, also giving an example of the occasional failure of first-principles methods to accurately describe a measured data set.

At the core of this chapter is the effect of anharmonicity on lattice thermal conduction, as it gives rise to three-phonon interactions that govern the phonon lifetimes/linewidths and account for the dominant intrinsic thermal resistance in many materials.

2.4 Phonon Lifetimes: Relaxation Time Approximation (RTA)

Calculation of the phonon lifetimes $\tau_{\vec{q}j}$ in Eq. 2 is the core challenge to realistic thermal transport predictions. A phonon lifetime corresponds to the average time for a phonon to exist before scattering and is often related to its mean free path given by $\ell_{\vec{q}j} = |\vec{v}_{\vec{q}j}| \tau_{\vec{q}j}$. It may vary considerably among modes and materials and can be quite sensitive to extrinsic crystal defects –the motivation behind nanostructuring in thermoelectrics (Singh and Terasaki 2008; Kanatzidis 2010). The main focus of discussion in this subsection is the intrinsic resistance due to three-phonon interactions; ab initio defect resistance calculations will be discussed in Sect. 5.

The phonon lifetimes will not be fully derived here; certain aspects will be discussed in detail; others will simply be mentioned. Within the relaxation time approximation (RTA – discussed in more detail below), the intrinsic phonon lifetimes are given as the sum of all individual scattering probabilities $W = (2\pi/\hbar)|\langle f| V_3 |i\rangle|^2 \delta(E_f - E_i)$ (Shankar 1994) determined from Fermi's golden rule with the lowest-order anharmonic potential as the perturbation (Srivastava 1990; Ziman and John 2001):

$$\frac{1}{\tau_{\vec{q}j}} = \sum_{\vec{q}'j'} \sum_{\vec{q}''j''} \left(W^{\text{coalesce}}_{\vec{q}j\vec{q}'j'\vec{q}''j''} + \frac{1}{2} W^{\text{decay}}_{\vec{q}j\vec{q}'j'\vec{q}''j''} \right) \tag{11}$$

$$W^{\text{coalesce}}_{\vec{q}j\vec{q}'j'\vec{q}''j''} = \frac{\pi\hbar}{4N} \frac{\left| \Psi_{-\vec{q}j-\vec{q}'j'\vec{q}''j''} \right|^2}{\omega_{\vec{q}j}\omega_{\vec{q}'j'}\omega_{\vec{q}''j''}} \left(n^0_{\vec{q}'j'} - n^0_{\vec{q}''j''} \right) \delta \left(\omega_{\vec{q}j} + \omega_{\vec{q}'j'} - \omega_{\vec{q}''j''} \right)$$
$$\Delta \left(\vec{q} + \vec{q}' - \left(\vec{q}'' + \vec{G} \right) \right) \tag{12}$$

$$W^{\text{decay}}_{\vec{q}j\vec{q}'j'\vec{q}''j''} = \frac{\pi\hbar}{4N} \frac{\left| \Psi_{-\vec{q}j\vec{q}'j'\vec{q}''j''} \right|^2}{\omega_{\vec{q}j}\omega_{\vec{q}'j'}\omega_{\vec{q}''j''}} \left(n^0_{\vec{q}'j'} + n^0_{\vec{q}''j''} + 1 \right) \delta \left(\omega_{\vec{q}j} - \omega_{\vec{q}'j'} - \omega_{\vec{q}''j''} \right)$$
$$\Delta \left(\vec{q} - \vec{q}' - \left(\vec{q}'' + \vec{G} \right) \right) \tag{13}$$

$$\Psi_{\vec{q}j\vec{q}'j'\vec{q}''j''} = \sum_{k} \sum_{l'k'} \sum_{l''k''} \sum_{\alpha\beta\gamma} \frac{\Phi^{0kl'k'l''k''}_{\alpha\beta\gamma}}{\sqrt{m_k m_{k'} m_{k''}}} \varepsilon_{\vec{q}jk\alpha} \varepsilon_{\vec{q}'j'k'\beta} \varepsilon_{\vec{q}''j''k''\gamma} e^{i\vec{q}'\cdot\vec{R}_{l'}} e^{i\vec{q}''\cdot\vec{R}_{l''}} \tag{14}$$

for three phonons that conserve energy and momentum (within a reciprocal lattice vector $\vec{G}$):

$$\omega_{\vec{q}j} \pm \omega_{\vec{q}'j'} = \omega_{\vec{q}''j''} \qquad \vec{q} \pm \vec{q}' = \vec{q}'' + \vec{G} \tag{15}$$

The $\pm$ corresponds to coalescence and decay processes, respectively, as depicted in Fig. 6.

The form and terms of Eqs. 11, 12, 13, and 14 derive from the quantum mechanical description of the displacements of the harmonic oscillators $u_{lk\alpha} = \sqrt{\hbar/2m_k N} \sum_{\vec{q}j} \varepsilon_{\vec{q}jk\alpha} e^{i\vec{q}\cdot\vec{R}_l} \left(\hat{a}_{\vec{q}j} + \hat{a}^*_{-\vec{q}j} \right) / \sqrt{\omega_{\vec{q}j}}$ (Ashcroft and Mermin 1976), which build Eq. 8 and act on the phonon states via Fermi's golden rule. N is the number of unit cells in the crystal; $\hat{a}_{\vec{q}j}$ and $\hat{a}^*_{\vec{q}j}$ are phonon creation and
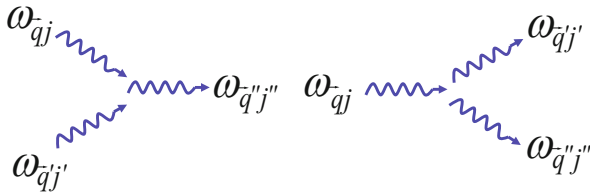
Fig. 6 Three-phonon processes allowed by energy conservation (three phonons cannot be simultaneously created nor destroyed). The left depicts a coalescence process; the right depicts a decay process

annihilation operators, respectively, such that $\hat{a}_{\vec{q}j} \left| n_{\vec{q}j} \right\rangle = \sqrt{n_{\vec{q}j} + 1} \left| n_{\vec{q}j} + 1 \right\rangle$ and $\hat{a}^*_{\vec{q}j} \left| n_{\vec{q}j} \right\rangle = \sqrt{n_{\vec{q}j}} \left| n_{\vec{q}j} - 1 \right\rangle$ (Ashcroft and Mermin 1976; Shankar 1994).

Scattering processes with $\vec{G} = 0$ are called normal scatterings as they conserve momentum; however, they cannot provide direct resistance to a flowing phonon distribution though treated as resistive within the RTA. Processes with $\vec{G} \neq 0$ are called Umklapp ("folding over") scatterings and are directly resistive (Ziman 2001; Peierls 2001). The main simplification presented by the RTA is that $1/\tau_{\vec{q}j}$ is independent of the distribution of the other modes. That is, the deviation from equilibrium induced by the temperature gradient of mode $\vec{q}j$ (from which $1/\tau_{\vec{q}j}$ is derived) does not depend on the distributions of the other phonon modes with which it interacts. Thus, equilibrium Bose distributions are given for phonons $\vec{q}'j'$ and $\vec{q}''j''$ in Eqs. 12 and 13 rather than the actual non-equilibrium distributions that arise in the presence of a temperature gradient. For many materials (particularly low κ materials), this approximation is valid (Lindsay et al. 2013a); however, for some cases (particularly high κ materials), this gives a bad representation of phonon interactions and transport (Lindsay et al. 2009, 2014; Cepellotti et al. 2015).

Figure 7 gives the calculated phonon scattering rates for KCl at $T = 300$ K (room temperature): normal, Umklapp, and phonon-isotope scatterings (Sect. 5). In this system at room temperature, the intrinsic scattering is dominant; however, as temperature decreases, the three-phonon scattering weakens via the Bose distribution factors in Eqs. 12 and 13, while phonon-isotope scattering (Eq. 27) is T independent.

Briefly summarizing, calculation of $\tau_{\vec{q}j}$ (and thus κ in Eq. 2) within the RTA requires a description of the phonons built from the harmonic IFCs, a description of the anharmonic IFCs (Sect. 3.1), and integrations over delta functions (Sect. 3.2). Up to this point, the Peierls-Boltzmann equation (PBE) (Boltzmann 1872; Peierls 1929, 2001) for describing phonon transport has been alluded, but not rigorously introduced or connected, to these lifetimes.

2.5 Peierls-Boltzmann Equation (PBE)

The Peierls-Boltzmann equation in its typical form (Ziman 2001; Peierls 2001):
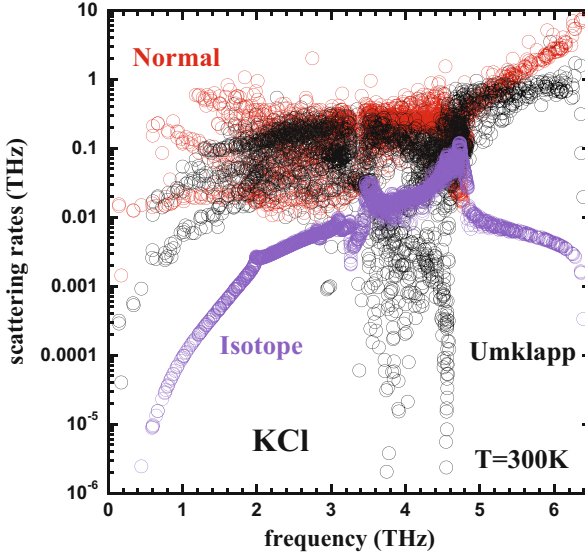
Fig. 7 Phonon scattering rates of KCl as determined by Eqs. 11, 12, 13, and 14 (intrinsic anharmonic rates) and Eq. 27 in Sect. 5 (phonon-isotope scattering from natural isotopic variation, purple circles). The intrinsic scattering is separated into normal scatterings (red circles) and Umklapp scatterings (black circles)

$$\vec{v}_{\vec{q}j} \cdot \vec{\nabla}T \frac{\partial n_{\vec{q}j}}{\partial T} = \left. \frac{\partial n_{\vec{q}j}}{\partial t} \right|_{scatterings} \tag{16}$$

describes the dynamics of a homogenous (no spatial variation beyond that induced by a small temperature gradient), steady-state (no time variation) distribution of phonons flowing in a material driven out of equilibrium by a temperature gradient – a temperature-driven distribution balanced by scattering. The solution of Eq. 16 gives the non-equilibrium distribution $n_{\vec{q}j}$ of each phonon mode in the presence of a temperature gradient and dependent on the non-equilibrium distribution of all the other modes. $n_{\vec{q}j}$ is used to determine each phonon lifetime as discussed later. The "scattering" term on the right-hand side of Eq. 16 is built from scattering processes of a similar form to Eqs. 12 and 13; however, the distributions are considered as unknowns to solve for rather than the $n_{\vec{q}j}^0$ terms presented there.

A significant assumption when solving the PBE is that the non-equilibrium distribution is the Bose distribution with a small correction linear in the temperature gradient (linearization of the PBE): $n_{\vec{q}j} = n_{\vec{q}j}^0 + n_{\vec{q}j}^1$ with $n_{\vec{q}j}^1 = \vec{\nabla}T \Omega_{\vec{q}j} \partial n_{\vec{q}j}^0 / \partial T$ where $\Omega_{\vec{q}j}$ is an unknown measure for how far each mode is perturbed from equilibrium (Srivastava 1990; Ziman 2001). (A helpful note: $\partial n_{\vec{q}j}^0 / \partial T = n_{\vec{q}j}^0 \left(n_{\vec{q}j}^0 + 1 \right) \hbar \omega_{\vec{q}j} / k_B T^2$.) Inserting these non-equilibrium distri-

butions in the scattering rates that construct Eq. 16, accounting for like terms and indistinguishable transitions, enforcing the collisional term to be zero for the equilibrium distribution, choosing an orientation for the temperature gradient along α, and keeping only linear terms in it, the PBE becomes (Broido et al. 2007):

$$
\begin{aligned}
\frac{v_{\vec{q}j\alpha}k_B T}{n_{\vec{q}j}^0 \left(n_{\vec{q}j}^0 + 1\right)} \frac{\partial n_{\vec{q}j}^0}{\partial T} = & \sum_{\vec{q}'j'} \sum_{\vec{q}''j''} \Bigg[W_{\vec{q}j\vec{q}'j'\vec{q}''j''}^{\text{coalesce}} \left(\Omega_{\vec{q}j} + \Omega_{\vec{q}'j'} - \Omega_{\vec{q}''j''}\right) \\
& + \frac{1}{2} W_{\vec{q}j\vec{q}'j'\vec{q}''j''}^{\text{decay}} \left(\Omega_{\vec{q}j} - \Omega_{\vec{q}'j'} - \Omega_{\vec{q}''j''}\right) \Bigg] \\
& + \frac{\Omega_{\vec{q}j}}{\tau_{\text{extrinsic}}}
\end{aligned}
\tag{17}
$$

This linearized form is solved to determine the deviations from equilibrium $\Omega_{\vec{q}j}$ for each phonon. Note that an extra scattering term is included in Eq. 17 for extrinsic phonon scattering mechanisms (isotopes, defects, boundaries, etc.) that can be incorporated at the RTA level, though not restricted to this approximation.

There are different ways to construct a solution of the coupled set of integral equations, including self-consistent iterative methods (Omini and Sparavigna 1996) and variational approaches (Ziman 2001; Fugallo et al. 2013). Given the solution of the PBE for $\Omega_{\vec{q}j}$, the lifetimes are:

$$
\tau_{\vec{q}j\alpha} = \frac{\Omega_{\vec{q}j}T}{\hbar\omega_{\vec{q}j}v_{\vec{q}j\alpha}}
\tag{18}
$$

In this way the calculated lifetimes from full solution of the PBE depend on the orientation of the temperature gradient (technically the deviation terms may also depend on the gradient $\Omega_{\vec{q}j} \rightarrow \Omega_{\vec{q}j\alpha}$). The lifetimes and deviation terms are related by Eq. 18 so as to reconstruct Eq. 2. Going back, the thermal energy current from Eq. 1 can also be written as the product of the energy carried by each mode, its speed, and the number of modes: $J_\alpha = \sum_{\vec{q}j} \hbar\omega_{\vec{q}j}v_{\vec{q}j\alpha}n_{\vec{q}j}/V_0$ (Ashcroft and Mermin 1976). Putting the linear form for the distribution in this and equating it with Eq. 1 gives Eq. 18.

In summary, given a set of harmonic and anharmonic IFCs, κ can be determined from a linearized form of the PBE with no adjustable fitting parameters for the intrinsic resistance. Practically speaking, wavevector grids to perform sums over the phonons must be defined within the first Brillouin zone, harmonic properties ($\omega_{\vec{q}j}$, $C_{\vec{q}j}$, $\vec{v}_{\vec{q}j}$, $n_{\vec{q}j}^0$) calculated via Eq. 5 for each grid point, numerical representations of the Dirac delta functions in Eqs. 12 and 13 constructed, and sums performed over these (phase space for scattering); and, if going beyond the RTA, these are plugged into Eq. 17 and solved for lifetimes to determine κ. Important points for the numerical implementation of these calculations will be discussed in the following section.

3 Numerical Recipes

Numerical implementation of the above theoretical discussion requires some careful thought and creativity, particularly regarding efficient calculation of interatomic forces, determination of all possible scattering configurations, and integrations over delta functions. Several software packages (Li et al. 2014; Tadano et al. 2014; Chernatynskiy and Phillpot 2015; Togo et al. 2015) have recently become available to calculate κ coupled with IFCs from density functional theory (Hohenberg and Kohn 1964; Kohn and Sham 1965), and each has different methods for addressing numerical details, some of which are discussed here.

3.1 Interatomic Force Constants (IFCs)

The only input parameters into the theoretical PBE transport formalism above are the harmonic and anharmonic IFCs. Thus, an accurate physical description of these is critically important for obtaining reasonable κ predictions. Unfortunately, this is also one of the numerical bottlenecks of these calculations, particularly when using DFT methods to extract the IFCs. Likely the most popular method for calculating them is numerical differentiation in real space of an interatomic potential defined by the electronic structure of the crystal (Srivastava 1990). In practice, a large supercell is defined from the basis unit cell using $\vec{R}_l$, specific atoms are strategically displaced from equilibrium by a small distance Δ, and calculations of the electronic structure, energies, and forces on all the atoms are performed. Δ should be significantly less than the interatomic spacing, but not so small that energies and forces cannot be calculated to desired precision. Furthermore, supercells must be large enough so that periodic images do not interact strongly.

Discussion here describes the calculation of the third-order anharmonic IFCs $\Phi_{\alpha\beta\gamma}^{0kl'k'l''k''}$; however, much of it can be applied to the determination of harmonic and even higher-order IFCs. Typical electronic structure calculations can give the forces acting on the atoms (Hellmann 1937; Feynman 1939; Baroni et al. 2001) relatively cheaply so $\Phi_{\alpha\beta\gamma}^{0kl'k'l''k''}$ is rewritten as:

$$\Phi_{\alpha\beta\gamma}^{0kl'k'l''k''} = -\frac{\partial^2 F_{l''k''\gamma}}{\partial u_{0k\alpha}\partial u_{l'k'\beta}} \approx -\frac{F_{l''k''\gamma}^{+,+} - F_{l''k''\gamma}^{+,-} - F_{l''k''\gamma}^{-,+} + F_{l''k''\gamma}^{-,-}}{4\Delta^2} \qquad (19)$$

where $F_{l''k''\gamma}$ is the force acting in the γth direction on the k''th atom in the l''th unit cell, $F_{l''k''\gamma}^{+,+}$ is the same but with $0k$ and $l'k'$ atoms perturbed a distance Δ in the positive $(+)$ α and positive β directions, respectively. In principle, there are an infinite number of IFCs each requiring four (one for each force in the numerator of Eq. 19) electronic structure calculations with large supercells, numerically intractable. Fortunately, IFCs become weaker with increasing distance of the atoms so an interaction cutoff distance is chosen (a spatial distance or number of nearest neighbor shells of the unit cell atoms) which truncates the number of anharmonic

IFCs. This cutoff should be chosen wisely to give a balance of computational efficiency and converged accuracy. From this truncated set, symmetry operations are employed to reduce the IFCs to a smaller set of independent/irreducible IFCs from which the entire truncated set can be assembled (Esfarjani and Stokes 2008). Then the minimum number of perturbed supercells required to assemble this irreducible set are constructed and forces calculated. As each calculation gives information of the forces acting on all the atoms in the supercell, many irreducible IFCs can be constructed from the same calculations, and there is a significant amount of redundant information – multiple calculations of the same irreducible IFC. This information can be used to check the numerical accuracy of calculations and can be averaged.

Physically meaningful IFCs not only satisfy derivative permutation and space group symmetry operations of the crystal lattice (which are guaranteed by operating with the irreducible IFCs) but also must satisfy various translational and rotational invariance constraints (Leibfried and Ludwig 1961; Gazis and Wallis 1966; Born and Huang 1998). Due to truncation of the anharmonic IFCs to a particular neighbor shell and numerical accuracy, translational invariance constraints given by the acoustic sum rule (ASR) (Leibfried and Ludwig 1961):

$$\sum_{l'k'} \Phi_{\alpha\beta\gamma}^{0kl'k'l''k''} = 0 = \sum_{l''k''} \Phi_{\alpha\beta\gamma}^{0kl'k'l''k''} \tag{20}$$

are not satisfied by IFCs directly from DFT supercell calculations. It is important to enforce these constraints by applying small changes to the irreducible DFT IFCs to obtain physically meaningful ones (Carrete et al. 2016); otherwise scattering rates from Eq. 11 likely will not have the proper low-frequency behavior, and calculated κ may be significantly underestimated, as demonstrated in Fig. 8 (Ge figure) adapted from Lindsay et al. (Lindsay et al. 2013a). Various numerical methods have been devised to solve optimization problems similar to enforcing symmetry constraints (Press et al. 1992; Esfarjani and Stokes 2008; Lindsay et al. 2013a; Mingo et al. 2014). Enforcing symmetry constraints beyond simple ASR can also be important for obtaining quadratic dispersion of flexural acoustic branches in 2D systems (Carrete et al. 2016).

Often DFT calculations are considered free of input parameters; however, there are significant choices which need consideration when performing lattice dynamical calculations: basis sets, representations of core and valence electrons, exchange and correlations, etc. (Giannozzi et al. 2009). Typical phonon transport calculations employ the local density (Kohn and Sham 1965) or generalized gradient approximations (Perdew et al. 1996) (LDA or GGA) using projector augmented wave (Blöchl 1994) or plane wave basis sets, with pseudopotentials to represent core electrons (Bachelet et al. 1982; Troullier and Martins 1991). An important point to note here is that LDA tends to overbind atoms (gives smaller structural lattice constants), while GGA tends to underbind atoms (Haas et al. 2009). Thus, LDA calculations sometimes overestimate optic phonon frequencies and acoustic velocities, while GGA calculations underestimate them. This can play a significant
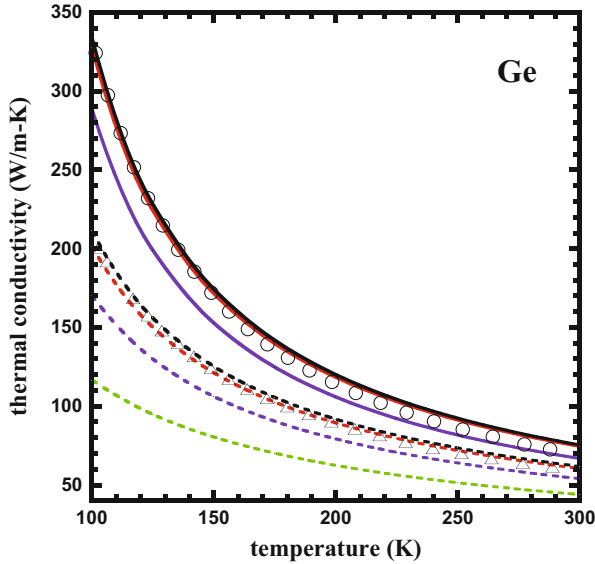
Fig. 8 Calculated thermal conductivity of Ge versus temperature adapted from Lindsay et al. (Lindsay et al. 2013a). Solid curves (calculated) and black circles (measured (Ozhogin et al. 1996)) for isotopically purified Ge, dashed curves (calculated) and black triangles (measured (Ozhogin et al. 1996)) for Ge with natural isotope variations. The black and red curves are derived from calculations that employed different techniques for applying translational invariance, point group, and derivative permutation symmetries simultaneously. The purple curves give calculations that employed a simple ASR, breaking other symmetry properties. The green curve gives calculations for which no translational invariance was enforced. (Adapted figure with permission from Lindsay L et al., PRB 87, 165,201 (2013) Copyright (2013) by the American Physical Society (Lindsay et al. 2013a))

role in κ calculations, especially if the phase space for phonon scatterings of heat-carrying acoustic phonons is not well-represented (Ziman 2001; Lindsay et al. 2012). If measured phonon frequencies are available from neutron, x-ray, or Raman spectroscopies, it is a good idea to make comparisons before undertaking full transport calculations. Another related consideration is that typical DFT calculations do not incorporate the effects of temperature, particularly thermal expansion and thermally induced disorder. Such features can be partially accounted for within quasiharmonic approximations (Srivastava 1990), and some work has presented more rigorous calculations of temperature effects related to vibrational behaviors and κ (Hellman and Broido 2014; Romero et al. 2015).

3.2 Integrations, Delta Functions, and Phase Space

Integrations over the various delta functions that arise from fundamental conservation conditions require careful consideration and give rise to significant variety

in the mechanics of different software packages for calculating κ. In terms of the scatterings in Eqs. 12 and 13, the challenge is to find all scattering processes that conserve (or come close to) energy and crystal momentum (Eq. 15). Examining more closely the scattering rates from Eq. 12 for normal coalescence processes, the Kronecker delta Δ eliminates the sum over $\vec{q}''$, and the sum over $\vec{q}'$ is converted to an integration over the first Brillouin zone that can make use of the Dirac delta for the energies δ:

$$
\frac{1}{\tau_{\vec{q}j}} = \sum_{j'j''} \frac{V_0}{(2\pi)^3} \int_{BZ} \frac{\pi\hbar}{4N} \frac{\left|\Psi_{-\vec{q}j-\vec{q}'j'(\vec{q}+\vec{q}')j''}\right|^2}{\omega_{\vec{q}j}\omega_{\vec{q}'j'}\omega_{(\vec{q}+\vec{q}')j''}} \left(n^0_{\vec{q}'j'} - n^0_{\vec{q}''j''}\right) d\vec{q}
$$

$$
\delta\left(\omega_{\vec{q}j} + \omega_{\vec{q}'j'} - \omega_{(\vec{q}+\vec{q}')j''}\right)
\tag{21}
$$

If a 2D integration, as with graphene or MoS_2, the density prefactor will be $A/(2\pi)^2$ (A = area of crystal), and if 1D integration as with nanotubes, it is $L/(2\pi)$ (L = length of crystal). In these lower dimensional cases, κ in Eq. 2 is scaled by the thickness chosen for the 2D sheet or nanotube wall, and care should be taken to use the same thickness factor when comparing with experiments or other calculations. The Dirac delta in Eq. 21 can be represented in a variety of forms, one being (Jackson 1999):

$$
\delta\left(\omega_{\vec{q}j} + \omega_{\vec{q}'j'} - \omega_{(\vec{q}+\vec{q}')j''}\right) = \sum_i \frac{\delta\left(q'_\alpha - q'_{\alpha i}\right)}{\left|v_{\vec{q}'j'\alpha} - v_{(\vec{q}+\vec{q}')j''\alpha}\right|_{q'_\alpha = q'_{\alpha i}}}
\tag{22}
$$

where the Dirac delta for energy conservation has been converted into a Dirac delta to eliminate one of the wavevector integrations. In this case, an algorithm must be employed to find all $q'_{\alpha i}$ for each set of $jj'j''\vec{q}q'_\beta q'_\gamma$ that satisfy energy conservation. Another popular form is an adaptive Gaussian smearing method (Yates et al. 2007; Li et al. 2014):

$$
\delta\left(\omega_{\vec{q}j} + \omega_{\vec{q}'j'} - \omega_{(\vec{q}+\vec{q}')j''}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\omega_{\vec{q}j} - \omega_{\vec{q}'j'} + \omega_{(\vec{q}+\vec{q}')j''}\right)^2/2\sigma^2}
\tag{23}
$$

where the smearing term $\sigma = a\left|\vec{v}_{\vec{q}'j'} - \vec{v}_{(\vec{q}+\vec{q}')j''}\right|\left|\Delta q'\right|$ is a function of the reciprocal grid spacing $\left|\Delta q'\right|$ and a is a parameter ~1 that can be adjusted to speed up convergence with increasing integration grid densities. The smaller σ is chosen, the better Eq. 23 represents a Dirac delta; however, if too small, no combination of grid points on a finite grid will provide contributions to the integral in Eq. 21. This method differs from Eq. 22 in two important ways: (1) all wavevector integrations must be made in Eq. 21 using the smearing method, while Eq. 22 reduces these by one (though with the addition of another sum); and (2) for Eq. 23, energy conservation never strictly occurs, only within a smearing parameter, while in Eq. 22 three-phonon processes are determined very near conservation conditions, within a predetermined accuracy. Another method for Brillouin zone integrations over

energy conserving delta functions is the tetrahedron method whose description can be found elsewhere (Lehmann and Taut 1972; Togo et al. 2015).

Inevitably, for numerical calculations of scattering rates and κ, sums and continuous integrations must be represented on finite grids, and clever choice of these may result in less computational effort for desired convergence accuracy. Monkhorst-Pack meshes (Monkhorst and Pack 1976) have the advantage of naturally coordinating with reciprocal lattice vectors of the crystal and being evenly spaced in reciprocal space. Thus, if $\vec{q}$ and $\vec{q}'$ are defined on the same mesh, then all $\vec{q}''$ determined from crystal momentum conservation also land on this mesh, and less frequency diagonalizations (Eq. 5) need to be made. Furthermore, if using Eq. 23 to represent the Dirac delta, then $\vec{q}$, $\vec{q}'$, and $\vec{q}''$ are all on the same mesh, and interpolation schemes to determine $\Omega_{\vec{q}'j'}$ and $\Omega_{\vec{q}''j''}$ for off-grid wavevectors are not necessary. However, such evenly spaced integration methods do not exploit fancier integration techniques (e.g., Gaussian quadratures (Press et al. 1992)) and do not rigorously enforce energy conservation.

The strength of each three-phonon scattering process (Eqs. 12 and 13) is dictated by a complicated combination of eigenvectors, frequencies, and anharmonic IFCs, varying for each mode involved. Another factor governing the overall scattering rate for each mode (Eq. 11) is the total number of processes allowed by the conservation conditions of Eq. 15 – scattering phase space (Ziman 2001). In fact, this phase space has been shown to be a leading indicator of a materials ability to conduct heat (Broido and Lindsay 2008; Lindsay 2016). The more scattering available in a system, the lower will be its κ. Thus, estimates of the available phase space for scattering provide important insights into the behaviors of lattice κ among various materials, without having to explicitly calculate anharmonic IFCs. Features of the phonon dispersions have been correlated with this phase space including bandgap between acoustic and optic phonon branches (Lindsay et al. 2012), relative closeness of the acoustic branches (Lindsay et al. 2013b), and bandwidth of the optic branches (Lee et al. 2014; Mukhopadhyay et al. 2016). Furthermore, it has been argued that in materials where this phase space is small, and thus three-phonon scattering weak, the effects of higher-order scattering may be more important as higher-order conservation conditions are less restrictive (Ziman 2001; Broido and Lindsay 2008; Feng and Ruan 2016; Feng et al. 2017).

4   Other Observables

The interplay of experiment and theory is critical to building basic insights and fundamental principles into physical phenomena, and yet comparison of measurement and calculations is not always straightforward in practice. Here we discuss the relationship of other measurement probes (Raman, neutrons, and x-rays) and first-principles calculations of lattice vibrations and anharmonicity beyond integrated conductivity values, getting at microscopic mode-dependent information.

A variety of interesting new ultrafast spectroscopic techniques have been developed recently toward gaining insights into thermal transport in materials. Raman thermometry is being employed to access transport in 2D materials (Balandin et al. 2008; Cai et al. 2010; Beechem et al. 2015), though with fairly large error bars due to determination of power absorption of different samples. Also, transient thermal grating (Johnson et al. 2013), time-domain (Cahill et al. 2014), and frequency-domain thermoreflectance (Schmidt et al. 2009) (TDTR and FDTR) techniques are being pioneered to examine κ in various bulk materials, some with capabilities to examine κ as a function of the mean free path of the heat carriers. To this end a discussion of thermal conductivity accumulation κ_L^{acc} with mean free path $\ell_{\vec{q}j} = |\vec{v}_{\vec{q}j}| \tau_{\vec{q}j}$ is given here.

κ_L^{acc} is defined as the contribution to the total κ from modes with $\ell_{\vec{q}j}$ less than some characteristic length L: $\kappa_L^{acc} = \sum_{\vec{q}j} \kappa_{\vec{q}j} \theta\left(L - \ell_{\vec{q}j}\right)$ where θ is the Heaviside step function. Note that the directional dependence of the conductivity terms are ignored here. L may be related to the film thickness, laser spot size, or grating period depending on the experiment. κ_L^{acc} provides information for how κ reduction might be achieved via micro- or nanostructuring in thermoelectric or barrier coating materials (Singh and Terasaki 2008; Kanatzidis 2010) and for quality single crystal sizes needed to achieve highest κ in materials for thermal management applications (Broido et al. 2013). Integration choices can play an important role in these calculations as low-frequency, long wavelength phonons with large $\ell_{\vec{q}j}$ are typically not sampled as well as smaller wavelength modes with smaller $\ell_{\vec{q}j}$, leading to unphysical "jumps" in the κ_L^{acc} (Minnich et al. 2011; Jiang et al. 2016) as seen in Fig. 9 (red symbols). In nearly isotropic materials, like prototypical zincblende compounds, integration grids based on polar or spherical coordinates provide significantly better sampling of the wavevector/frequency space and give more physically meaningful κ_L^{acc} curves as also demonstrated in Fig. 9 (blue symbols).

Inelastic x-ray and neutron scattering experiments provide key insights into phonon properties (particularly dispersions and lifetimes) throughout the Brillouin zone (Dorner 1982; Burkel 2000; Willis et al. 2009; Rueff 2010). Raman spectroscopy, though confined to optic modes near the Γ point, also provides important information of phonon behaviors, as well as insights into couplings with other material properties (Cardona 1975; Merlin et al. 2000). These techniques have guided our fundamental knowledge of materials properties and still bring to light novel vibrational, electronic, and magnetic behaviors. Figure 2 gives a comparison of the phonon dispersion of KCl as measured by inelastic neutron scattering and DFT calculations. Measured linewidths $\Gamma_{\vec{q}j}$ and lineshifts $\Delta_{\vec{q}j}$ as a function of temperature can also provide key benchmarking data for anharmonic calculations (Debernardi et al. 1995; Deinzer et al. 2003; Bonini et al. 2007). Comparisons of lifetimes and linewidths are often made with the simple expression $1/\tau_{\vec{q}j} = \Gamma_{\vec{q}j}/\hbar$ (Di Bartolo 1968; Kuball et al. 2000) where $\Gamma_{\vec{q}j}$ is determined from fitting Lorentzian or Cauchy distributions to measured data (Byron and Fuller 1992):
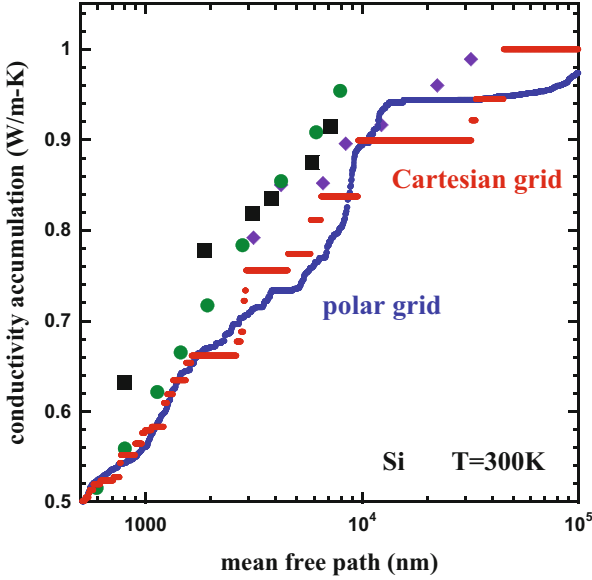
Fig. 9 Thermal conductivity accumulation with mean free path for Si at room temperature given by calculations using Cartesian integration (red points) and polar integration (blue points). Black squares (Jiang et al. 2016), green circles (Regner et al. 2013), and purple diamonds (Minnich et al. 2011; Ding et al. 2014) give measured data for comparison

$$\delta \left(E - \hbar \omega_{\vec{q} j} \right) = \frac{\Gamma_{\vec{q} j}}{\pi \left[\Gamma_{\vec{q} j}^2 + \left(E - \hbar \omega_{\vec{q} j} \right)^2 \right]} \tag{24}$$

again, E is the probe energy – in the case of neutron scattering, E is the neutron energy transfer. For strongly anharmonic materials, measured signals can deviate significantly from simple Lorentzian behavior, and calculation of lineshapes may be of interest. In this case, the phonon linewidth can be expressed explicitly as a function of the probe energy (Maradudin and Fein 1962; Cowley 1968):

$$\begin{aligned}
\Gamma_{\vec{q} j}(E) k = \frac{\pi \hbar}{16N} \sum_{\vec{q}' j'} \sum_{\vec{q}'' j''} & \frac{\left| \Psi_{\vec{q} j \vec{q}' j' \vec{q}'' j''} \right|^2}{\omega_{\vec{q} j} \omega_{\vec{q}' j'} \omega_{\vec{q}'' j''}} \\
& \times \left[\left(n_{\vec{q}' j'}^0 - n_{\vec{q}'' j''}^0 \right) \Delta \left(\vec{q} + \vec{q}' - \left(\vec{q}'' + \vec{G} \right) \right) \right. \\
& \left\{ \delta \left(\omega_{\vec{q}' j'} - \omega_{\vec{q}'' j''} + E \right) - \delta \left(\omega_{\vec{q}' j'} - \omega_{\vec{q}'' j''} - E \right) \right\} \\
& + \left(n_{\vec{q}' j'}^0 + n_{\vec{q}'' j''}^0 + 1 \right) \Delta \left(\vec{q} - \vec{q}' - \left(\vec{q}'' + \vec{G} \right) \right) \\
& \left. \left\{ \delta \left(\omega_{\vec{q}' j'} + \omega_{\vec{q}'' j''} - E \right) + \delta \left(\omega_{\vec{q}' j'} + \omega_{\vec{q}'' j''} + E \right) \right\} \right]
\end{aligned} \tag{25}$$

Note that this expression requires performing fairly complicated sums (very similar to κ calculations discussed above) for *each* probe energy. Furthermore,

taking into account frequency shifts $\Delta_{\vec{q}j}$ due to anharmonic interactions (Cowley 1968), the mode frequency itself may be a function of the probe energy. This may be of interest for calculating temperature- and pressure-dependent behaviors of Raman spectra (Cardona 1975; Burke and Herman 1993; Kuball et al. 2000).

For some materials where only small crystal sizes may be synthesized and powder samples are more accessible for neutron measurements, it is more convenient to measure the integrated phonon density of states (DOS), rather than obtain mode-specific information. For strongly anharmonic systems where representing the anharmonic lineshapes is important (Mukhopadhyay et al. 2016), this becomes challenging as Eq. 25 must be calculated and summed over every phonon mode in the first Brillouin zone.

5   Phonon-Defect Scattering

Inevitably, real materials are not perfect, infinite crystals; even the purest sample has boundaries and imperfections (point and extended). Many of these defects form during synthesis, when otherwise randomly oriented free atoms bind together to grow the crystal little by little. Defects can also be created within a material via external perturbations or particle irradiation. Phonon-defect scattering intimately depends on the nature of the defect, on the vibrational properties of the crystal and on the spatial distribution of these defects. Generally, this scattering becomes important at lower temperatures where intrinsic phonon-phonon scattering is weak. This section discusses different methods to describe phonon-defect scattering and κ from phenomenological models to recent parameter-free ab initio techniques.

Due to their simplicity, phenomenological models are the most widely used methods to incorporate phonon-defect scattering into the PBE. These models assume the phonon-defect scattering rates $1/\tau_{\vec{q}j}^{d}$ are proportional to some power n of the phonon frequency (Ziman 2001):

$$\frac{1}{\tau_{\vec{q}j}^{d}} = A\omega_{\vec{q}j}^{n} \tag{26}$$

The proportionality constant A is typically an adjustable parameter used to fit measurements, sometimes explicitly including T dependence, and n depends on the nature of the defect and the dimensions of the crystal. For a substitutional atom with mass or binding force variation, $n = 4$ (Klemens 1955); for strain due to dislocations, $n = 1$ (Klemens 1955; Carruthers 1959); for the core of dislocations, $n = 3$ (Klemens 1958); for dynamic dislocations "fluttering," $n = -1$ (Kneezel and Granato 1982); and for grain boundaries, $n = 2$ (Klemens 1955). The simplicity of Eq. 26 arises from several approximations that are compensated by the fitting parameter A. One of them assumes a linear phonon dispersion, which restricts the accuracy of those models to the limit of low temperature and frequency. In that limit,

Eq. 26 yields a well-defined temperature dependence of the thermal conductivity $\kappa \propto T^{3-n}$, with n the power of the frequency in Eq. 26 (Carruthers 1961; Ziman 2001).

At the next level of refinement, phonon-defect scattering is derived from first-order perturbation theory (Shankar 1994), and it is sometimes referred to as the Born approximation. In fact, many of the phenomenological expressions above were originally formulated from Fermi's golden rule in a similar way to the phonon lifetimes due to anharmonic perturbations in Sect. 2.4 (Klemens 1955, 1958; Carruthers 1961; Ratsifaritana and Klemens 1987). Consider a defective crystalline material whose total harmonic interatomic potential energy is described by $V_{\text{harm}} + V_d$, separated into contributions from the pure crystal V_{harm} and the defect V_d. If V_d is small relative to V_{harm}, the phonon-defect lifetimes $\tau_{\vec{q}j}^d$ are given as the sum of all individual scattering probabilities $W = (2\pi/\hbar)|\langle f|V_d|i\rangle|^2 \delta(E_f - E_i)$ (Shankar 1994) similar to Sect. 2.4. Applying this methodology to a system of randomly distributed independent mass defects (Tamura 1983, 1984) gives the standard phonon-isotope scattering included in ab initio calculations of κ:

$$\frac{1}{\tau_{\vec{q}j}^{\text{isotope}}} = \frac{\pi}{2N}\omega_{\vec{q}j}^2 \sum_k g_k \sum_{q'j'} \left|\hat{\varepsilon}_{\vec{q}jk} \cdot \varepsilon_{\vec{q}'j'k}^*\right|^2 \delta\left(\omega_{\vec{q}j} - \omega_{\vec{q}'j'}\right) \tag{27}$$

with $g_k = \sum_i f_{ik}(1 - m_{ik}/\overline{m}_k)^2$, f_{ik} and m_{ik} are the fraction and mass of the ith type of the kth atom, respectively, and $\overline{m}_k$ is the natural isotope average mass of this atom.

The major constraint of perturbative approaches – i.e., a *small* perturbation introduced by the defect in the host material – can be overcome using Green's function methodology. Calculations of phonon-defect scattering from this have demonstrated significant and important differences from those computed using perturbative approaches (Chen et al. 2013; Katcho et al. 2014; Katre et al. 2017; Kundu et al. 2011; Mingo et al. 2010; Protik et al. 2016; Wang et al. 2017, Polanco and Lindsay 2018). When only mass variation is considered, scattering rates from Eq. 27 are similar to those calculated using Green's function methodology at low frequencies, when the perturbation $V_d \propto \omega_{\vec{q}j}^2 (m_{ik} - \overline{m}_k)$ is small (Mingo et al. 2010; Kundu et al. 2011). However, structural and IFC variation around the defect can be more disruptive to phonon thermal transport. Phonon-vacancy scattering rates from Green's function methodologies for diamond, BAs, and SiC are nearly an order of magnitude larger than those from perturbative approaches, even in the low-frequency regime (Katcho et al. 2014; Protik et al. 2016; Katre et al. 2017). Green's function techniques have also been applied to extended defects (e.g., dislocations) (Wang et al. 2017), though computational effort increases significantly when going beyond localized defect structures. Figure 10 benchmarks the parameter-free calculations of κ from coupled phonon-defect scattering from Green's function methods with PBE transport, demonstrating reasonable agreement with measurements.
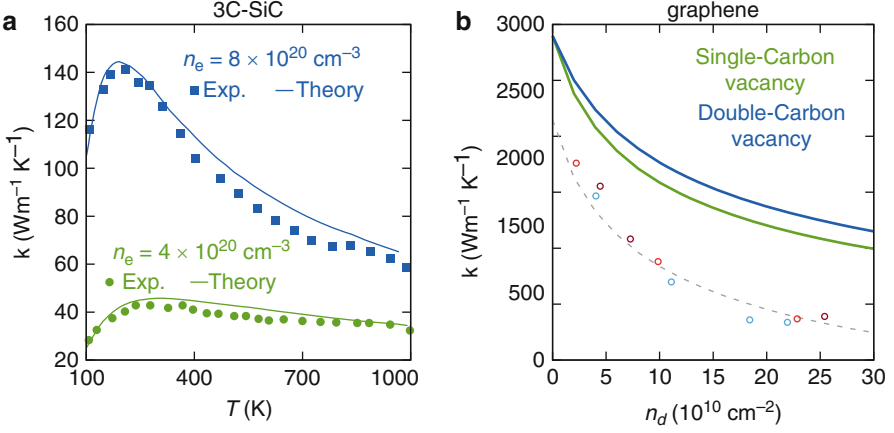
Fig. 10 Calculated κ via Peierls-Boltzmann transport and parameter-free phonon-defect scattering rates from the Green's function methodology. (Left) Calculations and measurements of κ for 3C-SiC doped with N_C substitutional defects (blue squares and blue curve) and codoped with N_C and B_C substitutional defects (green circles and green curve). The electron density n_e defines the defect density for the crystal doped with N_C and the difference between defect densities for the crystal codoped with N_C and B_C. (Adapted figure with permission from Katre A et al., PRL 119, 075902 (2017) Copyright (2017) by the American Physical Society (Katre et al. 2017)). (Right) Calculations (solid curves) and measurements (dots) of κ for graphene with single and double carbon vacancies. The qualitative trend of κ with defect density is in good agreement with measurements. (Reprinted figure with permission from Polanco C and Lindsay L, PRB 97, 014303 (2018) Copyright (2018) by the American Physical Society (Polanco and Lindsay 2018))

Green's function methodology is derived using the Lippmann-Schwinger equation (Lippmann and Schwinger 1950), which describes the exact probability of an initial state scattering into a final state due to a perturbation using an infinite recursive expansion. Considering all final states and applying the optical theorem (Economou 2006), the phonon scattering rates due to an independent distribution of point defects with defect density n_d can be expressed in terms of the scattering cross section $\sigma_{\vec{q}j}^d$ or the so-called T-matrix as (Economou 2006; Mingo et al. 2010):

$$\frac{1}{\tau_{\vec{q}j}^d} = n_d v_{\vec{q}j\alpha} \sigma_{\vec{q}j} = -n_d \frac{V_{\vec{q}j}}{\omega_{\vec{q}j}} \mathrm{Im} \left\{ \langle \vec{q}j | \, \tilde{T}_d \, | \vec{q}j \rangle \right\} \qquad (28)$$

Equation 28 is computed inside a small region with N_d atoms containing a single defect where the dynamical matrix deviates from that of the perfect crystal. This region will be referred as the device. The states $\vec{q}j$ are normalized vectors with $3N_d$ components, one per atomic degree of freedom of the device. The components associated to the unit cell l of the device are given by $\sqrt{1/N_d}\,\hat{\varepsilon}_{\vec{q}j}e^{i\vec{q}\cdot\vec{R}_l}$ assuming that $\hat{\varepsilon}_{\vec{q}j}$ is normalized. $V_{\vec{q}j}$ is the volume of the device, and $\tilde{T}_d$ is the T-matrix in the device, which has dimension $3N_d \times 3N_d$ and is given by a closed term or an equivalent infinite series (Economou 2006; Mingo et al. 2010):

$$\tilde{T}_d = \left[\tilde{I} - \tilde{D}_d \tilde{G}_0 \right]^{-1} \tilde{D}_d = \tilde{D}_d + \tilde{D}_d \tilde{G}_0 \tilde{D}_d + \tilde{D}_d \tilde{G}_0 \tilde{D}_d \tilde{G}_0 \tilde{D}_d + \dots \qquad (29)$$

Note that the tilde is used to represent $3N_d \times 3N_d$ matrices relating degrees of freedom inside the device. $\tilde{D}_d$ represents the contribution to the dynamical matrix from the defect so that the total dynamical matrix is split as $\tilde{D}_0 + \tilde{D}_d$, with $\tilde{D}_0$ the dynamical matrix of the pure crystal. $\tilde{I}$ is the identity matrix, and $\tilde{G}_0$ is the retarded Green function matrix for the perfect crystal inside the device region (Eq. 30).

For efficient computation of $\tilde{G}_0$, all but one Cartesian dimension are transformed to reciprocal space. In this way, the periodicity of the perfect crystal is used to decouple the system into independent "1D chains" that reduce the computational effort. In other words, the 3D crystal can be described as a 2D crystal in a plane, whose base unit cell is an infinite 1D atomic chain. The Green's function for the device of each chain is found using self-energy terms that account for the coupling of atoms inside the device to those outside it. Finally, $\tilde{G}_0$ results from transforming the matrices from reciprocal back to real space.

Transforming the xy plane to Fourier space (with the 1D chains along the z direction) turns the dynamical and Green's function matrices of the perfect crystal into block diagonal matrices labeled by the wavevector $\vec{q}_\perp$, which is defined inside a 2D Brillouin zone in the xy plane. We use the subscript $\perp$ to label these wavevectors because they are perpendicular to the 1D chains. This transformation simplifies the problem of finding Green's functions as each block can be dealt with separately from the others. Each block is associated with a 1D chain in reciprocal space along the z direction decoupled from the others. These chains will be labeled r-chains here to distinguish them from the chains in real space. After finding the device Green's function of each r-chain $\tilde{G}_{\vec{q}_\perp}$, $\tilde{G}_0$ is found using the inverse Fourier transform. The block of $\tilde{G}_0$ relating the degrees of freedom inside the device of the nth chain with those of the mth chain is given by (Zhang et al. 2006; Mingo et al. 2010):

$$\tilde{G}_0 (n, m) = \frac{1}{N_{\vec{q}_\perp}} \sum_{\vec{q}_\perp} \tilde{G}_{\vec{q}_\perp} e^{i \vec{q}_\perp \cdot (\vec{r}_{n\perp} - \vec{r}_{m\perp})} \qquad (30)$$

with $N_{\vec{q}_\perp}$ the number of wavevectors in the Brillouin zone of the 2D crystal and $\vec{r}_{n\perp}$ and $\vec{r}_{m\perp}$ the lattice position of the nth and mth chain, respectively. Each $\tilde{G}_{\vec{q}_\perp}$ is defined as (Zhang et al. 2006; Mingo et al. 2010):

$$\tilde{G}_{\vec{q}_\perp} = \left[\omega_{\vec{q}j}^2 \tilde{I} - \tilde{D}_{\vec{q}_\perp} - \tilde{\Sigma}_{\vec{q}_\perp}^{L} - \tilde{\Sigma}_{\vec{q}_\perp}^{R} \right]^{-1} \qquad (31)$$

All the matrices in Eq. 31 have dimension $3N_{cd} \times 3N_{cd}$, with N_{cd} the number of atoms inside the device region of a 1D chain or r-chain. $\tilde{D}_{\vec{q}_\perp}$ is the dynamical matrix of the r-chain labeled by $\vec{q}_\perp$, which is calculated by transforming the dynamical matrices of the 2D crystal to reciprocal space (Datta 2005; Zhang et al. 2006):

$$\tilde{D}_{\vec{q}\perp} = \sum_m \tilde{D}_{nm} e^{-i\vec{q}\perp \cdot (\vec{r}_{n\perp} - \vec{r}_{m\perp})} \tag{32}$$

$\tilde{D}_{nm}$ is the dynamical matrix relating device atoms of the nth chain with those of the mth chain, and m varies over all neighbor chains with non-zero interaction $\tilde{D}_{nm}$. $\tilde{\Sigma}^L_{\vec{q}\perp}$ and $\tilde{\Sigma}^R_{\vec{q}\perp}$ are the self-energy matrices accounting for the interaction of the device region of the r-chain with the regions to its left and right sides. The self-energy for the left (right) region $\tilde{\Sigma}^L_{\vec{q}\perp}$ $\left(\tilde{\Sigma}^R_{\vec{q}\perp} \right)$ is non-zero only on its top-left (bottom-right) corner, where the degrees of freedom of the device interact with those of the left (right) region. Those non-zero blocks are given by (Datta 2005; Zhang et al. 2007):

$$\text{top-left block of } \tilde{\Sigma}^L_{\vec{q}\perp} = \overline{C}^\dagger_{\vec{q}\perp} \overline{g}^L_{\vec{q}\perp} \overline{C}_{\vec{q}\perp} \tag{33}$$

$$\text{bottom-right block of } \tilde{\Sigma}^R_{\vec{q}\perp} = \overline{C}_{\vec{q}\perp} \overline{g}^R_{\vec{q}\perp} \overline{C}^\dagger_{\vec{q}\perp} \tag{34}$$

Quantities with a bar are matrices, usually with dimension smaller than that of the quantities with tilde. $\overline{g}^L_{\vec{q}\perp}$ and $\overline{g}^R_{\vec{q}\perp}$ are the surface Green's functions of the regions to the left and right of the device in the r-chain and are given by (Datta 2005):

$$\overline{g}^L_{\vec{q}\perp} = \left[\omega^2_{\vec{q}j} \overline{I} - \overline{B}_{\vec{q}\perp} + i\eta - \overline{C}^\dagger_{\vec{q}\perp} \overline{g}^L_{\vec{q}\perp} \overline{C}_{\vec{q}\perp} \right]^{-1} \tag{35}$$

$$\overline{g}^R_{\vec{q}\perp} = \left[\omega^2_{\vec{q}j} \overline{I} - \overline{B}_{\vec{q}\perp} + i\eta - \overline{C}_{\vec{q}\perp} \overline{g}^R_{\vec{q}\perp} \overline{C}^\dagger_{\vec{q}\perp} \right]^{-1} \tag{36}$$

with $i\eta$ an infinitesimal positive imaginary number that guarantees the causality of the Green's function (Datta 2005), hence termed "retarded." $\overline{B}_{\vec{q}\perp}$ and $\overline{C}_{\vec{q}\perp}$ are small matrices coming from the block tridiagonal nature of the dynamical matrix for each r-chain:

$$\overline{D}_{\vec{q}\perp} = \begin{pmatrix} \ddots & 0 & 0 & 0 & 0 \\ 0 & \overline{B}_{\vec{q}\perp} & \overline{C}_{\vec{q}\perp} & 0 & 0 \\ 0 & \overline{C}^\dagger_{\vec{q}\perp} & \overline{B}_{\vec{q}\perp} & \overline{C}_{\vec{q}\perp} & 0 \\ 0 & 0 & \overline{C}^\dagger_{\vec{q}\perp} & \overline{B}_{\vec{q}\perp} & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix} \tag{37}$$

The size of $\overline{B}_{\vec{q}\perp}$ and $\overline{C}_{\vec{q}\perp}$ depends on the longest interatomic interaction considered in the truncated set of harmonic IFCs that describe the perfect crystal. Similar to the discussion in Sect. 3.1, this truncated set should be the smallest possible that reproduces the phonon dispersion closely. This choice guarantees accuracy while keeping the computational load to a minimum. Assuming the size of $\overline{B}_{\vec{q}\perp}$ and $\overline{C}_{\vec{q}\perp}$ is

$s \times s$, the dimension of the surface Green's functions and non-zero blocks of the self-energies is $s \times s$. Note that for real material computations, efficient algorithms based on decimation or renormalization group techniques are used to find the surface Green's functions (Guinea et al. 1983; Rubio et al. 1984, 1985; Galperin et al. 2002).

To briefly summarize, the procedure to calculate the phonon-defect scattering rates from the Green's function methodology is the following: (1) Calculate the IFCs of the device without the defect and construct the dynamical matrix of the perfect device $\tilde{D}_0$ using Eq. 5. (2) Calculate the IFCs and corresponding dynamical matrix of the device with the defect. Then subtract this from $\tilde{D}_0$ to define the perturbation matrix $\tilde{D}_d$. (3) Describe the 3D crystal as a 2D crystal in a plane, whose base unit cell is an infinite 1D atomic chain. Then, using the IFCs for the perfect crystal, define the dynamical matrix $\tilde{D}_{nm}$ relating device atoms of the nth chain with those of the mth chain. (4) Define a set of wavevectors $\vec{q}_\perp$ inside the Brillouin zone of the 2D crystal, and use Eq. 32 to find the dynamical matrix of each r-chain $\tilde{D}_{\vec{q}_\perp}$. (5) Examine $\tilde{D}_{\vec{q}_\perp}$ and compare it with Eq. 37 to define $\overline{B}_{\vec{q}_\perp}$ and $\overline{C}_{\vec{q}_\perp}$. (6) Calculate the surface Green's functions from Eqs. 35 and 36 or using more efficient numerical techniques (Galperin et al. 2002). (7) Find the top-left and bottom-right blocks of the self-energies using Eqs. 33 and 34, and compute $\tilde{G}_{\vec{q}_\perp}$ for each $\vec{q}_\perp$. (8) Finally, build $\tilde{G}_0$ using Eq. 30, and calculate the T-matrix from Eq. 29 to determine the scattering rates using Eq. 28.

6   Summary

In summary, this chapter has presented details of the theoretical underpinnings of the phonon thermal transport capabilities that have recently become available for characterizing the flow of thermal energy in crystalline materials. More specifically, construction of intrinsic and extrinsic phonon scattering rates and other harmonic and anharmonic vibrational properties were discussed in terms of lattice thermal conductivity κ specifically in the context of the Peierls-Boltzmann equation. Discussion of integration methods, construction of interatomic force constants from density functional theory, and various numerical nuances were also presented, and first-principles phonon-defect scattering via Green's function methods was also discussed. Discussions of the calculations of vibrational properties and κ were set in the context of comparison with measured observables.

References

Andersson P (1985) Thermal conductivity under pressure and through phase transitions in solid alkali halides. I. Experimental results for KCl, KBr, KI, RbCl, RbBr and RbI. J Phys C Solid State Phys 18:3943

Ashcroft NW, Mermin ND (1976) Solid state physics. Holt, Rinehart/Winston

Bachelet GB, Hamann DR, Schlüter M (1982) Pseudopotentials that work: from H to Pu. Phys Rev B 26:4199–4228. https://doi.org/10.1103/PhysRevB.26.4199

Balandin AA, Ghosh S, Bao W et al (2008) Superior thermal conductivity of single-layer graphene. Nano Lett 8:902–907. https://doi.org/10.1021/nl0731872

Baroni S, de Gironcoli S, Dal Corso A, Giannozzi P (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73:515–562. https://doi.org/10.1103/RevModPhys.73.515

Beechem T, Yates L, Graham S (2015) Invited review article: error and uncertainty in Raman thermal conductivity measurements. Rev Sci Instrum 86:41101. https://doi.org/10.1063/1.4918623

Blöchl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953–17979

Boltzmann L (1872) Further studies on the balance among gas molecules. Sitzungsberichte Akad. Wiss., Vienna, part II, 66:275–370

Bonini N, Lazzeri M, Marzari N, Mauri F (2007) Phonon anharmonicities in graphite and graphene. Phys Rev Lett 99:176802

Born M, Huang K (1998) Dynamical theory of crystal lattices. Clarendon Press, Oxford

Broido DA, Lindsay L (2008) Three-phonon phase space and lattice thermal conductivity in semiconductors. J Phys Condens Matter 20:165209

Broido DA, Malorny M, Birner G et al (2007) Intrinsic lattice thermal conductivity of semiconductors from first principles. Appl Phys Lett 91:231922. https://doi.org/10.1063/1.2822891

Broido DA, Lindsay L, Reinecke TL (2013) Ab initio study of the unusual thermal transport properties of boron arsenide and related materials. Phys Rev B 88:214303

Burke HH, Herman IP (1993) Temperature dependence of Raman scattering in $Ge_{1-x} Si_x$ alloys. Phys Rev B 48:15016–15024

Burkel E (2000) Phonon spectroscopy by inelastic x-ray scattering. Reports Prog Phys 63:171

Byron FW, Fuller RW (1992) Mathematics of classical and quantum physics. Dover Publications, New York

Cahill DG, Pohl RO (1989) Heat flow and lattice vibrations in glasses. Solid State Commun 70:927–930. https://doi.org/10.1016/0038-1098(89)90630-3

Cahill DG, Watson SK, Pohl RO (1992) Lower limit to the thermal conductivity of disordered crystals. Phys Rev B 46:6131–6140

Cahill DG, Braun PV, Chen G et al (2014) Nanoscale thermal transport. II. 2003–2012. Appl Phys Rev 1:11305. https://doi.org/10.1063/1.4832615

Cai W, Moore AL, Zhu Y et al (2010) Thermal transport in suspended and supported monolayer graphene grown by chemical vapor deposition. Nano Lett 10:1645–1651. https://doi.org/10.1021/nl9041966

Cardona M (ed) (1975) Light scattering in solids. Springer, Berlin/Heidelberg

Carrete J, Li W, Lindsay L et al (2016) Physically founded phonon dispersions of few-layer materials and the case of borophene. Mater Res Lett 4:204–211. https://doi.org/10.1080/21663831.2016.1174163

Carruthers P (1959) Scattering of phonons by elastic strain fields and the thermal resistance of dislocations. Phys Rev 114:995–1001. https://doi.org/10.1103/PhysRev.114.995

Carruthers P (1961) Theory of thermal conductivity of solids at low temperatures. Rev Mod Phys 33:92–138. https://doi.org/10.1103/RevModPhys.33.92

Cepellotti A, Fugallo G, Paulatto L et al (2015) Phonon hydrodynamics in two-dimensional materials. Nat Commun 6:6400. https://doi.org/10.1038/ncomms7400

Chen P, Katcho NA, Feser JP et al (2013) Role of surface-segregation-driven intermixing on the thermal transport through planar Si/Ge Superlattices. Phys Rev Lett 111:115901

Chernatynskiy A, Phillpot SR (2015) Phonon transport simulator (PhonTS). Comput Phys Commun 192:196–204. https://doi.org/10.1016/j.cpc.2015.01.008

Copley JRD, Macpherson RW, Timusk T (1969) Lattice dynamics of potassium chloride. Phys Rev 182:965–972

Cowley RA (1968) Anharmonic crystals. Reports Prog Phys 31:123

Datta S (2005) Quantum transport atom to transistor, 2nd edn. Cambridge University Press, New York

Debernardi A, Baroni S, Molinari E (1995) Anharmonic phonon lifetimes in semiconductors from density-functional perturbation theory. Phys Rev Lett 75:1819–1822

Debye P (1912) Zur Theorie der spezifischen Wärmen. Ann Phys 344:789–839. https://doi.org/10.1002/andp.19123441404

Debye P (1914) Lectures on the kinetic theory of matter and electricity, vol 46. BG Teubner, Leipzig

Deinzer G, Birner G, Strauch D (2003) Ab initio calculation of the linewidth of various phonon modes in germanium and silicon. Phys Rev B 67:144304

Di Bartolo B (1968) Optical interactions in solids. Wiley, New York

Ding D, Chen X, Minnich AJ (2014) Radial quasiballistic transport in time-domain thermore-flectance studied using Monte Carlo simulations. Appl Phys Lett 104:143104. https://doi.org/10.1063/1.4870811

Dorner B (1982) Coherent inelastic neutron scattering in lattice dynamics. Springer, Berlin/Heidelberg

Economou EN (2006) Green's functions in quantum physics, 3rd edn. Springer, Berlin/Heidelberg

Einstein A (1911) Elementare Betrachtungen über die thermische Molekularbewegung in festen Körpern. Ann Phys 340:679–694. https://doi.org/10.1002/andp.19113400903

Esfarjani K, Stokes HT (2008) Method to extract anharmonic force constants from first principles calculations. Phys Rev B 77:144112

Fabian J, Allen PB (1997) Thermal expansion and Grüneisen parameters of amorphous silicon: a realistic model calculation. Phys Rev Lett 79:1885–1888

Feng T, Ruan X (2016) Quantum mechanical prediction of four-phonon scattering rates and reduced thermal conductivity of solids. Phys Rev B 93:45202

Feng T, Lindsay L, Ruan X (2017) Four-phonon scattering significantly reduces intrinsic thermal conductivity of solids. Phys Rev B 96:161201

Feynman RP (1939) Forces in molecules. Phys Rev 56:340–343

Fugallo G, Lazzeri M, Paulatto L, Mauri F (2013) Ab initio. Phys Rev B 88:45430

Galperin M, Toledo S, Nitzan A (2002) Numerical computation of tunneling fluxes. J Chem Phys 117:10817–10826. https://doi.org/10.1063/1.1522404

Gaskell DR (2008) Introduction to the thermodynamics of materials, 5th edn. Taylor & Francis, New York

Gazis DC, Wallis RF (1966) Conditions for rotational invariance of a harmonic lattice. Phys Rev 151:578–580. https://doi.org/10.1103/PhysRev.151.578

Giannozzi P, Baroni S, Bonini N et al (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502

Guinea F, Tejedor C, Flores F, Louis E (1983) Effective two-dimensional Hamiltonian at surfaces. Phys Rev B 28:4397–4402

Haas P, Tran F, Blaha P (2009) Calculation of the lattice constant of solids with semilocal functionals. Phys Rev B 79:85104

Hellman O, Broido DA (2014) Phonon thermal transport in Bi_2Te_3 from first principles. Phys Rev B 90:134309

Hellmann H (1937) Einführung in die Quantenchemie. Deuticke, Leipzig

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871

Jackson JD (1999) Classical electrodynamics. Wiley, Hoboken

Jiang P, Lindsay L, Koh YK (2016) Role of low-energy phonons with mean-free-paths >0.8 μm in heat conduction in silicon. J Appl Phys 119:245705. https://doi.org/10.1063/1.4954674

Johnson JA, Maznev AA, Cuffe J et al (2013) Direct measurement of room-temperature nondiffusive thermal transport over Micron distances in a silicon membrane. Phys Rev Lett 110:25901

Kanatzidis MG (2010) Nanostructured Thermoelectrics: the new paradigm? Chem Mater 22:648–659. https://doi.org/10.1021/cm902195j

Katcho NA, Carrete J, Li W, Mingo N (2014) Effect of nitrogen and vacancy defects on the thermal conductivity of diamond: an ab initio Green's function approach. Phys Rev B 90:94117

Katre A, Carrete J, Dongre B et al (2017) Exceptionally strong phonon scattering by B substitution in cubic SiC. Phys Rev Lett 119:75902. https://doi.org/10.1103/PhysRevLett.119.075902

Kittel C (2005) Introduction to solid state physics, 8th edn. Wiley

Klemens PG (1955) The scattering of low-frequency lattice waves by static imperfections. Proc Phys Soc Sect A 68:1113

Klemens PG (1958) Thermal conductivity and lattice vibrational modes. In: Seitz F, Turnbull DBT-SSP (eds) Solid state physics. Academic, New York, pp 1–98. https://www.sciencedirect.com/science/article/pii/S0081194708605512

Kneezel GA, Granato AV (1982) Effect of independent and coupled vibrations of dislocations on low-temperature thermal conductivity in alkali halides. Phys Rev B 25:2851–2866. https://doi.org/10.1103/PhysRevB.25.2851

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138

Kuball M, Hayes JM, Shi Y, Edgar JH (2000) Phonon lifetimes in bulk AlN and their temperature dependence. Appl Phys Lett 77:1958–1960. https://doi.org/10.1063/1.1311948

Kundu A, Mingo N, Broido DA, Stewart DA (2011) Role of light and heavy embedded nanoparticles on the thermal conductivity of SiGe alloys. Phys Rev B 84:125426

Lee S, Esfarjani K, Luo T et al (2014) Resonant bonding leads to low lattice thermal conductivity. Nat Commun 5:3525

Lehmann G, Taut M (1972) On the numerical calculation of the density of states and related properties. Phys Status Solidi 54:469–477. https://doi.org/10.1002/pssb.2220540211

Leibfried G, Ludwig W (1961) Theory of anharmonic effects in crystals. In: Seitz F, Turnbull DBT-SSP (eds) Solid state physics. Academic, New York, pp 275–444. https://www.sciencedirect.com/science/article/pii/S0081194708606566

Li W, Carrete JA, Katcho N, Mingo N (2014) ShengBTE: a solver of the Boltzmann transport equation for phonons. Comput Phys Commun 185:1747–1758. https://doi.org/10.1016/j.cpc.2014.02.015

Lindsay L (2016) First principles Peierls-Boltzmann phonon thermal transport: a topical review. Nanoscale Microscale Thermophys Eng 20:67–84. https://doi.org/10.1080/15567265.2016.1218576

Lindsay L, Broido DA, Mingo N (2009) Lattice thermal conductivity of single-walled carbon nanotubes: beyond the relaxation time approximation and phonon-phonon scattering selection rules. Phys Rev B 80:125407

Lindsay L, Broido DA, Reinecke TL (2012) Thermal conductivity and large isotope effect in GaN from first principles. Phys Rev Lett 109:95901

Lindsay L, Broido DA, Reinecke TL (2013a) Ab initio thermal transport in compound semiconductors. Phys Rev B 87:165201

Lindsay L, Broido DA, Reinecke TL (2013b) First-principles determination of ultrahigh thermal conductivity of boron arsenide: a competitor for diamond? Phys Rev Lett 111:25901

Lindsay L, Li W, Carrete J et al (2014) Phonon thermal transport in strained and unstrained graphene from first principles. Phys Rev B 89:155426

Lippmann BA, Schwinger J (1950) Variational principles for scattering processes. I. Phys Rev 79:469–480. https://doi.org/10.1103/PhysRev.79.469

Maradudin AA, Fein AE (1962) Scattering of neutrons by an anharmonic crystal. Phys Rev 128:2589–2608

McCarthy KA, Ballard SS (1960) Thermal conductivity of eight halide crystals in the temperature range 220°K to 390°K. J Appl Phys 31:1410–1412. https://doi.org/10.1063/1.1735853

Merlin R, Pinczuk A, Weber WH (2000) Overview of phonon Raman scattering in solids. In: Weber WH, Merlin R (eds) Raman scattering in materials science. Springer series in materials science, vol 42. Springer, Berlin/Heidelberg, pp 1–29

Mingo N, Esfarjani K, Broido DA, Stewart DA (2010) Cluster scattering effects on phonon conduction in graphene. Phys Rev B 81:45408

Mingo N, Stewart DA, Broido DA et al (2014) Ab initio thermal transport. In: Shindé SL, Srivastava GP (eds) Length-scale dependent phonon interactions. Springer, New York, pp 137–173

Minnich AJ, Johnson JA, Schmidt AJ et al (2011) Thermal conductivity spectroscopy technique to measure phonon mean free paths. Phys Rev Lett 107:95901

Monkhorst HJ, Pack JD (1976) Special points for Brillouin-zone integrations. Phys Rev B 13:5188–5192. https://doi.org/10.1103/PhysRevB.13.5188

Mukhopadhyay S, Lindsay L, Parker DS (2016) Optic phonon bandwidth and lattice thermal conductivity: the case of Li_2X (X = O, S, Se, Te). Phys Rev B 93:224301. https://doi.org/10.1103/PhysRevB.93.224301

Norwood MH, Briscoe CV (1958) Elastic constants of potassium iodide and potassium chloride. Phys Rev 112:45–48

Omini M, Sparavigna A (1996) Beyond the isotropic-model approximation in the theory of thermal conductivity. Phys Rev B 53:9064–9073. https://doi.org/10.1103/PhysRevB.53.9064

Ozhogin VI, Inyushkin AV, Taldenkov AN et al (1996) Isotope effect in the thermal conductivity of germanium single crystals. J Exp Theor Phys Lett 63:490–494. https://doi.org/10.1134/1.567053

Peech JM, Bower DA, Pohl RO (1967) Preparation of pure alkali halide crystals and some of their properties. J Appl Phys 38:2166–2171. https://doi.org/10.1063/1.1709847

Peierls R (1929) Zur kinetischen Theorie der Wärmeleitung in Kristallen. Ann Phys 395:1055–1101. https://doi.org/10.1002/andp.19293950803

Peierls RE (2001) Quantum theory of solids. Oxford University Press, Oxford

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Petrov A, Tsypkina N, Seleznev V (1976) Behaviour of lattice thermal conductivity of crystals at high temperatures. High Temp-High Press 8:537–543

Pettersson S (1987) Calculation of the thermal conductivity of alkali halide crystals. J Phys C Solid State Phys 20:1047

Polanco CA, Lindsay L (2018) Ab initio phonon point defect scattering and thermal transport in graphene. Phys Rev B 97:14303

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in FORTRAN: the art of scientific computing, 2nd edn. Cambridge University Press, New York

Protik NH, Carrete J, Katcho NA et al (2016) Ab initio study of the effect of vacancies on the thermal conductivity of boron arsenide. Phys Rev B 94:45207

Rapp JE, Merchant HD (1973) Thermal expansion of alkali halides from 70 to 570 K. J Appl Phys 44:3919–3923. https://doi.org/10.1063/1.1662872

Ratsifaritana CA, Klemens PG (1987) Scattering of phonons by vacancies. Int J Thermophys 8:737–750

Regner KT, Sellan DP, Su Z et al (2013) Broadband phonon mean free path contributions to thermal conductivity measured using frequency domain thermoreflectance. Nat Commun 4:1640

Romero AH, Gross EKU, Verstraete MJ, Hellman O (2015) Thermal conductivity in PbTe from first principles. Phys Rev B 91:214310

Rosenbaum RL, Chau C-K, Klein MV (1969) Thermal conductivity of alkali halide crystals containing the hydroxide ion. Phys Rev 186:852–864

Rubio J, Lopez Sancho MP, Sancho JML (1984) Quick iterative scheme for the calculation of transfer matrices: application to Mo (100). J Phys F Met Phys 14:1205

Rubio J, Lopez Sancho MP, Lopez Sancho JM, Sancho JML (1985) Highly convergent schemes for the calculation of bulk and surface Green functions. J Phys F Met Phys 15:851

Rueff J-P (2010) An introduction to inelastic X-ray scattering. In: Beaurepaire E, Bulou H, Scheurer F, Jean-Paul K (eds) Magnetism and synchrotron radiation. Springer Proceedings in Physics, vol 133. Springer, Berlin/Heidelberg, pp 263–277

Schmidt AJ, Cheaito R, Chiesa M (2009) A frequency-domain thermoreflectance method for the characterization of thermal properties. Rev Sci Instrum 80:94901. https://doi.org/10.1063/1.3212673

Shankar R (1994) Principles of quantum mechanics, 2nd edn. Springer, New York

Singh DJ, Terasaki I (2008) Nanostructuring and more. Nat Mater 7:616

Slack GA, Ross RG and Slack GA (1985) Thermal conductivity under pressure and through phase transitions in solid alkali halides. II. Theory. J Phys C Solid State Phys 18:3957

Southard JC, Nelson RA (1933) Low temperature specific heats. IV. The heat capacities of potassium chloride, potassium nitrate and sodium nitrate1. J Am Chem Soc 55:4865–4869. https://doi.org/10.1021/ja01339a022

Srivastava G (1990) The physics of phonons. Taylor & Francis Group, New York

Tadano T, Gohda Y, Tsuneyuki S (2014) Anharmonic force constants extracted from first-principles molecular dynamics: applications to heat transfer simulations. J Phys Condens Matter 26:225402

Tamura S (1983) Isotope scattering of dispersive phonons in Ge. Phys Rev B 27:858–866

Tamura S (1984) Isotope scattering of large-wave-vector phonons in GaAs and InSb: deformation-dipole and overlap-shell models. Phys Rev B 30:849–854. https://doi.org/10.1103/PhysRevB.30.849

Togo A, Chaput L, Tanaka I (2015) Distributions of phonon lifetimes in Brillouin zones. Phys Rev B 91:94306

Troullier N, Martins JL (1991) Efficient pseudopotentials for plane-wave calculations. Phys Rev B 43:1993–2006

Walker CT (1963) Thermal conductivity of some alkali halides containing *F* centers. Phys Rev 132:1963–1975

Wang T, Carrete J, van Roekeghem A et al (2017) *Ab initio* phonon scattering by dislocations. Phys Rev B 95:245304. https://doi.org/10.1103/PhysRevB.95.245304

Willis BTM, Bertram TM, Carlile CJ, Colin J (2009) Experimental neutron scattering. Oxford University Press, Oxford

Yates JR, Wang X, Vanderbilt D, Souza I (2007) Spectral and Fermi surface properties from Wannier interpolation. Phys Rev B 75:195121

Zhang W, Fisher TS, Mingo N (2006) Simulation of interfacial phonon transport in Si–Ge heterostructures using an atomistic green's function method. J Heat Transf 129:483–491

Zhang W, Fisher TS, Mingo N (2007) The atomistic Green's function method: an efficient simulation approach for nanoscale phonon transport. Numer Heat Transf Part B Fundam 51:333–349. https://doi.org/10.1080/10407790601144755

Ziman JM (2001) Electrons and phonons : the theory of transport phenomena in solids. Clarendon Press, Oxford

On the Kinetic Theory of Thermal Transport in Crystals

34

Andrea Cepellotti and Nicola Marzari

Contents

_____

A. Cepellotti (✉)
Department of Physics, University of California at Berkeley, Berkeley, CA, USA

Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Theory and Simulations of Materials (THEOS), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: andrea.cepellotti@berkeley.edu

N. Marzari
National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Theory and Simulations of Materials (THEOS), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: nicola.marzari@epfl.ch

Abstract

The phonon Boltzmann transport equation is a reliable model for describing heat transfer in dielectric crystals and, coupled with first-principles methods, has shown to provide an accurate description of heat transport. Early solutions, based on simplifying assumptions, such as the single-mode relaxation-time approximation, have fostered the description of thermal transport as a kinetic theory of the phonon gas. In this chapter, we summarize some recent advances in the field of semiclassical transport. In particular, we show how the exact solutions of the linearized Boltzmann transport equations can be framed in terms of relaxons, collective excitations of phonon populations defined as the eigenvectors of the scattering operator. We show how an exact kinetic theory of heat transport can be constructed using relaxons, rather than phonons, as the relevant heat carriers. We showcase this change of perspective with a first-principles study of the transport properties of graphene. We discuss the relaxons' mean free paths, velocities, relaxation times, and their contributions to thermal conductivity. We contrast these results with those, approximate, obtained in the single-mode relaxation time approximation, and show how in this case the approximation fails qualitatively and quantitatively. Consequently, also Matthiessen's rule becomes invalid. Then, we discuss how surface or edge scattering can be properly recast in terms of friction, giving rise to heat profiles that resemble the hydrodynamic flow of liquids. Finally, we show how the Laplace transform of the linearized Boltzmann equation allows the definition of collective excitations, termed transport waves. These excitations are characterized by dispersion relations, and, in the context of lattice transport, they correspond to second-sound modes at finite wavelength, leading to relaxons in the long-wavelength limit.

1   Introduction

One of the first successful microscopic models for explaining the thermal conductivity of crystals has been the phonon Boltzmann Transport Equation (BTE), introduced in one of the first works of Peierls (1929). Today, almost 90 years later, the BTE remains a state-of-the-art method for modeling the thermal conductivity of crystals, striking a good balance between computational complexity, agreement with experiments, and range of applicability. With respect to earlier work, advancements in methods, theory, and algorithms have vastly improved the accuracy of BTE-based models, with a majority of efforts aimed at removing any additional approximation in the results. To this aim, notable contributions have been the introduction of first-principles parameters (Broido et al. 2007) and the development of methods to solve exactly the linearized BTE (LBTE) (Omini and Sparavigna 1997). As a result, it is now possible to solve exactly a first-principles LBTE and obtain excellent agreement with experiments (Broido et al. 2005, 2007; Garg et al. 2011; Fugallo et al. 2013; Esfarjani et al. 2011; Ward et al. 2009; Luckyanova et al. 2012; Cahill et al. 2014; Donadio and Galli 2009; McGaughey and Kaviany 2004; Volz and Chen 1999; Zhang and Li 2005).

Having access to exact solutions of the BTE is particularly interesting for evaluating the quality of certain approximations. In particular, the pioneering work of Peierls relied on the single-mode relaxation time approximation (SMA) – below we will discuss it in detail – introduced as a way to simplify the description of scattering between phonons. Loosely speaking, the SMA makes the hypothesis that each phonon mode can relax to equilibrium independently, without being affected by the relaxation of other phonons. Peierls himself was aware of the limitations of the SMA and noted that, for example, the SMA fails to describe crystals at cryogenic temperatures under the conditions of Umklapp freezing, because the conservation of energy and crystal momentum can only be ensured by considering the relaxation of the whole ensemble (i.e., conservation is a statistical property and not a mode-specific condition). This problem has gathered renewed interest because of the recent observations that two-dimensional or layered materials fall in these conditions of Umklapp freezing even at room temperature (Lindsay et al. 2010; Cepellotti et al. 2015; Lee and Lindsay 2017; Lindsay and Broido 2011; Jain and McGaughey 2015; Fugallo et al. 2014), so that the SMA doesn't apply.

Nowadays, we have the possibility of solving numerically the complete linearized BTE, thanks to iterative (Omini and Sparavigna 1996, 1997; Broido et al. 2005), variational (Ziman 2001; Hamilton and Parrot 1969; Srivastava 1976; Fugallo et al. 2013), or diagonalization (Guyer and Krumhansl 1966; Hardy 1970; Chaput 2013; Cepellotti and Marzari 2016) methods. These methods improve the estimates of thermal conductivity, but the interpretation of the results has been object of debate. Most of the common understanding is based on considerations derived within the SMA: for example, that the thermal conductivity of crystals can be interpreted with a kinetic theory of the phonon gas (Ziman 2001). Kinetic gas theory models thermal conductivity as $k = Cv\lambda = Cv^2\tau$, with C the specific heat, v the velocity, λ the mean free path, and τ the relaxation time of the particles forming the gas; Peierls introduced the notion of the phonon gas after noting that the SMA expression of thermal conductivity has a very similar form, with the relevant particles being phonons. When going beyond the SMA, it's not obvious anymore whether one can still define phonon relaxation times or mean free paths, and interpretations based on the phonon gas model become questionable. As an alternative explanation, it has been proposed that collective phonon excitations should be considered the correct heat carriers (Fugallo et al. 2014; Gill-Comeau and Lewis 2015; Cepellotti et al. 2015; Lee et al. 2015; Barbarino et al. 2015; de Tomas et al. 2014; Cepellotti and Marzari 2016).

Recently, we have formalized the construction of these collective phonon excitations and shown that the phonon gas theory does not correspond to an exact interpretation of the LBTE (Cepellotti and Marzari 2016). To interpret the exact solutions of the LBTE, we have introduced *relaxons*, i.e., collective phonon excitations defined as the eigenvectors of the scattering matrix. We found that it is still possible to describe thermal conductivity exactly with a kinetic gas theory, where the relevant gas doesn't consist of phonons, but of relaxons. The eigenvalues of the scattering matrix represent the exact relaxation times of the system and,

together with the relaxon velocities and mean free paths, redefine the relevant time, velocity, and length scales of thermal transport. Surface scattering is also reinterpreted in light of these considerations. Phonon gas theory suggests that the mean free path of heat carriers is limited by the size of the sample. Instead, surface scattering can be interpreted in hydrodynamic terms, with friction at the material's borders slowing down the heat flux (Cepellotti and Marzari 2017a). Finally, a dispersion relation for heat waves (and more generally, transport waves) can be derived (Cepellotti and Marzari 2017b), with frequencies, wavevectors, and relaxation times determined by the LBTE and tightly connected with the dispersion relations of second sound.

In this chapter, we aim at summarizing the current state of research on relaxons, presenting the topic comprehensively. We will assume that the reader is familiar with semiclassical transport theories and refer to classic textbooks for more detailed explanations of the semiclassical approximation and the derivation of the BTE. The manuscript is structured as follows: In the first section, we introduce the BTE and show how it can be written in terms of quantities computable from first-principles or microscopic models. The second section discusses the SMA, rederiving the expressions for the phonon gas model and underlining the limitations of such approach. The third section expresses the rationale for finding exact relaxation times from the LBTE and introduces the concept of relaxons. In the fourth section, we describe how to obtain bulk thermal conductivities with applications to silicon and graphene. The failure of Matthiessen rule when dealing with the exact LBTE is discussed in the fifth section. The sixth section extends the discussion of thermal conductivity to systems of finite size, dealing in particular with the example of a graphene ribbon. The dispersion relations of relaxons and their relation with second sound are presented in Section VIII. Last, we briefly discuss the limitations of the LBTE and our outlook for further studies and conclusions.

2 Fundamentals of Semiclassical Transport

Let's assume that the semiclassical approximation holds in a material. For each phonon mode, it is possible to construct a phonon wavepacket characterized by a well-defined wavevector $\boldsymbol{q}$, a branch index s and space-time coordinates $(\boldsymbol{r}, t)$. To simplify equations, we use the shorthand notation $\mu = (\boldsymbol{q}, s)$ to label each phonon mode (in this way, integrals over the Brillouin zone and summations over the branch index are replaced by a single summation on μ). Under these conditions, we can think that each phonon wavepacket behaves as a quasiparticle and that is also characterized by an energy $\hbar\omega_\mu$ and velocity $v_\mu = \frac{\partial \omega_\mu}{\partial \boldsymbol{q}}$. Within the semiclassical approximation, heat transport can be described by the linearized Boltzmann Transport Equation (Ziman 2001):

$$\frac{\partial n_\mu(\boldsymbol{r}, t)}{\partial t} + \boldsymbol{v}_\mu \cdot \nabla n_\mu(\boldsymbol{r}, t) = -\frac{1}{\mathscr{V}} \sum_{\mu'} \Omega_{\mu\mu'} \Delta n_{\mu'}(\boldsymbol{r}, t) \,. \tag{1}$$

where n_μ is the population number of the phonon μ, $\mathscr{V}$ is a normalization volume, and $\Omega_{\mu\mu'}$ is the scattering matrix.

The LBTE is a balance equation: the left-hand side – the Liouville or diffusion operator – describes how the phonon wavepacket moves in space and time with velocity $\boldsymbol{v}_\mu$. If phonon interactions are neglected, say at the level of the harmonic lattice approximation, the wavepacket would be able to proceed indefinitely and lead to an infinite thermal conductivity. Instead, anharmonic contributions cause the phonon to deviate from this perfect motion, and scattering events introduce a source of resistance. In the LBTE, scattering is described with a matrix $\Omega_{\mu\mu'}$ containing the total scattering rate for the transition $\mu \to \mu'$ due to any deviation of the true phonon Hamiltonian from the harmonic approximation. We recall that the scattering operator is nonlinear in its most generic form; however, one typically limits the discussion to small deviations from equilibrium (the same hypothesis under which the semiclassical approximation holds). Therefore, the operator can be linearized and assumes a matrix representation. As a consequence of this linearization, the scattering matrix doesn't act on the non-equilibrium phonon population n_μ; instead, it acts on the (small) difference $\Delta n_\mu = n_\mu - \bar{n}_\mu$ between n_μ and the thermal equilibrium population identified by the Bose-Einstein distribution function $\bar{n}_\mu(\boldsymbol{r}, t) = (e^{\hbar\omega_\mu/k_B T(\boldsymbol{r},t)} - 1)^{-1}$, with ω_μ being the phonon frequency and $T(\boldsymbol{r}, t)$ the local temperature.

The phonon frequency and group velocity are easily constructed by diagonalizing the dynamical matrix, i.e., the second derivative of the total energy with respect to ionic displacements. The scattering matrix is by far the most complex quantity entering the LBTE and must be built including all the phonon interactions. In practice, we limit the discussion to the inclusion of three-phonon processes and isotopic scattering events that lead to the scattering matrix (Ziman 2001; Fugallo et al. 2013):

$$
\begin{aligned}
\Omega_{\mu\mu'}\bar{n}_{\mu'}(\bar{n}_{\mu'} + 1) = &\left[\sum_{\mu'',\mu'''} \left(P^{\mu''}_{\mu\mu'''} + \frac{1}{2} P^{\mu}_{\mu''\mu'''} \right) + \sum_{\mu'} P^{\mathrm{isot}}_{\mu\mu'} \right]\delta_{\mu\mu'} \\
&- \sum_{\mu''}(P^{\mu'}_{\mu\mu''} - P^{\mu''}_{\mu\mu'} + P^{\mu}_{\mu'\mu''}) + P^{\mathrm{isot}}_{\mu\mu'} .
\end{aligned}
\tag{2}
$$

In Eq. 2, we split the matrix into a diagonal and an off-diagonal part, as it is convenient for the discussion of the SMA. The P coefficients indicate the transition rates for three-phonon and phonon-isotope scattering processes, which are obtained with the Fermi golden rule. In particular, the transition rate for three-phonon processes is (Fugallo et al. 2013; Paulatto et al. 2013; Ziman 2001; Vanderbilt et al. 1984; Debernardi et al. 1995; Lazzeri and de Gironcoli 2002):

$$
\begin{aligned}
P^{\mu''}_{\mu\mu'} = &\frac{2\pi}{N\hbar^2} \sum_{\boldsymbol{G}} |V^{(3)}(\mu, \mu', -\mu'')|^2 \\
&\times \bar{n}_\mu \bar{n}_{\mu'}(\bar{n}_{\mu''} + 1)\delta_{\boldsymbol{q}+\boldsymbol{q}'-\boldsymbol{q}'',\boldsymbol{G}} \\
&\times \delta(\hbar\omega_\mu + \hbar\omega_{\mu'} - \hbar\omega_{\mu''}) ,
\end{aligned}
\tag{3}
$$

where N is the number of phonon wavevectors $\boldsymbol{q}$, $\boldsymbol{G}$ is a reciprocal lattice vector, and $V^{(3)}$ is the third derivative of the unit cell energy $\mathscr{E}^{\text{cell}}$ with respect to atomic displacements

$$V^{(3)}(\mu, \mu', \mu") = \frac{\partial^3 \mathscr{E}^{\text{cell}}}{\partial X_\mu \partial X_{\mu'} \partial X_{\mu''}} \ , \tag{4}$$

with:

$$X_\mu = \frac{1}{N} \sum_{lb\alpha} \sqrt{\frac{2 M_b \omega_\mu}{\hbar}} z_\mu^{*b\alpha} u_{b\alpha}(\boldsymbol{R}_l) e^{-i\boldsymbol{q} \cdot \boldsymbol{R}_l} \ , \tag{5}$$

where b is an index running on the basis of atoms in the unit cell, $\boldsymbol{R}_l$ is a Bravais lattice vector identifying the lth unit cell inside the crystal, α is a cartesian index, M_b is the mass of atom b, z_μ is the phonon polarization vector, * indicates the complex conjugate, and u is the vector of atomic displacements. The scattering rate for a phonon-isotope scattering event is (Garg et al. 2011; Ziman 2001; Abeles 1963):

$$P_{\mu\mu'}^{\text{isot}} = \frac{\pi}{2N} \omega_\mu \omega_{\mu'} \left[\bar{n}_\mu \bar{n}_{\mu'} + \frac{\bar{n}_\mu + \bar{n}_{\mu'}}{2} \right]$$
$$\times \sum_b g_b \left| \sum_\alpha z_\mu^{*,b\alpha} z_{\mu'}^{b\alpha} \right|^2 \delta(\omega_\mu - \omega_{\mu'}) \ , \tag{6}$$

where $g_b = \frac{(M_b - \langle M_b \rangle)^2}{\langle M_b \rangle^2}$.

All of these quantities are needed to build the LBTE. In the examples that will be made in the following sections, we have obtained both the dynamical matrix and the anharmonic matrix $V^{(3)}$ from density-functional perturbation theory (Baroni et al. 1987, 2001; Giannozzi et al. 1991; Debernardi et al. 1995; Paulatto et al. 2013; Lazzeri and de Gironcoli 2002; Gonze and Vigneron 1989), with an implementation freely available in the open-source software distribution Quantum ESPRESSO (Giannozzi et al. 2009). As a result, the LBTE is free of adjustable parameters and can be used directly to compare with experimental results. We note in passing that the LBTE is not tied to first-principles simulations, and these parameters can be obtained with any other microscopic model (e.g., tight-binding methods, empirical potentials, etc.) capable of providing the total energy as a function of ionic positions.

As a last numerical point, we note that every numerical implementation of the theory needs to overcome the difficulty of the Dirac delta imposing the energy conservation. In the results shown here, we follow the prescription of Fugallo et al. (2013) and substitute the Dirac delta with a Gaussian smearing, chosen in such a way to best reproduce the numerical value of thermal conductivity in the limit of an infinite number of phonon wavevectors and zero smearing width.

For later convenience, we express the left-hand side of Eq. (1) in terms of $\Delta n_\mu = n_\mu - \bar{n}_\mu$:

$$\frac{\partial \bar{n}_\mu}{\partial T}\left(\frac{\partial T(\boldsymbol{r},t)}{\partial t} + \boldsymbol{v}_\mu \cdot \nabla T(\boldsymbol{r},t)\right) + \frac{\partial(\Delta n_\mu(\boldsymbol{r},t))}{\partial t} + \boldsymbol{v}_\mu \cdot \nabla(\Delta n_\mu(\boldsymbol{r},t)) =$$
$$= -\frac{1}{\mathscr{V}}\sum_{\mu'}\Omega_{\mu\mu'}\Delta n_{\mu'}(\boldsymbol{r},t) ; \tag{7}$$

where T is the reference temperature at which the scattering operator has been linearized. To write this equation, we used the fact that the Bose-Einstein distribution depends on space-time coordinates only through temperature. In this form, we have now explicitly put the equation in terms of the unknown Δn_μ and the parameter T; for each temperature, the LBTE needs to be solved to obtain Δn_μ. This complete expression shows that the LBTE becomes a set of coupled linear differential equations. Their solution has proven to be a daunting challenge, and the rest of the discussion will focus on finding solutions to a few simple problems. However, we note that the solution Δn_μ grants access to the out-of-equilibrium properties. In particular it can be shown, within the harmonic approximation (Hardy 1963), that each phonon transfers an energy flux equal to $\hbar\omega_\mu\boldsymbol{v}_\mu$; hence the total energy (heat) flux of the lattice is:

$$\boldsymbol{Q}(\boldsymbol{r},t) = \frac{1}{\mathscr{V}}\sum_\mu \hbar\omega_\mu\boldsymbol{v}_\mu n_\mu(\boldsymbol{r},t) . \tag{8}$$

Therefore, the knowledge of the out-of-equilibrium phonon population allows to compute the heat flux at each point of the material and to describe its evolution in time, all via microscopic degrees of freedom. Eventually, thermal conductivity is obtained from the relation $\boldsymbol{Q} = -k\nabla T$, connecting the heat flux generated in response to an applied temperature gradient ∇T. The great challenge, however, is the calculation of the deviation of phonon populations from equilibrium, and the simplest approximated way to find an expression for Δn_μ is by means of the SMA.

3   Single-Mode Relaxation Time Approximation

The first historical attempts at finding a solution of the LBTE have relied on the SMA, with the aim of simplifying the scattering matrix to a form such that the analytical solution of the LBTE becomes trivial. Mathematically, the SMA approximates to zero all the out-of-diagonal matrix elements of Ω, so that the right-hand side of the LBTE becomes:

$$\frac{1}{\mathscr{V}}\sum_{\mu'}\Omega_{\mu\mu'}\Delta n_{\mu'}(\boldsymbol{r},t) \approx \frac{\Delta n_\mu(\boldsymbol{r},t)}{\tau_\mu^{\text{SMA}}} , \tag{9}$$

where we label the diagonal terms of the scattering matrix as $\frac{1}{\tau_\mu^{\text{SMA}}} = \Omega_{\mu\mu}$.

To show that τ_μ^{SMA} is a relaxation time, we now suppose that the crystal is at thermal equilibrium ($T(\boldsymbol{r}, t) = T =$ const), so that the first two terms of Eq. 7 disappear. Every phonon at time t_0 is populated according to the Bose-Einstein distribution, except for one phonon mode μ, which we excite to a number $n_\mu(t_0)$ throughout space (in this way, there are no spatial gradients). The SMA-LBTE that describes this problem is simply:

$$\frac{\partial \Delta n_\mu(t)}{\partial t} = -\frac{\Delta n_\mu(t)}{\tau_\mu^{\text{SMA}}} , \qquad (10)$$

which describes how the system relaxes to equilibrium. Clearly, it will thermalize with an exponential decay of the excitation number $\Delta n_\mu(t) = \Delta n_\mu(t_0)e^{-t/\tau_\mu^{\text{SMA}}}$, where the relaxation time τ_μ^{SMA} gives the characteristic time of such decay and, if positive, ensures that at large-enough times the system will relax back to equilibrium ($\Delta n_\mu = 0$). Therefore, the diagonal matrix elements determine the characteristic time of such decay and have been identified as phonon relaxation times; we note though that this conclusion has to rely on the introduction of the SMA.

Since the SMA scattering matrix is diagonal, many problems can be tackled easily, such as the calculation of the bulk thermal conductivity. To characterize this property, we apply an infinitesimally small temperature gradient ∇T, which is spatially homogeneous and independent of time: $\nabla T(\boldsymbol{r}, t) = \nabla T =$ const. Since we limit the study to steady state, the distribution Δn_μ originated in response to this perturbation is independent of time. Moreover, we can expect Δn_μ to be independent from the spatial coordinates as well, thanks to the translational invariance of the bulk crystal. The SMA-LBTE to be solved becomes:

$$\frac{\partial \bar{n}_\mu}{\partial T} \boldsymbol{v}_\mu \cdot \nabla T = -\frac{\Delta n_\mu}{\tau_\mu^{\text{SMA}}} . \qquad (11)$$

Trivially, we can solve for Δn_μ by multiplying the equation by $-\tau_\mu^{\text{SMA}}$ (in fact, this simplicity is the main justification for this approximation). The thermal conductivity tensor k^{ij} follows and in the SMA is equal to:

$$k^{ij} = -(\nabla T)_j^{-1} Q_i = -(\nabla T)_j^{-1} \frac{1}{\mathcal{V}} \sum_\mu \hbar\omega_\mu v_\mu^i \Delta n_\mu$$

$$\approx \frac{1}{\mathcal{V}} \sum_\mu \frac{\partial \bar{n}_\mu}{\partial T} \hbar\omega_\mu v_\mu^i v_\mu^j \tau_\mu^{\text{SMA}} = \frac{1}{\mathcal{V}} \sum_\mu C_\mu v_\mu^i (\Lambda_\mu^j)^{\text{SMA}} , \qquad (12)$$

where i and j represent Cartesian coordinates, $C_\mu = \frac{1}{k_B T^2} \bar{n}_\mu(\bar{n}_\mu + 1)(\hbar\omega_\mu)^2 = \frac{\partial \bar{n}_\mu}{\partial T} \hbar\omega_\mu$ is the specific heat, and $(\Lambda_\mu^j)^{\text{SMA}} = v_\mu^j \tau_\mu^{\text{SMA}}$ is the mean free path of the phonons. This expression has the same structure of the thermal conductivity obtained from the kinetic gas theory ($k = Cv\Lambda$). Therefore, Peierls interpreted the

semiclassical thermal transport theory in terms of a phonon gas, where each phonon contributes to the heat flux with a heat capacity C_μ and travels for a distance Λ_μ^{SMA} at velocity v_μ before thermalizing to equilibrium by the action of scattering on a time scale set by τ_μ^{SMA}.

This conclusion is justified within the SMA, but we cannot blindly apply it to the general solution of the BTE. In the next section, we will extend the same reasoning to the complete scattering matrix, and we will see that exact relaxation times can be defined but are not given by the diagonal elements. In fact, in the present derivation, we neglected the out-of-diagonal matrix elements. Physically, these are necessary for an accurate description of the energy transfer between phonons. For example, consider the case of Eq. 10 where we studied the time evolution of a single excited phonon. The energy change from equilibrium is determined by the phonon excitation number and evolves in time as $\Delta E_\mu = \hbar \omega_\mu \Delta n_\mu(t_0) e^{-t/\tau_\mu^{\text{SMA}}}$. It must be noted that the energy is not conserved, contrary to the expectations and despite the fact that each transition in the scattering matrix conserves energy (through a number of Dirac deltas). While the diagonal matrix elements correctly describe the rate at which energy leaves a phonon mode, the out-of-diagonal matrix elements describe how energy is transferred to other phonons (Fugallo et al. 2013). Thus, the SMA is not able to describe the correct redistribution of energy upon scattering; to this aim, one has to include the out-of-diagonal matrix elements.

4   Relaxons

To find exact relaxation times, we simply proceed in the same way as shown above but retaining the full expression of the scattering matrix. We consider a crystal kept at constant temperature $T(\mathbf{r}, t) =$const and excite a phonon mode homogeneously in space – as we did for the SMA – such that the evolution of the phonon population only depends on time but not on space. The LBTE is simplified to the following differential equation (a generalization of Eq. 10):

$$\frac{\partial \Delta n_\mu(t)}{\partial t} = -\frac{1}{\mathcal{V}} \sum_{\mu'} \Omega_{\mu\mu'} \Delta n_{\mu'}(t) . \tag{13}$$

Obviously, the SMA solution $\Delta n_\mu(t) = \Delta n_\mu(t_0) e^{-t \Omega_{\mu\mu}}$ is not the correct solution of this more general case. Nevertheless, the problem is still a set of coupled linear differential equations that can be solved analytically.

To solve these, we have to diagonalize the scattering matrix. From Eq. 2, one can verify that the scattering matrix $\Omega_{\mu\mu'}$ is not symmetric, i.e., $\Omega_{\mu\mu'} \neq \Omega_{\mu'\mu}$, which is inconvenient for diagonalization. The asymmetry however is caused just by a factor $\bar{n}_\mu(\bar{n}_\mu + 1)$ and the following scaling transformations (Hardy 1965, 1970; Krumhansl 1965; Chaput 2013):

$$\tilde{\Omega}_{\mu\mu'} = \Omega_{\mu\mu'} \sqrt{\frac{\bar{n}_{\mu'}(\bar{n}_{\mu'} + 1)}{\bar{n}_{\mu}(\bar{n}_{\mu} + 1)}} \;, \tag{14}$$

$$\Delta\tilde{n}_{\mu} = (\bar{n}_{\mu}(\bar{n}_{\mu} + 1))^{-\frac{1}{2}} \Delta n_{\mu} \tag{15}$$

allow to recast Eq. 13 in terms of a symmetric matrix $\tilde{\Omega}_{\mu\mu'}$ and of $\Delta\tilde{n}_{\mu}$. We can thus diagonalize $\tilde{\Omega}_{\mu\mu'}$ and find its complete set of eigenvectors and eigenvalues:

$$\frac{1}{\mathcal{V}} \sum_{\mu'} \tilde{\Omega}_{\mu\mu'} \theta_{\mu'}^{\alpha} = \frac{1}{\tau_{\alpha}} \theta_{\mu}^{\alpha} \;, \tag{16}$$

where α is an eigenvalue index, θ_{μ}^{α} is an eigenvector, and τ_{α} its eigenvalue. To impose normalization of the eigenvectors, we also define a scalar product $\langle\alpha|\alpha'\rangle \equiv \frac{1}{\mathcal{V}} \sum_{\mu} \theta_{\mu}^{\alpha} \theta_{\mu}^{\alpha'}$ and note that the completeness relation ensures that $\langle\alpha|\alpha'\rangle = \delta_{\alpha\alpha'}$. As a technical note aside, we note that Fugallo et al. (2013) stressed that the calculation of scattering rates requires a numerical implementation of the Dirac delta that ensures the energy conservation: such implementation must be made so that the scattering matrix remains symmetric, or the results of the diagonalization may deviate from what we will discuss in the following.

At the present stage, we know very little about the eigenvalue spectrum of the scattering matrix, because the diagonalization of this large matrix hasn't been attempted until recently, and even then, it has been done only in a handful of materials (Chaput 2013; Cepellotti and Marzari 2016; Togo et al. 2015). However, since $\tilde{\Omega}_{\mu\mu'}$ is a real symmetric matrix, all its eigenvalues are real numbers (Hardy 1970). Moreover, $\tilde{\Omega}_{\mu\mu'}$ is also a semi-positive definite matrix (Hardy 1965, 1970; Fugallo et al. 2013), that is, the eigenvalues must be equal to either zero or any positive number. We further add that for the case of graphene and silicon, later discussed in detail, we find only strictly positive eigenvalues.

Returning to the problem of thermal relaxation, we can now take advantage of the complete basis set of eigenvectors of the scattering matrix. Any response $\Delta\tilde{n}_{\mu}$ can be written as a linear combination of the θ_{μ}^{α} eigenvectors (Hardy 1970):

$$\Delta\tilde{n}_{\mu}(\boldsymbol{r}, t) = \sum_{\alpha} f_{\alpha}(\boldsymbol{r}, t) \theta_{\mu}^{\alpha} \;, \tag{17}$$

where $f_{\alpha}(\boldsymbol{r}, t)$ are a set of real coefficients. If we write Eq. 13 in the symmetrized form and then rotate it in the basis of eigenvectors, the equation describing thermalization becomes:

$$\frac{\partial f_{\alpha}(\boldsymbol{r}, t)}{\partial t} = -\frac{f_{\alpha}(\boldsymbol{r}, t)}{\tau_{\alpha}} \;. \tag{18}$$

The LBTE in this set of coordinates has the same structure of the SMA (Eq. 10), and therefore the relaxation of the system toward equilibrium is described by $f_\alpha(t) = f_\alpha(t_0)e^{-t/\tau_\alpha}$. We now recognize that the eigenvalues τ_α describe the characteristic time scale of thermal relaxation. At variance with the SMA result, the scattering operator has not been approximated, and therefore τ_α are the exact time scales of the thermal relaxation as described by the LBTE. The semi-positiveness of the scattering matrix ensures that the exponential factor decays to zero or, in the case of zero eigenvalues, the component is a stationary solution.

At variance with the SMA, the quantities relaxing to equilibrium are not phonons anymore, but the factors f_α that project the phonon population onto the eigenvectors of the scattering matrix. To stress that phonons do not have a relaxation time, note that the solution of Eq. 18 can be rotated back into the phonon basis, finding $\Delta n_\mu(t) = \sum_\alpha f_\alpha(t_0)\theta_\mu^\alpha e^{-t/\tau_\alpha}$. The thermal relaxation of a phonon is characterized by N different exponential relaxation channels (N being the number of phonon states in the system, equal to the number of eigenvalues), and thus it's not possible to identify a single characteristic decay time for a phonon population (we stress that the LBTE abstract the microscopic dynamics of the phonons into that of phonon populations, and it's the time scale of these phonon populations that we are describing now). Since the eigenvectors θ_μ^α can have any sign, the phonon thermalization $\Delta n_\mu(t)$ can behave as any decaying curve and not simply as an exponential. As a result, phonon relaxation times do not make sense anymore when the complete LBTE is considered. We note in passing that this statement is in contrast with several others (e.g., Broido et al. (2005) and others citing it), that define an exact phonon relaxation time, however without proving any connection with the actual time evolution of the phonon populations.

Looking now more carefully at the eigenvectors θ_μ^α, we note that each eigenvector has been used to express the phonon populations in terms of the coefficients f_α; in other terms, each eigenvector at fixed α identifies a distribution function of phonon occupation numbers, i.e., a collective excitation of phonons. To simplify the discussion, we will refer to the eigenvectors of the scattering matrix as *relaxons*, this name indicating that these are entities characterized by exact relaxation times. By definition, relaxons are invariant under the action of the scattering operator: each is thus a collective excitation of phonons that interact/scatter with each other, but each relaxon does not interact/scatter with other relaxons. In addition, the coefficients f_α can be interpreted as relaxon populations, and the evolution described by Eq. 18 is the time evolution of the relaxon populations. Since the LBTE has been expressed in terms of deviations from equilibrium, a zero relaxon population corresponds to thermal equilibrium. This interpretation is schematically shown in Fig. 1: we define relaxons as collective phonon excitations, such that the scattering interactions become decoupled, and the problem of phonon scattering is recast into a formulation of noninteracting relaxons, which only exist in an out-of-equilibrium situation and disappear at long times when thermal equilibrium is restored.
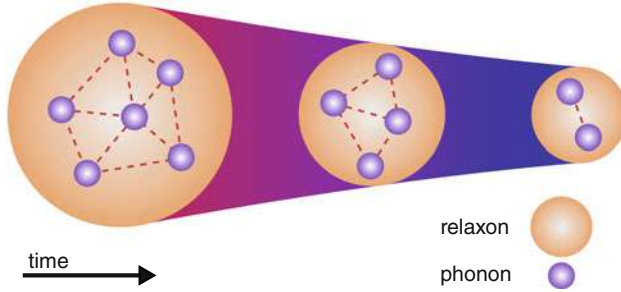
Fig. 1 Graphical representation of relaxon thermalization. Each relaxon is a collective excitation of phonons: phonons interact among themselves by the action of scattering, but relaxons do not scatter against each other by construction. Each relaxon thermalizes decaying exponentially to equilibrium, at a rate given by its exact relaxation times. (Picture from Cepellotti and Marzari 2016)

An interesting property of relaxons is that they have a well-defined parity (Hardy 1970), a result that follows from the symmetry of the scattering matrix $\tilde{\Omega}_{\mu\mu'} = \tilde{\Omega}_{-\mu-\mu'}$, where $-\mu = (-\boldsymbol{q}, s)$, so that every relaxon is either even or odd, i.e., $\theta_\mu^\alpha = \pm\theta_{-\mu}^\alpha$. As a consequence, the parity distinguishes the contributions to the energy change ΔE:

$$\Delta E = \frac{1}{\mathscr{V}} \sum_\mu \hbar\omega_\mu \Delta n_\mu = \frac{1}{\mathscr{V}} \sum_\mu \hbar\omega_\mu \sqrt{\bar{n}_\mu(\bar{n}_\mu + 1)} \Delta\tilde{n}_\mu$$

$$= \frac{1}{\mathscr{V}} \sum_\mu \theta_\mu^0 \sqrt{k_B T^2 C} \sum_\alpha f_\alpha \theta_\mu^\alpha$$

$$= \sqrt{k_B T^2 C} \sum_\alpha f_\alpha \langle 0|\alpha\rangle \ , \tag{19}$$

where we introduced the total crystal heat capacity $C = \frac{1}{\mathscr{V}} \sum_\mu C_\mu$ and the auxiliary vector $|0\rangle$ as:

$$|0\rangle = \theta_\mu^0 = \frac{\sqrt{\bar{n}_\mu(\bar{n}_\mu + 1)}\hbar\omega_\mu}{\sqrt{k_B T^2 C}} \ . \tag{20}$$

The parity of the eigenvectors, together with the even parity of $\bar{n}_\mu$ and ω_μ, ensures that only relaxons that are even contribute to changes in energy, whereas relaxons that are odd conserve energy. The distribution θ_μ^0 introduced here can be obtained from the Bose-Einstein distribution: to do so, one must take the first-order term in the Taylor expansion of $\bar{n}_\mu(T + \delta T)$, symmetrize the distribution using Eq. (15), and normalize the distribution.

The heat flux can be expressed in terms of relaxons, finding:

$$\boldsymbol{Q} = \frac{1}{\mathcal{V}} \sum_{\mu} \hbar \omega_{\mu} v_{\mu}^{j} \Delta n_{\mu}$$

$$= \sqrt{k_{B} T^{2} C} \sum_{\alpha} \boldsymbol{V}_{\alpha} f_{\alpha} \,, \tag{21}$$

where we introduce a relaxon velocity $\boldsymbol{V}_{\alpha} = \boldsymbol{V}_{0\alpha} = \langle 0|v|\alpha\rangle = \frac{1}{\mathcal{V}} \sum_{\mu} \theta_{\mu}^{0} \boldsymbol{v}_{\mu} \theta_{\mu}^{\alpha}$. Here, only relaxons of odd parity have a non-zero contribution to heat flux; even relaxons do not contribute. Therefore, the parity of the spectrum distinguishes between excitations contributing to energy changes (even) and relaxons contributing to energy fluxes (odd).

Having now found that relaxons are conveniently characterized by the exact relaxation times of the system, it is natural to extend the study to a more complex case than the thermal relaxation of Eq. 13. For a more general case, we must start from the full LBTE of Eq. 7, symmetrize it with Eq. 15, and change the basis to relaxons using Eq. 17. After various algebraic manipulations, the LBTE in the relaxon basis becomes:

$$\sqrt{\frac{C}{k_{B} T^{2}}} \left(\frac{\partial T(\boldsymbol{r}, t)}{\partial t} \langle 0|\alpha\rangle + \nabla T(\boldsymbol{r}, t) \cdot \boldsymbol{V}_{\alpha} \right) + \frac{\partial f_{\alpha}(\boldsymbol{r}, t)}{\partial t} + \sum_{\alpha'} \boldsymbol{V}_{\alpha\alpha'} \cdot \nabla f_{\alpha'}(\boldsymbol{r}, t)$$

$$= -\frac{f_{\alpha}(\boldsymbol{r}, t)}{\tau_{\alpha}} \,, \tag{22}$$

where $\boldsymbol{V}_{\alpha\alpha'} = \frac{1}{\mathcal{V}} \sum_{\mu} \theta_{\mu}^{\alpha} \boldsymbol{v}_{\mu} \theta_{\mu}^{\alpha'} \equiv \langle \alpha|v|\alpha'\rangle$ and $\boldsymbol{V}_{\alpha} = \boldsymbol{V}_{0\alpha} = \langle 0|v|\alpha\rangle$. $\boldsymbol{V}_{\alpha\alpha'}$ derives from the action of the diffusion operator on the deviation from equilibrium, while $\boldsymbol{V}_{\alpha}$ derives from the action of the diffusion operator on the equilibrium distribution.

The LBTE in the relaxon basis is now expressed in terms of the unknown relaxon occupation number, which replaces the phonon occupation number. At a first glance, not much has been gained by this transformation. First of all, the basis change is just a rotation of coordinates that doesn't add or modify the information encoded in the LBTE. Moreover, while the scattering operator in the relaxon basis is, by construction, simply diagonal, the drifting operator is not diagonal anymore, as it was instead in the phonon basis. We will show however that the challenging aspects of the LBTE haven't simply been moved from the scattering to the diffusion operator but that this choice of basis allows for identifying analytical solutions of the LBTE and additional physical insights.

5 Bulk Conductivity

The steady-state thermal conductivity in the bulk crystal can be obtained readily from the LBTE in the relaxon basis. In full analogy with the way bulk conductivity

has been derived within the SMA, we apply a constant temperature gradient that doesn't depend in space and time. Equation 22 can be simplified by setting time derivatives to zero, as long as we limit to stationary conditions, and taking advantage of the translational periodicity of a lattice, we can set $f_\alpha(\mathbf{r}, t) = f_\alpha$, as the response of the system must be homogeneous in space. The resulting LBTE is simply:

$$\sqrt{\frac{C}{k_B T^2}} \nabla T \cdot \mathbf{V}_\alpha = -\frac{f_\alpha}{\tau_\alpha} . \tag{23}$$

This equation has the same formal structure of the SMA bulk problem (Eq. 11) and is trivially solved in terms of f_α, simply multiplying the equation by $-\tau_\alpha$. Using the definition of thermal conductivity and the expression of heat flux of Eq. 21, the bulk conductivity of the LBTE can be expressed exactly as:

$$k^{ij} = \frac{-1}{\mathcal{V} \nabla_i T} \sum_\mu \hbar \omega_\mu v_\mu^j \Delta n_\mu = -\sum_\alpha f_\alpha^i \sqrt{k_B T^2 C} V_\alpha^j$$

$$= \sum_\alpha C V_\alpha^i V_\alpha^j \tau_\alpha = \sum_\alpha C V_\alpha^i \Lambda_\alpha^j , \tag{24}$$

where Λ_α^j is the α relaxon mean free path in direction j. In the same way that the bulk LBTE hasn't formally changed from the SMA to the relaxon basis, the formal structure of thermal conductivity itself hasn't changed and is again reminiscent of a kinetic theory of gases. Therefore, this result prompts the interpretation of thermal transport in bulk crystals as a gas of relaxons that travel at velocity V_α for a distance Λ_α, after which they equilibrate. The specific heat indicates that each relaxon changes the local temperature by an amount δT after spending an amount of energy equal to $\delta E = C \delta T$. Therefore, the kinetic theory of gases still applies, but the objects forming the gas have been redefined.

One difference with respect to the SMA is the specific heat: while each phonon has its own specific heat C_μ, the present equations indicate that each relaxon is characterized by an identical specific heat C, equal to the specific heat of the whole crystal. The difference stems from the fact that temperature, being an equilibrium property, cannot be changed by a single phonon but by a distribution of them. In detail, in order to change temperature from T to $T + \delta T$, it is necessary that the phonon distribution changes from $\bar{n}(T)$ to $\bar{n}(T + \delta T)$: for this reason, the relevant velocity V_α describes the phonon group velocities of relaxon α projected into the deviation θ_μ^0 (recall the definition $V_\alpha = \langle 0|v|\alpha \rangle$), and the component relevant to thermal transport is the one that contributes to the phonon population change $\bar{n}(T + \delta T) - \bar{n}(T)$.

5.1   Examples: Graphene and Silicon

To showcase the model discussed above for bulk thermal conductivity, we present here, as also discussed in Cepellotti and Marzari (2016), the bulk thermal conductivity of silicon and graphene. These two materials have been chosen for two opposite reasons. Silicon is a well-known material, there aren't debates on the value of its thermal conductivity, and the SMA does a good job in predicting the bulk thermal conductivity from first principles (138 W/mK) with respect to the exact conductivity (141 W/mk from Eq. 24), which is close to the experimental value (144 W/mK). In contrast, there is still no wide consensus (neither between experiments nor between theoretical predictions) on the thermal conductivity of graphene, and more importantly the SMA dramatically underestimates the exact thermal conductivity (Lindsay et al. 2010), in our calculations $k^{SMA} = 495$ W/mK against an exact $k = 3894$ W/mK at room temperature.

The LBTE is built from first principles, calculating harmonic and anharmonic force constants using density-functional perturbation theory (Baroni et al. 1987, 2001; Giannozzi et al. 1991; Debernardi et al. 1995; Paulatto et al. 2013; Lazzeri and de Gironcoli 2002; Gonze and Vigneron 1989) as implemented in the Quantum ESPRESSO distribution (Giannozzi et al. 2009), the scattering matrix contains three-phonon and isotope-phonon interactions. Linear algebra operations require massively parallel linear algebra routines; in particular, we take advantage of the ScaLAPACK library (Blackford et al. 1997). For the rest of this section, all quantities are computed at a temperature of 300 K, and, taking advantage of the isotropy of the materials, by thermal conductivity we mean here the single irreducible component of the thermal conductivity tensor.

The diagonalization of Eq. (16) gives all the relaxons θ_μ^α. After desymmetrizing the eigenvectors with Eq. 15, they can be used to check the contributions of different phonon modes to a given relaxon. For example, in Fig. 2, we plot the longest-lived relaxon of graphene as a function of the phonon index $\mu = (\boldsymbol{q}, s)$. For the sake of simplicity, we only show the contribution to such relaxon from the three acoustic phonon branches (the optical contributions being much smaller). As the relaxon describes deviations of phonon populations from equilibrium, the picture shows that contributions to this specific relaxons come from deviations of long-wavelength ZA phonons and receives also contributions from LA and TA branches (with a much more complicated dependence on the Brillouin zone). It's important to note in this picture that, being a collective excitation, each relaxon receives contribution from most phonon modes, and the loci of phonon deviations do not seem to have a simple structure, except for the most relevant and dominant case of the ZA branch.

In Fig. 3, we illustrate the difference between the exact relaxation times (red dots) and phonon lifetimes (blue dots) for both materials. In particular, we plot the percentage contribution to thermal conductivity (k for relaxons, k^{SMA} for phonons) as a function of the exact/phonon relaxation time or lifetime.

In silicon, where the discrepancy between the SMA and the exact solution is small, we observe that the shape of the two distribution is rather similar,

Fig. 2 For graphene at room temperature, the relaxon θ_μ^α with the longest relaxation time ($\alpha=1$) is plotted against the phonon index $\mu = (\boldsymbol{q}, s)$, where we restrict the branch index s to the three out-of-plane/transverse/longitudinal acoustic modes (ZA, TA, and LA, respectively), among the six that are present. The hexagonal shape is the Brillouin zone of graphene, while height and colors indicate the difference of phonon population with respect to thermal equilibrium (red for overpopulated modes and blue for depopulated ones). It can be seen that contributions come from phonons across the whole Brillouin zone for different branches. The fine structure of the ridges is a numerical artifact due to discrete Brillouin zone sampling. (Picture from Cepellotti and Marzari 2016)
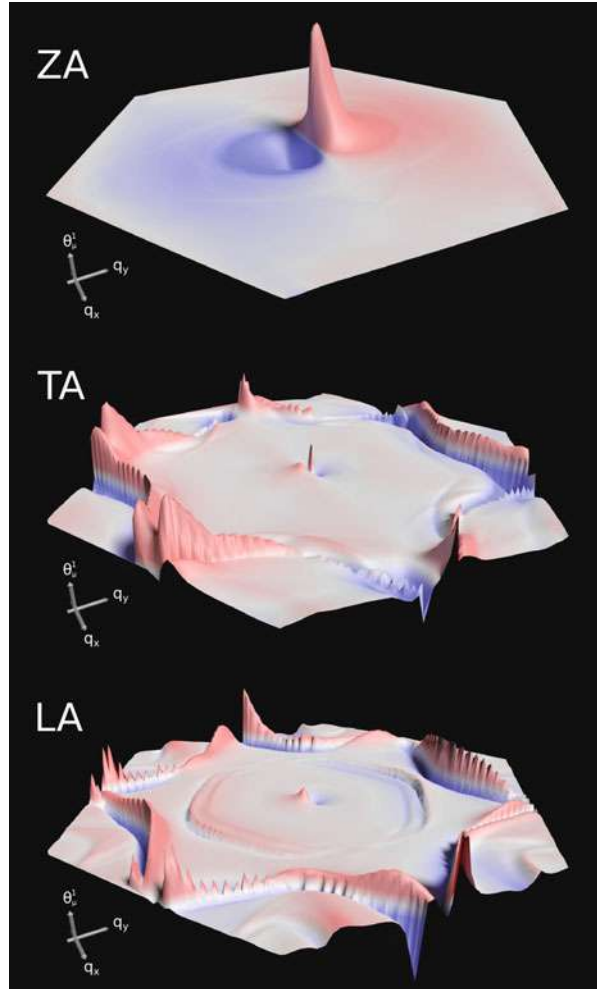
as well as the range of time scale covered by the two solutions. There are no lifetimes/relaxation times below 7–8 ps: this is the time scale set by the largest scattering rate available in the scattering matrix (limited to three-phonon and phonon-isotope processes). As the phonon/relaxon lives for a longer time, the more it contributes to thermal conductivity. Although the details of the distributions are different (and compressed in the picture by the logarithmic scale), the relaxon picture here doesn't change much the relevant time scale of the transport dynamics. We note in passing that we truncate in the picture several modes that have negligible contributions to the thermal conductivity.

The situation is radically different for graphene. Here, k^{SMA} is mostly determined by phonons of characteristic lifetimes of approximately 10–100 ps, while the largest

Fig. 3 Contribution to thermal conductivity as a function of relaxation times. Blue dots represent results at the level of the SMA phonon gas theory; red dots show results in terms of eigenvalues. Thermal conductivity is computed within the SMA for phonons, exactly for relaxons; to have comparable results, the contribution to thermal conductivity is plotted as a percentage. In silicon, the two theories give similar results. Relaxons in graphene instead show that the exact time scale of relaxation is much larger than what predicted at the SMA level. The peak of contributions to SMA thermal conductivity comes at a time scale of about 10–100 ps, whereas the largest exact contributions have a time scale larger than 1000 ps. The shaded area is a guide to the eyes, to recall that a finer integration of the Brillouin zone would lead to a continuous spectrum, whereas relaxon spectrum is discrete. (Picture from Cepellotti and Marzari 2016)
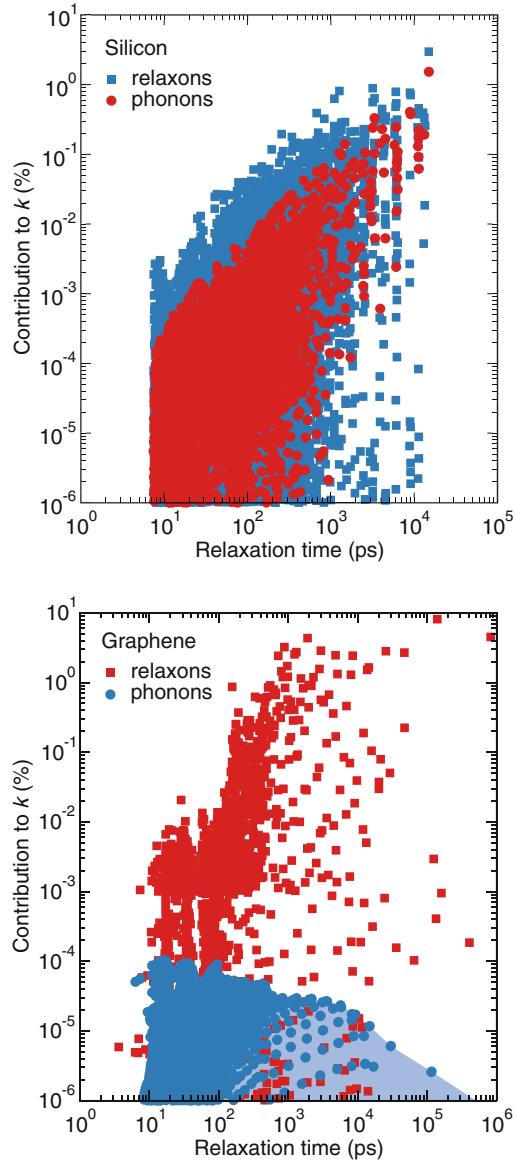
contributions to the exact thermal conductivity come from relaxons with lifetimes above $1\,\mu$s and up to $1\,$ms. As a consequence, we can see that the SMA fails to estimate the time scale of heat transport by orders of magnitude. The physical meaning of this large discrepancy is related with the role of the off-diagonal scattering matrix elements. The time scale at which energy leaves a phonon is of

the order of 10–100 ps, but the heat flux is not thermalized after a single scattering event. Heat flux is instead transferred to other modes and can propagate over several scattering events on a much larger time scale than that of a single phonon scattering event. The exact time scale of this collective participation of phonon modes to the heat flux is set by the eigenvalues of the scattering matrix. In contrast, the similarity of the exact and SMA results for silicon suggests that in this material heat flux is quickly dissipated after a few scattering events.

As an artifact of the finite sampling of k-points in the Brillouin zone, the phonon lifetimes don't form a continuous distribution in the picture; however, by increasing the integration sampling, we would see the blue dots join together to form a continuous area. From the numerical results, it appears that relaxons do not necessarily form a continuous spectrum. In fact, our convergence studies on graphene indicate that the largest relaxation times form a discrete spectrum. It's not unusual in condensed matter physics that the eigenvalue spectrum of a continuous matrix can lead to a discrete eigenvalue spectrum, but we don't have yet a deeper insight of why it can happen here.

In Fig. 4, we study the percentage contribution to k or k^{SMA} for relaxons and phonons, respectively, as a function of the velocities $|V_\alpha^x|$ or $|v_\mu^x|$. For this property, both silicon and graphene (i.e., regardless of the effectiveness of the SMA) show notable differences between phonon group velocities and relaxon velocities. It appears therefore that the SMA overestimates the speed at which heat flux propagates through the material; in particular, the largest contributions come from phonons that have approximately the speed of sound. Instead, the relaxon velocity V_α is always much smaller than the speed of sound by one or two orders of magnitude. The result is interpreted by the fact that a relaxon couples together an ensemble of phonons that transfer heat flux. Since heat flux doesn't live on a single phonon, its average velocity will have to be averaged over several phonon modes; given that the speed of sound is typically the largest value, the resulting velocity will be smaller. In other words, the instantaneous velocity is the phonon group velocity, but the statistical velocity at which heat flux propagates in the bulk is much smaller.

Finally, the mean free paths shown in Fig. 5 combine the features of velocities and relaxation times of the previous figures, to illustrate the length scales at which heat flux dissipation occurs. In particular, the orders of magnitude of difference in the velocities and relaxation times are now propagated to the mean free paths. In silicon, the mean free paths Λ_α are much smaller than the SMA counterparts, but the distribution has a comparable shape. The differences between the two theories are more pronounced in graphene, where the distribution has also a different shape. In both cases, however, the contribution to the exact thermal conductivity is roughly proportional to the relaxon mean free path. Comparing the two materials, the exact mean free path shows that the heat flux dissipation takes place at a much larger value in graphene (10 microns) than in silicon (less than 1 micron).

It's worth noting that a typical tool to study the effect of nanostructuring is to plot the cumulative curve of conductivity as a function of the mean free path, as a direct postprocessing of the presented figure. The common understanding is that nanostructuring in a material will be effective to decrease thermal conductivity when

Fig. 4 Contribution to exact (SMA) thermal conductivity in percentage, as a function of the exact (phonon) relaxation time of the system at room temperature in graphene. Within the SMA, the velocity scale of thermal transport is well approximated by the group velocity of phonons, in particular, the largest contributions come from phonons at the speed of sound, i.e., about 20 (8) km/s in graphene (silicon). In the relaxon picture, heat flux is allowed to be shuttled through several phonon modes, resulting in an average velocity of heat flux that is on average much smaller than the speed of sound. (Picture from Cepellotti and Marzari 2016)
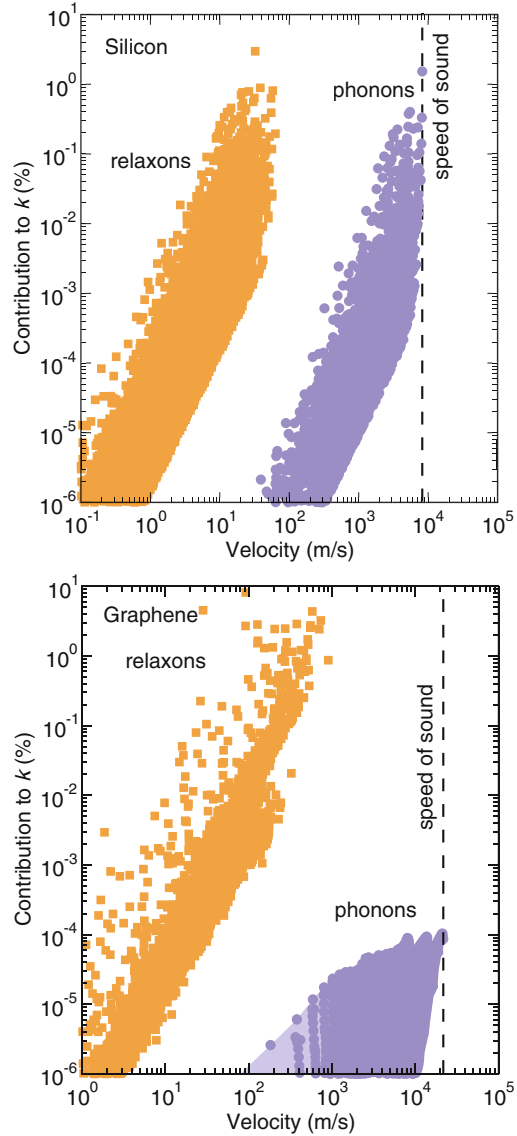
it has a length scale comparable with the mean free path. We stress however that these properties have been studied here only in the context of a bulk system, and in the next sections, we will see that the relation between surface scattering and mean free paths, while correct within the SMA, doesn't hold exactly.

Fig. 5 Analysis of phonon and relaxon mean free paths and their contributions to the thermal conductivity of graphene at room temperature. The relaxon mean free paths in silicon are a couple of orders of magnitude smaller than the values predicted by SMA phonon mean free paths. In graphene, the distribution of phonon and relaxon mean free paths are rather different, with the peak contribution from relaxon mean free paths at values of $10\,\mu$, larger than typical phonon mean free paths. This quantity estimates the distance that heat flux can propagate inside the material before dissipation. (Picture from Cepellotti and Marzari 2016)
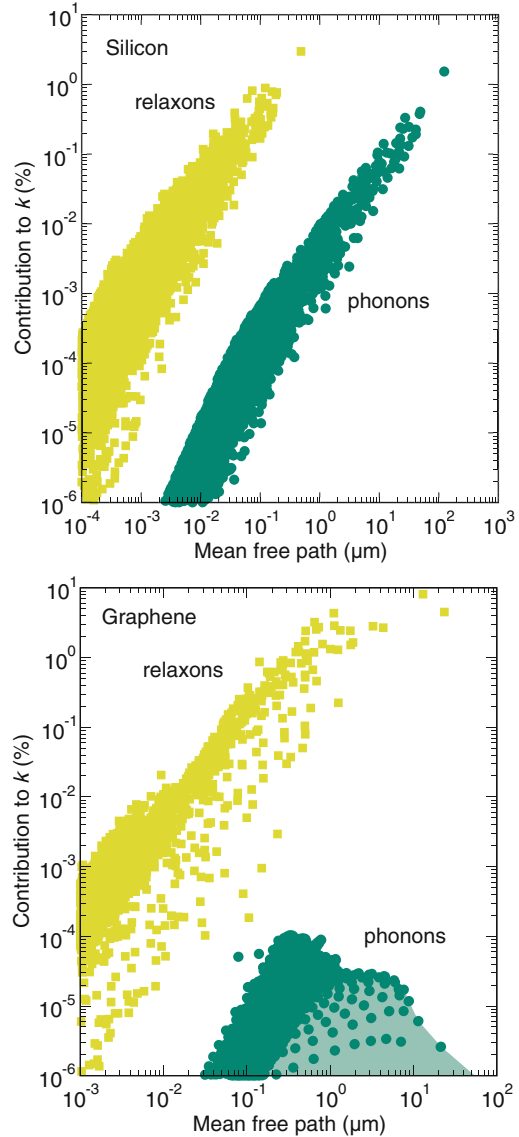
6 Matthiessen Rule

In its original form, the Matthiessen rule states that the total resistivity ($\frac{1}{k}$) arising from two separate scattering mechanisms can be obtained as the sum of the resistivity of the two separate mechanisms, that is:

$$\frac{1}{k} = \frac{1}{k_1} + \frac{1}{k_2} \ . \tag{25}$$

The rule comes handy as it allows to separate the contribution to thermal conductivity from different scattering sources. However, it has long been known that the Matthiessen rule is only an approximation, the correct expression being (see Ziman (2001) and Cepellotti and Marzari (2016) for proofs):

$$\frac{1}{k} \leq \frac{1}{k_1} + \frac{1}{k_2} \ . \tag{26}$$

The historical relevance of this rule was much more important a few decades ago, when it was not unusual to approximate the whole scattering matrix with a single scalar quantity (the single relaxation time approximation, a further simplification of the single-mode relaxation time approximation). In the limit of the scattering matrix being approximated with a single scalar, the Matthiessen rule becomes exact. Nowadays the single relaxation time approximation sees less usage in numerical simulations but is nevertheless often used to obtain estimates (e.g., in experiments).

It is interesting to test, as shown in Fig. 6, the extent of the violation of the Matthiessen rule. In particular, we compare the exact conductivity with results in which we selectively turned off normal, Umklapp, or isotopic scattering events and add them separately with the Matthiessen rule. As granted by Eq. 26, the exact thermal conductivity is always smaller than all other curves. In most cases, the Matthiessen rule overestimates the total thermal conductivity of the system. The only case where the Matthiessen estimate gets closer to the exact result is when a kind of scattering event can be neglected. For example, the conductivity obtained splitting the isotopic from the three-phonon interactions ($k^{-1} = (k^{N+U})^{-1} + (k^I)^{-1}$) gets close to the exact answer only at large temperature when the isotopic scattering is negligible and $(k^I)^{-1} \approx 0$.

A common extension of the Matthiessen rule found in literature is the property:

$$\frac{1}{\tau_\mu^{\text{SMA}}} = \frac{1}{\tau_\mu^{\text{SMA}-1}} + \frac{1}{\tau_\mu^{\text{SMA}-2}} \ . \tag{27}$$

This equation reflects how the total phonon relaxation time of the system can be found as the inverse summation of different scattering mechanisms. Therefore, one can clearly distinguish the contribution to the phonon relaxation time coming from, say, three-phonon scattering and isotopic scattering events. This property derives from the fact that different scattering events are treated as independent processes (correlations are neglected); therefore the scattering rates can be summed in the scattering matrix. In the SMA limit, this relation holds exactly and allows for a clear-cut separation of contribution to thermal conductivity from different sources of scattering.

When dealing with exact relaxation times of the system, also this generalized Matthiessen rule doesn't hold anymore. In fact, the eigenvalues of the sum of two
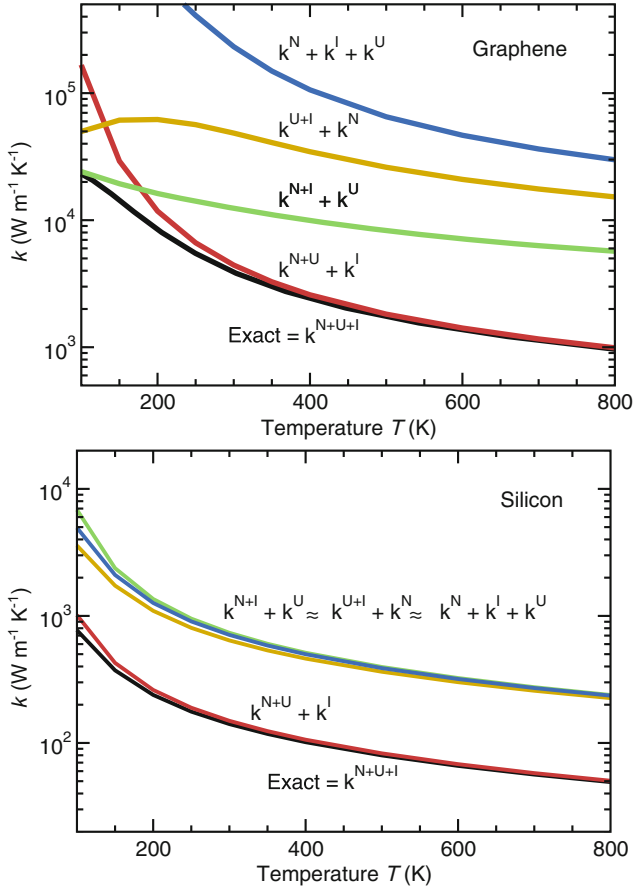
Fig. 6 Study of the failure of Matthiessen rule in graphene and silicon. The total thermal conductivity (black line) is compared to conductivities computed through various Matthiessen sums, where the total sum is given by the sum of the reciprocals (by $k^N + k^U + k^I$ we mean $k^{-1} = (k^N)^{-1} + (k^U)^{-1} + (k^I)^{-1}$). Due to the correlation between scattering events, there is no decomposition for which the Matthiessen rule is obeyed at all temperatures. The only curve that approaches the exact thermal conductivity is the decomposition $(k^{N+U} + k^I)$, just when the effect of isotopic scattering is negligible

matrices are not equal to the sum of the eigenvalues of the two separate matrices (except in the SMA limit where the matrix is diagonal). Therefore, the exact solution of the LBTE doesn't have a case where the Matthiessen rule can be applied rigorously. As a result, it's not simple to define how much a single scattering process contributes to thermal conductivity, because the total thermal conductivity includes the interplay of the different scattering mechanisms. For example, the redistribution of phonon populations after, say, a normal scattering event indirectly affects the heat flux dissipation that takes place with an Umklapp or isotopic process.

From another perspective, however, Eq. 26 ensures that the inclusion of an additional scattering mechanism lowers the thermal conductivity. In fact, if we were to include the effect of four-phonon interactions, Eq. 26 guarantees that either (1) the thermal conductivity will be left unchanged if the interaction is negligible, or (2) the thermal conductivity will be smaller. Therefore, one should typically expect to overestimate the true thermal conductivity of a system.

7 Surface-Limited Conductivity

Nanostructuring a material provides an effective way to lower the thermal conductivity of a system. The most common understanding of such reduction is based on the idea that the heat carrier mean free path is limited by the size of the system (von De Haas and Biermasz 1938; Casimir 1938). In fact, one would argue naively that the heat carrier simply cannot travel over a distance that is larger than the sample size; therefore, adopting Matthiessen rule, the effective total relaxation time in a system of size L is:

$$\frac{1}{\tau^{\text{eff}}} = \frac{1}{\tau^{\text{intr}}} + \frac{v}{L} \, , \tag{28}$$

where the lifetimes may be referred to the phonon μ or the relaxon α and v is the carrier velocity along the thermal gradient. If L is much smaller than the intrinsic mean free path $v\tau^{\text{intr}}$, the thermal conductivity as given by Eqs. 12 or 24 will be linear in L and is therefore reduced. Within the SMA, one can indeed put this reasoning on solid grounds (see, e.g., Ziman 2001, Carruthers 1961, and Cepellotti and Marzari 2016) and thus agree with the intuition above. However, we will now show that the exact thermal conductivity in presence of a surface isn't explained by a relation of the form of Eq. 28. As argued in Cepellotti and Marzari (2017a), the impact of surfaces on thermal conductivity should be interpreted instead hydrodynamically, in terms of a viscous effect.

We note also that the majority of recent studies trying to describe surface-limited thermal conductivity with exact solvers of the LBTE (e.g., Li et al. 2014, Fugallo et al. 2013, 2014, and Omini and Sparavigna 1996) mix their results with concepts relying on the SMA. In particular, the following expression, that we will refer to as the Surf-RTA, has been used to model surface scattering:

$$\boldsymbol{v}_\mu \cdot \nabla T \frac{\partial \bar{n}_\mu}{\partial T} = -\frac{1}{\mathscr{V}} \sum_{\mu'} \left(\Omega_{\mu\mu'} + \frac{v_\mu}{L} \delta_{\mu\mu'} \right) \Delta n_{\mu'}(\boldsymbol{r}, t) \, . \tag{29}$$

Here, the BTE looks like a bulk problem but with a correction on the diagonal matrix elements. Note that the correction $\frac{v_\mu}{L}\delta_{\mu\mu'}$ also appears in various declinations, such as $\frac{|v_\mu|}{L}\delta_{\mu\mu'}$ or $\frac{v_\mu}{FL}\delta_{\mu\mu'}$. In any case, this expression is empirically justified by noting that (1) the diagonal matrix elements of the scattering matrix are phonon

lifetimes, (2) the SMA theory of surface scattering rigorously proves that the phonon relaxation time must be corrected with a factor $\frac{v_\mu}{L}$, and hence (3) an exact treatment of surface scattering should be achieved by adding the term $\frac{v_\mu}{L}\delta_{\mu\mu'}$. Unfortunately, the justification of this approach is only by analogy; we will see below that the results obtained from Eq. 29 strongly deviate from an exact description of surface scattering.

The main challenge for describing this phenomenon is that, in contrast with intrinsic events such as phonon-phonon interactions, the reflection/absorption of a phonon only takes place at the sample's surfaces. As argued by Ziman (2001) (who reports a few unsuccessful attempts of adding space-dependent scattering terms), one cannot describe this by adding a space-dependent term to the scattering matrix. Instead, the correct way to describe surface scattering (Ziman 2001; Carruthers 1961) is to solve the LBTE as a differential equation: after the integration on the spatial degrees of freedom, the integration constant is fixed by suitable boundary conditions. The following study is nothing but a generalization of the derivation of Eq. 28 made within the SMA by Carruther (1961) to the case beyond the SMA; the mathematical derivation is similar, but the analysis of the results differs notably.

We can start the derivation (Cepellotti and Marzari 2017a) from the LBTE in the relaxon basis (Eq. 22). Limiting the study to the steady state and a constant gradient of temperature, time derivatives are neglected and the LBTE becomes:

$$\sqrt{\frac{C}{k_B T^2}}\nabla T \cdot \boldsymbol{V}_\alpha + \sum_{\alpha'} \boldsymbol{V}_{\alpha\alpha'} \cdot \nabla f_{\alpha'}(\boldsymbol{r}) = -\frac{f_\alpha(\boldsymbol{r})}{\tau_\alpha} \ . \tag{30}$$

In contrast to the bulk problem, the periodicity of the lattice is broken by the surface, so that the relaxon population retains the dependence on spatial coordinates. The resulting LBTE is a set of coupled linear differential equations (rather than the linear algebra problem for the bulk). In general, it is hard to impose the boundary conditions for a material of a generic shape, and methods like finite-elements may be more suitable. Here, instead, we focus on a simple geometry that allows for a simple analytical solution and that illustrates the qualitative aspects of surface scattering treated exactly. In particular, we will focus on an infinitely long ribbon of graphene of width L. Cartesian coordinates are chosen such that x spans the length of the system and y the width, such that the surfaces are located on the lines $y = \pm\frac{L}{2}$. Finally, the temperature gradient is along the ribbon's length, so that the LBTE reduces to:

$$\sqrt{\frac{C}{k_B T^2}}V_\alpha^{(x)}\nabla T + \sum_{\alpha'} V_{\alpha\alpha'}^{(y)}\frac{\partial f_{\alpha'}(y)}{\partial y} = -\frac{f_\alpha(y)}{\tau_\alpha} \ . \tag{31}$$

To make progress, we multiply the LBTE by τ_α, to note that the derivative term has a factor $\tau_\alpha V_{\alpha\alpha'}$, which identifies an asymmetric matrix. To recast the problem

in a symmetrized form, we transform the relaxon populations as $f_\alpha = g_\alpha \sqrt{\tau_\alpha}$. The LBTE becomes:

$$\sum_{\alpha'} \Lambda_{\alpha\alpha'}^{(y)} \frac{\partial g_{\alpha'}}{\partial y} + g_\alpha = g_\alpha^\infty \tag{32}$$

where $\Lambda_{\alpha\alpha'}^{(y)} = \sqrt{\tau_\alpha} V_{\alpha\alpha'}^{(y)} \sqrt{\tau_{\alpha'}}$ and $g_\alpha^\infty = -\sqrt{\frac{C\tau_\alpha}{k_B T^2}} V_\alpha^{(x)} \nabla T$ is the solution of the bulk problem (save for a factor $\sqrt{\tau}_\alpha$, it's identical to Eq. 23). The matrix $\Lambda_{\alpha\alpha'}$ is symmetric (recall the definition $V_{\alpha\alpha'} = \frac{1}{\mathcal{V}} \sum_\mu \theta_\mu^\alpha v_\mu \theta_\mu^{\alpha'}$) and real, so that we can diagonalize it and write:

$$\sum_{\alpha'} \Lambda_{\alpha\alpha'}^{(y)} \psi_{\alpha' i} = \lambda_i^{(y)} \psi_{\alpha i} , \tag{33}$$

finding the eigenvalues $\lambda_i^{(y)}$ and the eigenvectors $\psi_{\alpha i}$ of the matrix $\Lambda_{\alpha\alpha'}^{(y)}$. Unlike the eigenvalues of the scattering matrix, the eigenvalues $\lambda_y^{(y)}$ can have any sign; in fact, note that the eigenvalues of $V_{\alpha\alpha'}$ are simply the phonon group velocities v_μ, which also can have any sign. In the numerical example we tested, we find a symmetry in the eigenvalue spectrum, such that both $\lambda_i^{(y)}$ and $-\lambda_i^{(y)}$ are eigenvalues. Since the eigenfunctions $\psi_{\alpha i}$ form a complete basis set, we can rotate the LBTE in this new basis and express the scaled relaxon populations g_α as a linear combination of the eigenvectors $\psi_{\alpha i}$:

$$g_\alpha = \sum_i \psi_{\alpha i} d_i , \tag{34}$$

with d_i a set of real coefficients. Like we did for relaxons, we interpret the eigenvectors $\psi_{\alpha i}$ as another kind of collective excitation and d_i as a population number. After a few algebraic manipulations, the LBTE can be rotated in this new basis, to find:

$$\lambda_i^{(y)} \frac{\partial d_i}{\partial y} + d_i = d_i^\infty , \tag{35}$$

with d_i^∞ such that $g_\alpha^\infty = \sum_\alpha \psi_{\alpha i} d_i^\infty$. Eventually, the LBTE has been recast into a simple form, using a new degree of freedom identified by the eigenvectors $\psi_{\alpha i}$ that decouple all terms and bring the LBTE to a diagonal form.

Since $\lambda_i^{(y)}$ is any real number, there are three classes of solutions. The simplest case is $\lambda_i^{(y)} = 0$, when the population d_i is identical to that of the bulk system. This suggests the interpretation that the mode $\psi_{\alpha i}$ is traveling parallel to the surface: as it never scatters against the surface, the ribbon will appear as a bulk to this mode. In the other two cases, when $\lambda_i^{(y)} > 0$ or < 0, the excitation $\psi_{\alpha i}$ scatters against

the surface and its population deviates from bulk as $d_i = d_i^{\infty} + c_i e^{-y/\lambda_i^{(y)}}$, where c_i is an integration constant to be determined from boundary conditions. To fix c_i, we must know the population d_i at one point.

Let's suppose that $\lambda_i^{(y)} > 0$: the mode $\psi_{\alpha i}$ is moving in the positive y-direction, suggesting that we should fix the boundary condition at the surface $y = -\frac{L}{2}$ where it last scattered. Ziman (2001) discusses a similar problem within the SMA and justifies his choice with arguments of causality, arguing that the phonon population must be determined only by events in the past. Moreover, there's also a more practical necessity: to fix the boundary condition on the opposite surface may cause an exponential divergence. In general, the boundary condition will look like $d_i(y = -\frac{L}{2}) = \delta_i$, so that the solution becomes:

$$d_i = d_i^{\infty} + (\delta_i - d_i^{\infty})e^{-(y+\frac{L}{2})/\lambda_i^{(y)}} . \tag{36}$$

The case of opposite sign $\lambda_i^{(y)} < 0$ is similar (but opposite) and leads to the expression:

$$d_i = d_i^{\infty} + (\delta_i - d_i^{\infty})e^{-(y-\frac{L}{2})/\lambda_i^{(y)}} . \tag{37}$$

Therefore, we formally obtained the solution of the LBTE, resorting to an additional change of basis set, which allowed us to find an analytical solution of the LBTE in a 2D ribbon. There exists collective excitations $\psi_{\alpha i}$ that travel parallel to the surface and behave as bulk, as well as modes that scatter against the border, and thermalize to the bulk limit after they traveled away from the surface for a characteristic distance $\lambda_i^{(y)}$.

We have however introduced an additional parameter δ_i that doesn't come from the LBTE itself. Ziman (2001), when discussing surface scattering within the SMA, proposed to associate this boundary condition with the phonon reflectivity at the boundary. In fact, one can construct a matrix $\mathscr{R}_{\mu}^{\mu'}$ describing how a phonon μ can be reflected into μ' after the impact on the surface. In principle, this matrix can describe any kind of surface scattering: for example, $\mathscr{R} = 0$ describes a perfectly absorbing surface, $\mathscr{R} = \text{constant}$ describes a diffusive phonon scattering, and specular scattering is identified by suitably selecting the angles between μ and μ'. Such reflection coefficients can be computed from microscopic models, such as molecular dynamics. However, the information on these quantities is scarce in the literature, and therefore, in order to avoid further complexity, we drastically simplify the discussion and make the simplest assumption possible, known as the Casimir limit: the surface is perfectly absorbing and therefore $\mathscr{R} = 0$ for any mode. To relate this coefficient with the boundary condition, Ziman wrote a balance equation so that the number of phonons leaving a surface is equal to the number of reflected phonons. Applying this same idea to the modes $\psi_{\alpha i}$, we first rotate the matrix of the phonon reflection coefficients in the basis of modes $\psi_{\alpha i}$ as $\mathscr{R}_i^j = \sum_{\mu\mu'\alpha\beta} \psi_{\alpha i} \theta_{\mu}^{\alpha} \mathscr{R}_{\mu}^{\mu'} \theta_{\mu'}^{\beta} \psi_{\beta i}$. Then, the balance equation is:

$$\delta_i = d_i\left(y = -\frac{W}{2}\right)\Bigg|_{\lambda_i^{(y)} > 0} = \sum_j \mathscr{R}_i^j d_j\left(y = -\frac{W}{2}\right)\Bigg|_{\lambda_j^{(y)} < 0}, \tag{38}$$

expressing the boundary condition in terms of microscopic quantities. In the Casimir limit, we simply have $\delta_i = 0$.

The thermal conductivity is obtained, as for the bulk case, as a by-product of the LBTE once the solution is known. In particular, the xx component of the thermal conductivity tensor is obtained by rotating Eq. 21 in the basis of modes $\psi_{\alpha i}$:

$$k(y) = -\frac{Q}{\nabla T} = \frac{-1}{\mathscr{V}\nabla T}\sum_\mu \hbar\omega_\mu v_\mu \Delta n_\mu = -\sum_\alpha f_\alpha(y)\sqrt{k_B T^2 C}\, V_\alpha^{(x)}$$

$$= -\sum_\alpha g_\alpha(y)\sqrt{\tau_\alpha k_B T^2 C}\, V_\alpha^{(x)}$$

$$= \sum_\alpha C V_\alpha^{(x)} \Lambda_\alpha^{(x)} - k_B T^2 \sum_{\alpha\beta}(\delta_\alpha - g_\alpha^\infty)\left(\sum_{\lambda_i^{(y)} > 0}\psi_{\alpha i}\, e^{-\frac{y+\frac{W}{2}}{\lambda_i^{(y)}}}\,\psi_{i\beta}^T + \right.$$

$$\left. \sum_{\lambda_i^{(y)} < 0}\psi_{\alpha i}\, e^{-\frac{y-\frac{W}{2}}{\lambda_i^{(y)}}}\,\psi_{i\beta}^T\right)(\delta_\beta - g_\beta^\infty) = k^\infty - \Delta k^{\text{surf}}(y), \tag{39}$$

where $\psi_{i\beta}^T$ is the matrix transpose of $\psi_{\beta i}$ and $\delta_\alpha = \sum_\alpha \psi_{\alpha i}\delta_i$. Hence we found the thermal conductivity of a 2D ribbon, consisting of a bulk contribution and a surface term, the latter decreasing thermal conductivity (as well as heat flux) by an amount that depends on the position inside the sample. The space dependency appears through a series of exponential factors: the largest suppression happens close to the surfaces and then relaxes to the bulk value when leaving the surface. In particular, there are two series of exponentials, each describing the suppression of thermal conductivity by one of the two ribbon surfaces. At large enough distance from both surfaces – that is, further than all the characteristic lengths $\lambda_\alpha^{(y)}$ – surface scattering becomes negligible, and the thermal conductivity approaches its bulk value.

First, we note that the length scale at which nanostructuring is effective in decreasing heat flux is given by the lengths $\lambda_\alpha^{(y)}$. In fact, if the ribbon is smaller than any of these characteristic lengths, there will be some non-negligible exponentials, and thus the thermal conductivity is suppressed. Otherwise, the surface scattering will decrease heat flux only on a small layer close to the surface, and the average conductivity will be as in the bulk.

This behavior of heat flux is reminiscent of the motion of fluids through a pipe in the laminar regime. Also for real fluids, the mass flux slows down in proximity of the surface and grows going further away from the surface. In that case, the underlying quantity for the suppression of mass flux is the liquid viscosity. Thus, we

suggest to interpret the eigenvalues $\lambda_i^{(y)}$ as *friction lengths* and the surface effects as a hydrodynamic slowdown of the heat flux by friction at the borders.

We can now note that Eq. 39 doesn't describe the reduction of thermal conductivity by surface scattering as a renormalization of the relaxation times or mean free paths, as predicted instead by the SMA (see Eq. 28). While relaxon mean free paths Λ_α describe the characteristic length of the decay of heat flux along the direction of the thermal gradient by action of intrinsic scattering events, the surface defines a decay of heat flux with respect to the position of the surface. Moreover, while the former case describes an exponential decay of the collective excitation θ_μ^α, the surface causes an exponential decay of the modes $\psi_{\alpha i}$. Due to the presence of a rotation between the two basis sets, it's not possible to interpret surface scattering by a simple rescaling of the relaxon mean free path. As a result, the interpretation of surface scattering based on the SMA doesn't apply anymore; by chance, the SMA allows to describe the problem as a renormalization of the mean free path, but this is only possible when the scattering matrix is diagonal and the correlation between phonons can be neglected (in fact, only the phonon basis is needed, and one doesn't have to introduce θ_μ^α or $\psi_{\alpha i}$). The interpretation of surface scattering in terms of friction instead encompasses both the full LBTE and the SMA limit, where in the latter, case friction lengths accidentally coincide with the phonon mean free paths.

We note in passing that the analogy of the heat flow with the Poiseuille flow of a real liquid is a tool to illustrate the case study but lacks a true one-to-one correspondence. Besides the obvious differences such as mass/heat and pressure/temperature, we note that the Poiseuille description is made on a single degree of freedom, the liquid density, whereas the LBTE has more microscopic information, given that it describes the population number of N phonons. Moreover, the underlying Navier-Stokes equations are quite different from the LBTE, so that even the detailed behavior of the profile of flux as a function of the distance from the surface is quantitatively different (a polynomial decay vs an exponential series). Therefore, one should think at the similarity with real fluid exclusively as a tool to gain some intuition on heat flux.

Before moving on to an example, we also argue against a limitation of this model of surface scattering. The LBTE written in Eq. 30 assumes that lattice properties such as force constants and anharmonic couplings do not change over the extent of the crystal. This assumption isn't correct for the first few layers of atoms at a surface, since the surface introduces new effects such as local relaxations, steps, defects, and oxidations that will modify surface lattice properties. To obtain a better description of these effects, which may be non-negligible for ribbon widths at the nano-scale, one should treat the system as a one-dimensional problem, compute force constants over the entire section of the ribbon, and study the details of surface phonons (e.g., using the insertion method Bungaro et al. 1996). This approach however can only be done for sufficiently small widths, due to computational constraints; if the ribbon is large and if the modification of the lattice properties is small compared to the bulk, then the model presented above might be preferable.

7.1 Examples

We will now apply the model of surface scattering to a ribbon of graphene. The computational parameters are the same used for the previous example on the bulk thermal conductivity of graphene; in addition, if not specified otherwise, we fix the ribbon width at $2\,\mu$m and the temperature at 300 K. We note that there is a remaining degree of freedom, consisting in the orientation of the surface with respect to the crystal lattice vectors. This choice doesn't modify the results for the bulk, because the symmetry of graphene guarantees that the bulk thermal conductivity tensor is isotropic. In the presence of a surface, and in particular for small sizes, some symmetries are broken, and one may expect anisotropies. Here, we will not investigate these effects any further and fix the x coordinate to be along the zig-zig direction of graphene, focusing instead on the qualitative aspects of the model.

In Fig. 7, we plot the average contribution of mode i to Δk^{surf}, i.e., the quantity

$$k_B T^2 \sum_{\alpha\beta} g_\alpha^\infty \psi_{\alpha i} \psi_{i\beta}^T g_\beta^\infty \frac{1}{W} \int e^{-(y \pm \frac{W}{2})/\lambda_i^{(y)}}\, dy \;,$$

as a function of the friction lengths $\lambda_\alpha^{(y)}$. In detail, we plot the contribution of mode i in percentage w.r.t. to the bulk conductivity, in order to compare with the percentage
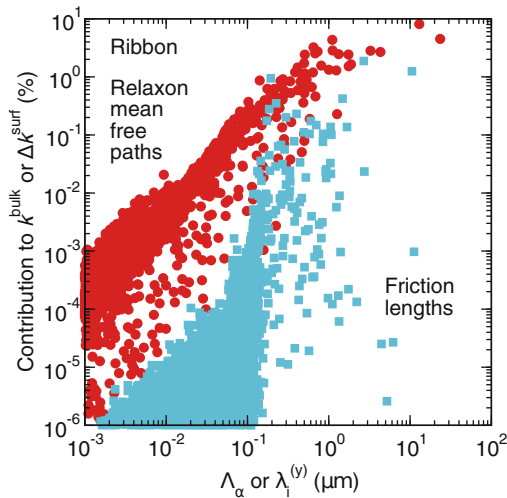
Fig. 7 Friction lengths (see text for definition) on the horizontal axis in a graphene ribbon and their average contribution to thermal conductivity reduction (vertical axis) in a system of size $2\,\mu$m. All friction length contribute to a nonnegative reduction of the bulk conductivity (i.e., all values in the graph are positive). Friction lengths, determining the length scale at which the surfaces perturb the heat flux, are contrasted with the relaxon mean free paths, which describe the length scale at which heat flux is dissipated. Since the largest friction length is about $10\,\mu$m, one can expect nanostructuring in graphene to be effective at characteristic length scales of $10\,\mu$m or below

contribution to bulk thermal conductivity from relaxon α (red dots in Fig. 7 coincide with those in Fig. 5). Because friction lengths can have any sign and since we found a symmetry such that both $\lambda_\alpha^{(y)}$ and $-\lambda_\alpha^{(y)}$ are eigenvalues of $\Lambda_{\alpha\beta}$, we only plot the positive part of the spectrum of friction lengths.

The largest friction length is at $10\,\mu m$ approximately, indicating that the reduction of thermal conductivity happens within a distance of $10\,\mu m$ from the surface. A ribbon larger than $10\,\mu m$ may well be approximated by its bulk counterpart, neglecting the contribution from the surface layers. As a result, the spectrum of eigenvalues $\lambda_i^{(y)}$, and its largest eigenvalue in particular, provides at a glance the characteristic size at which nanostructuring is effective at reducing thermal conductivity. By contrasting the friction lengths with relaxon mean free paths, we intend to underline how the two quantities are not related. Not only is their physical meaning different – decay distance of heat flux along the thermal gradient vs the decay distance of heat flux from the surface – but also the distribution of these two quantities differs and does not need to coincide.

This similarity with the Poiseuille flow is best observed in Fig. 8, by looking at the profile of thermal conductivity along the ribbon section. The thermal conductivity of the ribbon (of width $2\,\mu m$ at $300\,K$) never quite reaches the bulk values, reported in figure for comparison. This is expected from the previous analysis of Fig. 7, since there are friction lengths between 2 and $10\,\mu m$ capable of disrupting heat flux. Thermal conductivity reaches a maximum at the ribbon center, the furthest distance from surfaces, with values about 10% smaller than the bulk value. Thermal conductivity (and thus heat flux) gets smaller when approaching the surfaces, similarly to the slowing of the mass flux of a real fluid in proximity of a wall, with most of the change happening within half a micron from the surface. We note how the minimum value of thermal conductivity at the surfaces is different from zero. In fact, Casimir's boundary condition – a perfectly absorbing surface – doesn't actually coincide with a nonslip condition: this is another departure from

Fig. 8 Profile of thermal conductivity along a section of a graphene ribbon (blue line) compared with the bulk conductivity (dashed red line). In a sample of finite size, thermal conductivity is not an intrinsic property of the material but depends on the sample shape and on the distance from the surface. Thermal conductivity reaches a maximum at the center of the ribbon (the furthest distance from surfaces) and decreases to a non-zero minimum at the surfaces
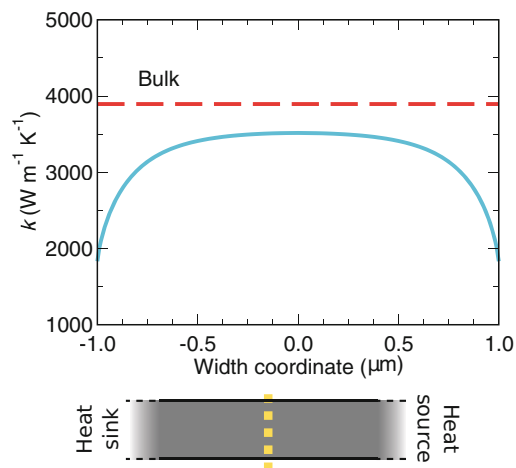
the hydrodynamics of the Navier-Stokes equations. The perfect absorption only guarantees that there won't be modes $\psi_{\alpha i}$ traveling away (rightward in the plot) from the surface at $y = -1\,\mu m$ (and vice versa for the opposite surface). This doesn't prevent modes $\psi_{\alpha i}$ traveling toward (leftward in the plot) the surface at $y = -1\,\mu m$ to conduct heat in proximity of that surface. Therefore, the model allows for a non-zero thermal conductivity at the surface; the nonslip condition can only be applied to a subset of modes $\psi_{\alpha i}$. Finally, we remark that the residual value of conductivity at the surface strongly depends on the boundary condition chosen. The case of the perfectly absorbing surface used here represents the case with the strongest reduction from the bulk value; a softer boundary condition that allows for reflection will instead increase thermal conductivity. In fact, the limiting case of a perfect mirror should leave the thermal conductivity equal to the bulk values at all points inside the ribbon.

It should be noted how thermal conductivity is not an intrinsic property of the material. Extrinsic properties now determine its value: here we have seen how the distance from the surface modifies the actual value of thermal conductivity; we have also shown in Cepellotti and Marzari (2017a) how a different shape (in that case, a trench) can also modify results, and we noted how even the orientation of crystal axis with respect to the surface affects the results and how the microscopic details of the surface modify the reflection of modes at the surface. Nevertheless, it is convenient to tag a nanostructure with an average thermal conductivity $\bar{k} = \frac{1}{W} \int k(y)\,dy$. This quantity is a good descriptor of the amount of heat flux that can travel through the ribbon at that specific size W; however, we are not allowed to use this in a different structure and cannot be used rigorously as a parameter entering heat diffusion equations. In fact, the thermal conductivity appearing in diffusion equations is an intrinsic constant property of the material and should not depend on extrinsic properties such as position. For small nanostructures therefore the correct approach is to avoid diffusion equations and use instead a microscopic model such as the LBTE.

In Fig. 9, we report the average thermal conductivity $\bar{k}$ as a function of the ribbon width at room temperature. We also contrast the results of this model with the Surf-RTA, i.e., the (incorrect) model of surface scattering where the scattering matrix is modified by adding a term $\frac{|v_{\mu}^{(y)}|}{L}\delta_{\mu\mu'}$ (similarly to Eq. 29). Results are unsurprising on the far right and far left of the graph: for small widths, thermal conductivity goes to zero, and for large widths the bulk value is recovered. The large difference appears in the region where the transition from the two limits takes place. For example, results from the two models differ by about 1000 W/mK in the range 0.1–10 μm, at length scales that are relevant for current experimental applications. Also, the Surf-RTA reaches the bulk value at a much slower rate than the exact solution. This has led a few – including us – to conclude that nanostructuring in graphene is relevant up to the millimeter length scale (or more) (Fugallo et al. 2014; Kuang et al. 2016); this result is more of an artifact of the model used. We remark that here we do not compare with experiments intentionally, the primary reason being that the geometry used here doesn't correspond to those of published experiments, i.e., Corbino disks (Balandin et al. 2008) or ribbons of comparable length and width (Xu et al. 2014).
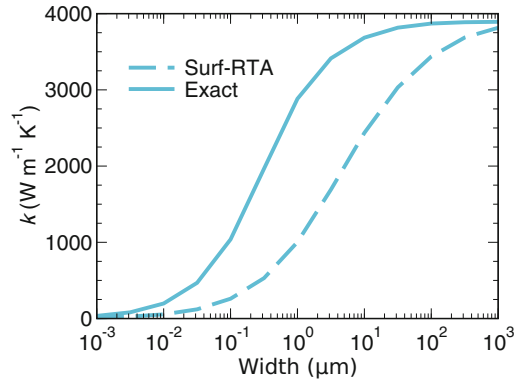
Fig. 9 Average thermal conductivity of a graphene ribbon at room temperature as a function of the ribbon width. The results of the present model (solid lines) are compared with the Surf-RTA results (dotted lines): although the qualitative behavior is similar, the two theories give widely different results. In particular, the Surf-RTA overestimates the impact of nanostructuring on the thermal conductivity
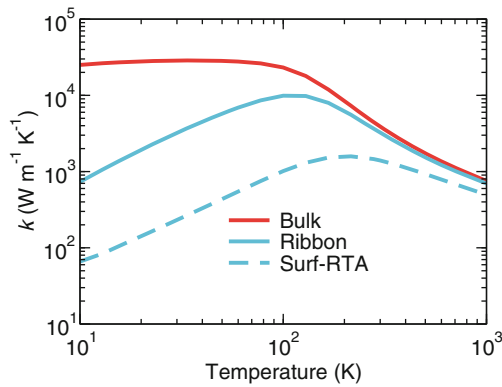
Fig. 10 Average thermal conductivity $\bar{k}$ of a graphene ribbon (width $2\,\mu m$) as a function of temperature. For comparison, the red curve shows the conductivity of an infinite sheet and the dashed blue line shows the values obtained with the approximated Surf-RTA description of surface scattering. The model discussed here gives large improvements over the Surf-RTA, in particular at low temperatures where the surface gives the dominant contribution to thermal conductivity and the estimates are revised by an order of magnitude

Finally, we fix the width at $2\,\mu m$ and show in Fig. 10 how thermal conductivity varies as a function of temperature, comparing differences between the bulk conductivity, the ribbon, and the Surf-RTA conductivities. The bulk conductivity displays two distinct regimes: above $100\,K$ the thermal conductivity is driven by three-phonon interactions that give approximately a $1/T$ decrease of thermal conductivity (Ziman 2001). At lower temperatures, the bulk conductivity is dominated by phonon-isotope scatterings. The introduction of a surface causes a decrease of

thermal conductivity, especially in the low temperature regime, and one would expect the temperature dependence at low temperature to scale as T^3, i.e., the temperature dependence of the specific heat (Ziman 2001; Carruthers 1961). At high temperatures, friction lengths become smaller, and so the thermal conductivity of the ribbon approaches the bulk limit. Although also in this case results from the present model of surface scattering and the Surf-RTA are qualitatively similar, the differences are in the magnitude of the conductivity. In particular, the two estimates differ by an order of magnitude at low temperature, where surface scattering is dominant. Even at large temperatures, we can see how the Surf-RTA never achieves bulk values and systematically underestimates thermal conductivity, whereas the friction model quickly reaches the bulk values around room temperature. As a technical note, temperatures below 10 K have not been shown because of the difficulty in converging results which require a very dense sampling of the Brillouin zone.

To conclude, we note that the regime between 10 and 100 K isn't exactly a ballistic regime. A ballistic regime requires thermal conductivity to be completely determined by surface scattering: one can approximate the scattering matrix to zero, reduce the LBTE to the diffusion operator and compute the resulting thermal conductivity for the system of finite size. Here, as noted in Cepellotti et al. (2015) and Lee et al. (2015), the regime is not ballistic but of the Poiseuille kind. This means that, although surfaces reduce thermal conductivity, the existence of intrinsic scattering events cannot be neglected. In fact, and very notably, the thermal conductivity of the ribbon of graphene in presence of a surface is larger than the ballistic estimate and highlights how scattering between phonons can enhance the heat flux with respect to a system of noninteracting phonons.

8 Dispersion Relations for Relaxons and Second Sound

The derivation of relaxons given in Sect. 4 doesn't make it obvious whether relaxons are a mathematical tool to solve the LBTE or whether they can also lead to actual excitations of the crystal. Much of this question still needs to be answered, given the early stage of research on the topic. However, we will now show that we can define a dispersion relation for collective phonon excitations starting from the LBTE, as derived in Cepellotti and Marzari (2017b). We will see that relaxons, as introduced in Sect. 4, are the long-wavelength limit of these collective phonon excitations present in the LBTE. We will also see that such excitations are closely related with oscillations of energy and of temperature in a crystal, which sometimes has been experimentally observed as second sound.

To derive the dispersion relations, we must start again from the LBTE in the phonon basis (Eq. 7). We now limit the discussion to a system at thermal equilibrium, so that the temperature of the material is constant and equal to $T(\boldsymbol{r}, t) = T$. With this simplification, the LBTE reduces to:

$$\frac{\partial \Delta n_\mu(\mathbf{x}, t)}{\partial t} + \mathbf{v}_\mu \cdot \nabla(\Delta n_\mu(\mathbf{x}, t)) = -\frac{1}{\mathscr{V}} \sum_{\mu'} \Omega_{\mu\mu'} \Delta n_{\mu'}(\mathbf{x}, t) . \tag{40}$$

While in Sect. 4 we focused exclusively on the entities that are left invariant by the action of the scattering operator, we now seek to consider solutions that are stationary to the whole LBTE. In particular, we look for collective excitations in the form of a plane wave of phonon population numbers:

$$\Delta n_\mu(\mathbf{r}, t) = \text{Re}\left(I_\mu e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}\right) , \tag{41}$$

where I_μ is the amplitude of the wave on mode μ, $\mathbf{k}$ is the wavevector, and ω is the frequency of such excitation (note the difference between ω and the phonon frequency ω_μ). To fix the value of the frequency and wavevector, we must substitute this ansatz in the LBTE – equivalent to taking the Laplace transform of the LBTE – obtaining:

$$-i\omega I_\mu + i I_\mu \mathbf{k} \cdot \mathbf{v}_\mu + \frac{1}{\mathscr{V}} \sum_{\mu'} \Omega_{\mu\mu'} I_{\mu'} = 0 . \tag{42}$$

Rearranging terms, we can cast Eq. 42 as an eigenvalue problem:

$$\sum_{\mu'} B_{\mu\mu'}(\mathbf{k}) I_{\mu'}^\alpha(\mathbf{k}) = \omega_\alpha(\mathbf{k}) I_\mu^\alpha(\mathbf{k}) , \tag{43}$$

where

$$B_{\mu\mu'} = \mathbf{k} \cdot \mathbf{v}_\mu \delta_{\mu\mu'} - \frac{i}{\mathscr{V}} \Omega_{\mu\mu'} , \tag{44}$$

with α a newly introduced eigenvalue index. For each wavevector $\mathbf{k}$, this equation identifies an eigenvalue problem that determines the allowed frequencies. In other terms, the Fourier transform of the LBTE fixes the dispersion relation of the collective phonon excitations.

To understand the connection with the relaxons introduced previously, we consider the long-wavelength limit $\mathbf{k} = 0$. In this case, the B matrix reduces (besides a phase factor) to the scattering matrix, and the equation becomes almost equivalent to the relaxon eigenvalue equation introduced in Eq. 16. The last difference to be fixed is to recast the problem in terms of symmetric matrices. To do so, we take the transformations of Eqs. 14 and 15: I and B transform like Δn and Ω, respectively, and re-express Eqs. 40, 41, 42, 43, and 44 in terms of symmetrized quantities. All quantities from now on refers to the symmetrized ones.

Looking at the mathematical aspects of the problem, the matrix B is symmetric, complex, and non-Hermitian; only in the long-wavelength limit the matrix is Hermitian. There is therefore a key difference with respect to the relaxon problem.

Here, the eigenvalues are a complex quantity and can be written as:

$$\omega_\alpha(\boldsymbol{k}) = \overline{\omega}_\alpha(\boldsymbol{k}) - \frac{i}{\tau_\alpha(\boldsymbol{k})} \ . \tag{45}$$

The real part $\overline{\omega}_\alpha(\boldsymbol{k})$ is the oscillation frequency of the collective excitation, and the imaginary part is the decay time $\tau_\alpha(\boldsymbol{k})$. This interpretation becomes apparent after rewriting the solution in terms of waves of phonon populations:

$$\Delta n_\mu^\alpha(\boldsymbol{k}) = |I_\mu^\alpha(\boldsymbol{k})| e^{-t/\tau_\alpha(\boldsymbol{k})} \sin(\boldsymbol{k} \cdot \boldsymbol{x} - \overline{\omega}_\alpha(\boldsymbol{k})t + \phi) \ . \tag{46}$$

where ϕ is a phase.

It's easy to see that each of these solutions causes an oscillation of energy (i.e., heat). Note in fact that the energy of the crystal E is changed by an oscillation Δn_μ by an amount ΔE such that:

$$E(\boldsymbol{x}, t) = \frac{1}{\mathscr{V}} \sum_\mu \left(\bar{n}_\mu + \frac{1}{2}\right) \hbar\omega_\mu + \frac{1}{\mathscr{V}} \sum_\mu \Delta n_\mu(\boldsymbol{x}, t) \hbar\omega_\mu = E_0 + \Delta E(\boldsymbol{x}, t) \ ,$$

where E_0 is the energy at thermal equilibrium. Since changes of energy and temperature are related by the specific heat with $\delta E = C\delta T$, it follows that a damped wave of Δn_μ causes a damped wave of temperature. The emergence of temperature waves in a crystal is also known as second sound, with the relation $\omega(\boldsymbol{k})$ being its dispersion relation. Improving over earlier theoretical works (Guyer and Krumhansl 1964; Chester 1963; Hardy 1970; Enz 1968), this model gives an infinite number of second sound modes (in the thermodynamic limit) where at most three had been found, the full dispersion relation (often neglected), without approximations of the LBTE (in particular, avoiding the SMA).

We noted a number of properties in our numerical examples that lack an analytic proof. In particular, the stability of the solution requires $\tau_\alpha(\boldsymbol{k}) > 0$; otherwise the time evolution will cause a divergence in the phonon population number. We also found the symmetry $\overline{\omega}_\alpha(\boldsymbol{k}) - \frac{i}{\tau_\alpha(\boldsymbol{k})} = -\overline{\omega}_\alpha(-\boldsymbol{k}) - \frac{i}{\tau_\alpha(-\boldsymbol{k})}$. The non-Hermitian character of B implies that there is a distinction between left and right eigenvectors, which complicates both proofs and the calculation of group velocities for the collective excitation, since the Feynman theorem doesn't apply anymore.

Returning to the relaxon problem studied before, this can be understood as a special case where the eigenvalues are purely a decay time, with zero oscillation frequency and zero wavevector: it is for this reason that no dispersion relation appeared in the previous discussion. Now, by looking at a more general eigenvalue problem, we constructed a set of collective excitations that encompasses relaxons, as defined before, and all other forms of wavelike excitations of the system. Given the connection between collective phonon excitations and changes of temperature, it is now more intuitive to understand why the long-wavelength limit is the most convenient for the previous study of bulk conductivity. Excitations at finite

momentum determine a change in temperature of wavelike form that are more suitable to describe nonuniform temperature perturbations; in the limit of long-wavelengths, they become a constant gradient, which is the same temperature perturbation that was used in the previous sections.

8.1   Dispersion Relations in Graphene

We still use graphene to make a numerical example of the dispersion relation introduced above (Cepellotti and Marzari 2017b). Despite keeping the same computational parameters of the previous examples (same force constants, same number of k-points, etc.), the diagonalization of the non-Hermitian matrix is much more expensive than the diagonalization of a real and symmetric matrix. Nevertheless, current computational resources are powerful enough to diagonalize the B matrix along the high-symmetry lines of graphene. However, the computational costs prevented us to conduct, for the time being, a thorough convergence study of the spectrum, unlike the simpler case of the diagonalization of the scattering matrix. Therefore, the exact value of oscillating frequencies and decay times may not be well converged with respect to q-point sampling and the broadening of the energy-conserving Dirac delta function, but this shouldn't qualitatively alter the following analysis.

In Fig. 11, we plot the decay time versus the oscillation frequency for a few wavevectors along the line $\Gamma \rightarrow M$ in the Brillouin zone, computed at a temperature
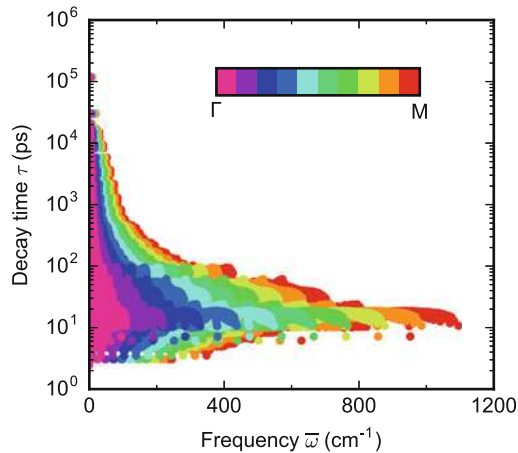
Fig. 11 Eigenvalues $\omega = \bar{\omega} - \frac{i}{\tau}$ of graphene as a function of the oscillation frequency $\bar{\omega}$ (horizontal axis) and the relaxation time τ (vertical axis); colors identify different k-points in the Brillouin zone lying between Γ and the edge M. The allowed oscillation frequencies and decay times are both bounded. The change of wavevector affects more significantly the oscillation frequency: in the limit of $k = 0$, all frequencies must vanish, and the range of allowed frequencies is approximately linear in the wavevector. (Figure from Cepellotti and Marzari 2017b)

of 300 K. The first thing to note is that each wavevector identifies N collective excitations (N being equal to the number of phonon modes of the system), and therefore there is much more information to be analyzed than before.

As discussed above, the eigenvalues of B at $\boldsymbol{k} = \Gamma$ are characterized by a zero oscillation frequency and are pure decay times: this limit reproduces the eigenvalues of relaxons introduced in Sect. 4. The oscillation frequency is approximately linear in the wavevector $\boldsymbol{k}$ and tends to grow as the distance from Γ is increased. This result suggests that in Eq. 44, the contribution to B from scattering can be considered as a perturbation with respect to the diffusive term. Last, the frequency of oscillation can be comparable, at finite wavevectors, with the frequency of phonons. At the M point, these can be as large as $1100 \, \text{cm}^{-1}$, which compares with the largest frequency of phonons (about $1600 \, \text{cm}^{-1}$). The absolute value of the oscillating frequency of transport waves (at most $1100 \, \text{cm}^{-1}$) is comparable with the oscillation frequency of phonons, which range from 0 to $1600 \, \text{cm}^{-1}$. It must be noted however that results of this model at small wavevectors should not be trusted: the semiclassical approximation used to derive the LBTE requires a temperature perturbation to be slowly varying with respect to the interatomic distances. When the wavelength of the collective excitation is comparable to the lattice parameter (i.e., $q \approx M$), the semiclassical approximation fails and the predictions based on the LBTE may not be valid. Nevertheless, the semiclassical approximation holds for $q \to 0$, and results are reliable at large wavelengths.

The decay times of collective excitations span a range of values between a few picoseconds up to about 10^5 ps. This range of values is similar to that of phonon lifetimes which, as also analyzed in Fig. 4, mostly takes place at the range of values between 10 and 100 ps. Similarly, the largest portion of collective excitations have decay times in this region between 10 and 100 ps, which forms almost a continuum of points. There are however a few eigenvalues with extremely long decay times, which are reminiscent of the discrete part of spectrum of relaxation times discussed for graphene in Sect. 4. We also note that there are no relaxation times smaller than a few picoseconds; In fact, the smallest transition rates found in the scattering matrix for this example are of a few picoseconds; it is therefore impossible for collective phonon excitations to decay faster than these lower bounds.

In Fig. 12 we plot the eigenvalue spectrum along a high-symmetry path of the Brillouin zone. Since, at a fixed frequency, more than one eigenvalue may exist, the color code is used to map the largest relaxation time found at a certain wavevector and frequency. From this picture, we observe that the largest relaxation time is approximately linear with the modulus of the wavevector $\boldsymbol{k}$, that the wave frequency is comparable with the phonon frequency, and that the frequency can be well described by a linear dispersion. The waves with the largest decay times are primarily found in the small frequency region, with high frequencies being also characterized by a faster decay. The steplike features in the colored areas are numerical artifacts, due to the fact that only 25 wavevectors are computed along the high-symmetry path.

Since it has been put forward that second sound may exist in graphene (Cepellotti et al. 2015; Lee et al. 2015), it is reasonable to speculate that the longest-lived
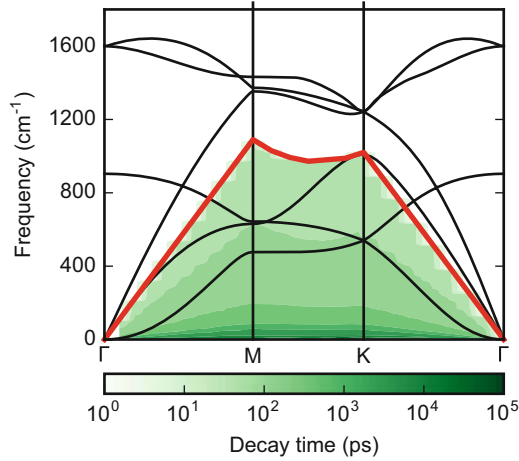
Fig. 12 Continuum spectrum of dispersion relations for temperature waves in graphene at room temperature (shades of green); the red line marks the largest value for the allowed oscillation frequency, and all frequencies lying below are valid solutions. Phonon dispersions at zero temperature are also plotted (black lines). The different shades of green indicate the decay times of the temperature waves, showing how such modes can be longer lived, especially in the low-frequency region

modes will be the most likely candidates for experimental observation. If in fact a perturbation may excite collective excitations at a certain frequency or wavevector, it is likely that a continuum of collective excitations will be perturbed: therefore it may not be simple to observe a well-defined temperature wave. After a certain amount of time from the perturbation, the only surviving modes are those with the longest relaxation times, and it would be possible to recover a wavelike behavior of temperature. We stress however that at the present time, there is still no reported experimental observation of second sound in graphene.

9   Conclusions

In this chapter, we studied extensively the LBTE; while this is done in the context of heat transport, most or all of these concepts could find application to other quantities: electrons, spins, or other quasiparticles. The single-mode relaxation time approximation has been discussed, noting that it fails at describing the correct dynamics of energy fluxes, and in particular it doesn't guarantee energy conservation. A correct description can be achieved by solving the LBTE without approximations to the scattering operator. Thanks to computational advances, it is now possible to construct direct solutions of the LBTE. To this aim, we introduced relaxons, defined as the eigenvectors of the scattering matrix and that collective

phonon excitations are characterized by well-defined relaxation times. The exact bulk thermal conductivity can be expressed exactly in terms of a kinetic gas theory, but the relevant gas of excitations are the relaxons, not the phonons; in this sense, relaxons are the correct heat carriers in a crystal.

The direct approach to solving the LBTE also allows us to improve the description of surface scattering. Whereas conventional explanations interpret the presence of a surface as a way to limit the distance traveled by the heat carrier, the exact solution of the LBTE indicates that the surface slows down the heat flux hydrodynamically. In particular, friction at the materials' borders slows down heat flux to a minimum, and the bulk limit is recovered as the distance from the surface is increased. Heat flux is found to depend on the distance from the surface, in the form of a series of exponentials, with well-defined friction lengths that characterize the length scale at which nanostructuring is an effective tool to disrupt thermal conductivity.

Eventually, we have shown how it is possible to construct a dispersion relation of collective phonon excitations, such that each mode is characterized by a wavevector, a frequency, and a decay time. The long-wavelength, zero-frequency limit of such modes is found to coincide with the definition of relaxons. We find that the excitation of these collective excitations induces a wavelike change of temperature profile in the material, which suggests a link with the theory of second sound.

The theoretical modeling has been supported with first-principles simulations of the LBTE, illustrating how the revised theory can change the time, length, and velocity scale of thermal transport, discussed here for two prototypical materials such as silicon and graphene.

Before concluding, we stress that all the results reported here hold within the validity of the LBTE, limited in particular by the semiclassical approximation. There exist a number of cases where one should expect this hypothesis to fail. For example, the LBTE only holds for small perturbations out of equilibrium, but this may not necessarily be respected in presence of strong temperature differences. Moreover, we should not expect this approximation to hold in conditions where the phonon spectral function have features so broad that one cannot identify a phonon wavepacket of well-defined energy. Under all circumstances such that the semiclassical approximation doesn't hold, none of our conclusions hold rigorously, because our starting point, the LBTE, is not justified. Whether our conclusions hold beyond the semiclassical approximations remains to be understood and constitutes an exciting outlook for further studies. Despite this criticism, we believe that the semiclassical approximation has still great relevance in today's simulations, as proved by the excellent agreement between experimental results and first-principles LBTE studies and the insights that can give to our understanding on the physics of thermal transport.

References

Abeles B (1963) Lattice thermal conductivity of disordered semiconductor alloys at high temperatures. Phys Rev 131:1906–1911

Balandin AA, Ghosh S, Bao W, Calizo I, Teweldebrhan D, Miao F, Lau CN (2008) Superior thermal conductivity of single-layer graphene. Nano Lett 8(3):902–907

Barbarino G, Melis C, Colombo L (2015) Intrinsic thermal conductivity in monolayer graphene is ultimately upper limited: a direct estimation by atomistic simulations. Phys Rev B 91:035416

Baroni S, Giannozzi P, Testa A (1987) Green's-function approach to linear response in solids. Phys Rev Lett 58:1861–1864

Baroni S, de Gironcoli S, Dal Corso A, Giannozzi P (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73:515–562

Blackford LS, Choi J, Cleary A, D'Azevedo E, Demmel J, Dhillon I, Dongarra J, Hammarling S, Henry G, Petitet A, Stanley K, Walker D, Whaley RC (1997) ScaLAPACK users' guide. Society for Industrial and Applied Mathematics, Philadelphia

Broido DA, Ward A, Mingo N (2005) Lattice thermal conductivity of silicon from empirical interatomic potentials. Phys Rev B 72:014308

Broido DA, Malorny M, Birner G, Mingo M, Stewart DA (2007) Intrinsic lattice thermal conductivity of semiconductors from first principles. Appl Phys Lett 91(23):231922

Bungaro C, de Gironcoli S, Baroni S (1996) Theory of the anomalous Rayleigh dispersion at h/w(110) surfaces. Phys Rev Lett 77:2491–2494

Cahill DG, Braun PV, Chen G, Clarke DR, Fan S, Goodson KE, Keblinski P, King WP, Mahan GD, Majumdar A, Maris HJ, Phillpot SR, Pop E, Shi L (2014) Nanoscale thermal transport. II. 2003–2012. Appl Phys Rev 1(1):011305

Carruthers P (1961) Theory of thermal conductivity of solids at low temperatures. Rev Mod Phys 33:92–138

Casimir HBG (1938) Note on the conduction of heat in crystals. Physica 5(6):495–500

Cepellotti A, Marzari N (2016) Thermal transport as a kinetic theory of relaxons. Phys Rev X 6:041013

Cepellotti A, Marzari N (2017a) Boltzmann transport in nanostructures as a friction effect. Nano Lett 17(8):4675–4682. PMID:28661694

Cepellotti A, Marzari N (2017b) Transport waves as crystal excitations. Phys Rev Mater 1:045406

Cepellotti A, Fugallo G, Paulatto L, Lazzeri M, Mauri F, Marzari N (2015) Phonon hydrodynamics in two-dimensional materials. Nat Commun 6:6400

Chaput L (2013) Direct solution to the linearized phonon Boltzmann equation. Phys Rev Lett 110:265506

Chester M (1963) Second sound in solids. Phys Rev 131:2013–2015

Debernardi A, Baroni S, Molinari E (1995) Anharmonic phonon lifetimes in semiconductors from density-functional perturbation theory. Phys Rev Lett 75:1819–1822

de Tomas C, Cantarero A, Lopeandia AF, Alvarez FX (2014) From kinetic to collective behavior in thermal transport on semiconductors and semiconductor nanostructures. J Appl Phys 115(16):164314

Donadio D, Galli G (2009) Atomistic simulations of heat transport in silicon nanowires. Phys Rev Lett 102:195901

Enz CP (1968) One-particle densities, thermal propagation, and second sound in dielectric crystals. Ann Phys 46(1):114–173

Esfarjani K, Chen G, Stokes HT (2011) Heat transport in silicon from first-principles calculations. Phys Rev B 84:085204

Fugallo G, Lazzeri M, Paulatto L, Mauri F (2013) Ab Initio variational approach for evaluating lattice thermal conductivity. Phys Rev B 88:045430

Fugallo G, Cepellotti A, Paulatto L, Lazzeri M, Marzari N, Mauri F (2014) Thermal conductivity of graphene and graphite: collective excitations and mean free paths. Nano Lett 14(11): 6109–6114

Garg J, Bonini N, Kozinsky B, Marzari N (2011) Role of disorder and anharmonicity in the thermal conductivity of silicon-germanium alloys: a first-principles study. Phys Rev Lett 106:045901

Giannozzi P, de Gironcoli S, Pavone P, Baroni S (1991) Ab Initio calculation of phonon dispersions in semiconductors. Phys Rev B 43:7231–7242

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussi C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM (2009) Quantum espresso: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502

Gill-Comeau M, Lewis LJ (2015) On the importance of collective excitations for thermal transport in graphene. Appl Phys Lett 106(19):193104

Gonze X, Vigneron J-P (1989) Density-functional approach to nonlinear-response coefficients of solids. Phys Rev B 39:13120–13128

Guyer RA, Krumhansl JA (1964) Dispersion relation for second sound in solids. Phys Rev 133:A1411–A1417

Guyer RA, Krumhansl JA (1966) Solution of the linearized phonon Boltzmann equation. Phys Rev 148:766–778

Hamilton RAH, Parrot JE (1969) Variational calculation of the thermal conductivity of germanium. Phys Rev 178:1284–1292

Hardy RJ (1963) Energy-flux operator for a lattice. Phys Rev 132:168–177

Hardy RJ (1965) Lowest-order contribution to the lattice thermal conductivity. J Math Phys 6(11):1749–1761

Hardy RJ (1970) Phonon Boltzmann equation and second sound in solids. Phys Rev B 2:1193–1207

Jain A, McGaughey AJH (2015) Strongly anisotropic in-plane thermal transport in single-layer black phosphorene. Sci Rep 5:8501

Krumhansl JA (1965) Thermal conductivity of insulating crystals in the presence of normal processes. Proc Phys Soc 85(5):921–930

Kuang Y, Lindsay L, Shi S, Wang X, Huang B (2016) Thermal conductivity of graphene mediated by strain and size. Int J Heat Mass Trans 101(Supplement C):772–778

Lazzeri M, de Gironcoli S (2002) First-principles study of the thermal expansion of be(1010). Phys Rev B 65:245402

Lee S, Lindsay L (2017) Hydrodynamic phonon drift and second sound in a (20,20) single-wall carbon nanotube. Phys Rev B 95:184304

Lee S, Broido D, Esfarjani K, Chen G (2015) Hydrodynamic phonon transport in suspended graphene. Nat Commun 6:6290

Li W, Carrete J, Katcho NA, Mingo N (2014) Shengbte: a solver of the Boltzmann transport equation for phonons. Comput Phys Commun 185(6):1747–1758

Lindsay L, Broido DA (2011) Enhanced thermal conductivity and isotope effect in single-layer hexagonal boron nitride. Phys Rev B 84:155421

Lindsay L, Broido DA, Mingo N (2010) Flexural phonons and thermal transport in graphene. Phys Rev B 82:115427

Luckyanova MN, Garg J, Esfarjani K, Jandl A, Bulsara MT, Schmidt AJ, Minnich AJ, Chen S, Dresselhaus MS, Ren Z, Fitzgerald EA, Chen G (2012) Coherent phonon heat conduction in superlattices. Science 338(6109):936–939

McGaughey AJH, Kaviany M (2004) Quantitative validation of the Boltzmann transport equation phonon thermal conductivity model under the single-mode relaxation time approximation. Phys Rev B 69:094303

Omini M, Sparavigna A (1996) Beyond the isotropic-model approximation in the theory of thermal conductivity. Phys Rev B 53:9064–9073

Omini M, Sparavigna A (1997) Heat transport in dielectric solids with diamond structure. Nuovo Cimento D 19(10):1537

Paulatto L, Mauri F, Lazzeri M (2013) Anharmonic properties from a generalized third-order ab initio approach: theory and applications to graphite and graphene. Phys Rev B 87:214303

Peierls R (1929) Zur kinetischen Theorie der Wärmeleitung in Kristallen. Ann Phys 395(8):1055–1101

Srivastava GP (1976) Derivation and calculation of complementary variational principles for the lattice thermal conductivity. J Phys C Solid State Phys 9(16):3037

Togo A, Chaput L, Tanaka I (2015) Distributions of phonon lifetimes in brillouin zones. Phys Rev B 91:094306

Vanderbilt D, Louie SG, Cohen ML (1984) Calculation of phonon-phonon interactions and the absence of two-phonon bound states in diamond. Phys Rev Lett 53:1477–1480

Volz SG, Chen G (1999) Molecular dynamics simulation of thermal conductivity of silicon nanowires. Appl Phys Lett 75(14):2056–2058

von De Haas WJ, Biermasz T (1938) Die Waermeleitfaehigkeit von Kristallen bei tiefen Temperaturen. Physica 5(4):320–324

Ward A, Broido DA, Stewart DA, Deinzer G (2009) Ab initio theory of the lattice thermal conductivity in diamond. Phys Rev B 80:125203

Xu X, Pereira LFC, Wang Y, Wu J, Zhang K, Zhao X, Bae S, Bui CT, Xie R, Thong JTL, Hong BH, Loh KP, Donadio D, Li B, Özyilmaz B (2014) Length-dependent thermal conductivity in suspended single-layer graphene. Nat Commun 5:3689

Zhang G, Li B (2005) Thermal conductivity of nanotubes revisited: effects of chirality, isotope impurity, tube length, and temperature. J Chem Phys 123(11):114714

Ziman JM (2001) Electrons and phonons: the theory of transport phenomena in solids. Oxford classic texts in the physical sciences. Oxford University Press, Oxford

Heat Transport in Insulators from Ab Initio Green-Kubo Theory

35

Stefano Baroni, Riccardo Bertossa, Loris Ercole, Federico Grasselli, and Aris Marcolongo

Contents

S. Baroni (✉)
SISSA – Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy

CNR Istituto Officina dei Materiali, Trieste, Italy
e-mail: stefano.baroni@sissa.it

R. Bertossa · L. Ercole · F. Grasselli
SISSA, Trieste, Italy

A. Marcolongo
Cognitive Computing and Computational Sciences Department, IBM Research, Zürich, Switzerland

THEOS-MARVEL, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

809

Abstract

The Green-Kubo theory of thermal transport has long been considered incompatible with modern simulation methods based on electronic-structure theory, because it is based on such concepts as energy density and current, which are ill-defined at the quantum-mechanical level. Besides, experience with classical simulations indicates that the estimate of heat-transport coefficients requires analyzing molecular trajectories that are more than one order of magnitude longer than deemed feasible using ab initio molecular dynamics. In this paper we report on recent theoretical advances that are allowing one to overcome these two obstacles. First, a general gauge invariance principle has been established, stating that thermal conductivity is insensitive to many details of the microscopic expression for the energy density and current from which it is derived, thus permitting to establish a rigorous expression for the energy flux from Density-Functional Theory, from which the conductivity can be computed in practice. Second, a novel data analysis method based on the statistical theory of time series has been proposed, which allows one to considerably reduce the simulation time required to achieve a target accuracy on the computed conductivity. These concepts are illustrated in detail, starting from a pedagogical introduction to the Green-Kubo theory of linear response and transport, and demonstrated with a few applications done with both classical and quantum-mechanical simulation methods.

1   Introduction

Heat conduction in insulators is determined by the dynamics of the atomic nuclei, the electrons following adiabatically in their ground state: a regime that we will refer to as atomic or *adiabatic heat transport*. When a quasiparticle picture of the heat carriers (phonons) is adequate, a kinetic approach to heat conduction based on the Boltzmann transport equation (Peierls 1929) has been demonstrated (Broido et al. 2007) and successfully applied to crystalline insulators (Zhou et al. 2016), leveraging phonon frequencies and lifetimes from density-functional perturbation theory (Baroni et al. 1987, 2001; Giannozzi et al. 1991; Debernardi et al. 1995). As the temperature increases, anharmonic effects eventually make the quasiparticle picture break down well below melting (Turney et al. 2009), while the Boltzmann transport equation simply does not apply to glasses and liquids, where phonon dispersions are not even defined.

The simulation of thermal transport in glasses, liquids, and high-temperature crystalline solids thus requires the full deployment of the statistical mechanics of hydrodynamic fluctuations (Kadanoff and Martin 1963), started in the 1930s by Lars Onsager (1931a; 1931b) and culminated in the 1950s with the Green-Kubo (GK) theory of linear response (Green 1952, 1954; Kubo 1957; Kubo et al. 1957). According to this theory, the thermal conductivity, κ, can be expressed in terms of time correlations of the heat current, J, as:

$$\kappa \propto \int_0^\infty \langle J(t) J(0) \rangle \, dt, \tag{1}$$

where the brackets indicate ensemble averages over trajectories, which are accessible in principle to molecular dynamics (MD) simulations. In spite of the beauty, rigor, and broad scope of the GK theory, its adoption in the ab initio simulation of heat transport has stumbled at two seemingly insurmountable hurdles: first and foremost, the heat current entering Eq. (1) is ill-defined at the atomic scale, thus allegedly thwarting its implementation from first principles; second but not less important, experience in classical simulations, where a definition of the heat current is readily available in terms of suitably defined atomic energies (Irving and Kirkwood 1950), indicates that taming its statistical fluctuations requires so long MD simulations, as to be unaffordable using ab initio (AI) MD (Car and Parrinello 1985; Marx and Hutter 2009). As a matter of fact, no AIMD simulations of adiabatic heat transport based on GK theory have appeared until the spurious nature of these hurdles was recently revealed by Marcolongo et al. (2016) and Ercole et al. (2016, 2017).

The first difficulty was overcome by revealing a general *gauge invariance* principle for thermal transport, stemming from the hydrodynamic nature of energy fluctuations (see below), which makes transport coefficients independent of the microscopic expression of the energy density and current that are to a large extent ill-defined (Marcolongo et al. 2016; Ercole et al. 2016). Building on this principle, an expression for the energy current was derived from density-functional theory (Hohenberg and Kohn 1964; Kohn and Sham 1965), which allows simulating heat transport within the GK formalism, using equilibrium AIMD. The second difficulty was addressed using advanced statistical methods for the spectral analysis of stationary time series, which provide an asymptotically unbiased and consistent estimate of the power spectrum of the energy current, meaning that both the bias and the statistical error of the estimated conductivity can be made arbitrarily and controllably small in the limit of long simulation time (Ercole et al. 2017). The combination of these theoretical and methodological innovations makes the ab initio simulation of heat transport feasible in principle and affordable in practice. In this chapter, we review the efforts that have led to these achievements, starting with a brief review of the relevant theoretical concepts and illustrate them with relevant benchmarks and an application to a realistic model of liquid water.

2   Green-Kubo Theory of Heat Transport

2.1   Hydrodynamic Variables

The macroscopic processes occurring in condensed matter are often described in terms of *extensive variables*. By definition, the value that such a variable assumes for a system is the sum of the values it has for each of its subsystems. This property allows one to express an extensive variable, A, as the integral of a suitably defined density, $a(\mathbf{r})$, as:

$$A[\Omega] = \int_{\Omega} a(\mathbf{r}) d\mathbf{r}, \tag{2}$$

where Ω is the system volume. Here and in the following boldfaces indicate 3D vectors and Greek subscripts label Cartesian components: $\mathbf{u} = \{u_\alpha\} = \{u_1, u_2, u_3\}$. When an extensive quantity is locally conserved, a current density, $\boldsymbol{j}(\mathbf{r}, t)$, can be associated with its density in such a way that the two of them satisfy the continuity equation:

$$\frac{\partial a(\mathbf{r}, t)}{\partial t} = -\nabla \cdot \boldsymbol{j}(\mathbf{r}, t), \tag{3}$$

where $\nabla \cdot \boldsymbol{j}$ indicates partial differentiation and the middle dot a scalar product (a divergence in this case). In the following, the densities and current densities of conserved quantities will be called *conserved densities* and *conserved currents* for short. The space Fourier transform of Eq. (3) reads:

$$\dot{\tilde{a}}(\mathbf{q}, t) = -i\mathbf{q} \cdot \tilde{\boldsymbol{j}}(\mathbf{q}, t), \tag{4}$$

where the overdot indicates a time derivative and the tilde a Fourier transform, so that the longer the wavelength, the slower is the dynamics of a conserved density. We conclude that for long-enough wavelengths, conserved densities are adiabatically decoupled from all the other (zillions of) fast atomic degrees of freedom. Note that in this chapter, we are using the concept of *adiabatic decoupling* in two distinct senses, depending on the context, to indicate the decoupling of electronic from nuclear degrees of freedom and that of hydrodynamic variables from fast atomic ones.

The long-wavelength Fourier components of conserved densities are called *hydrodynamic variables*. In macroscopically homogeneous systems, different wavelengths are decoupled from each other, while, as we have seen, the long wavelengths are adiabatically decoupled from all the other degrees of freedom. Let us suppose there are Q conserved extensive variables. In the case of a mono-atomic fluid, for instance, $Q = 5$, corresponding to mass (or particle number), energy, and the three components of the momentum. In order to simplify the notation, we set the value of the conserved quantities equal to zero, $A^i = 0$, so that their densities, $a^i(\mathbf{r})$, directly refer to the departure from equilibrium, and we indicate by $\boldsymbol{j}^i(\mathbf{r}, t)$ the corresponding currents. At equilibrium, all the conserved densities and currents vanish. Off equilibrium, it will be assumed that the wavelength and the time scale of the disturbances are so long that thermal equilibrium still holds *locally*. That is to say, a local temperature, pressure, and chemical potential can be defined, such that, when combined with the densities of extensive variable, they satisfy a local equation of state.

For small-enough deviations from equilibrium, the time derivatives of conserved densities are linear combinations of the densities themselves. In the frequency/wavevector domains, this condition can be expressed as

$$-i\omega \tilde{a}^i(\mathbf{q}, \omega) = \sum_j \tilde{\Lambda}^{ij}(\mathbf{q}, \omega)\tilde{a}^j(\mathbf{q}, \omega), \tag{5}$$

where the tilde indicates now a space-time Fourier transform: $\tilde{a}(\mathbf{q}, \omega) = \int e^{-i(\mathbf{q}\cdot\mathbf{r}-\omega t)} a(\mathbf{r}, t) d\mathbf{r} dt$. By combining Eq. (5) with the time Fourier transform of Eq. (4), we obtain the so-called constitutive equations for the (longitudinal components of the) conserved currents:

$$\tilde{j}^i(\mathbf{q}, \omega) = i \frac{\mathbf{q}}{q^2} \sum_j \tilde{\Lambda}^{ij}(\mathbf{q}, \omega) \tilde{a}^j(\mathbf{q}, \omega). \tag{6}$$

In isotropic media, the $\tilde{\Lambda}$s are spherically symmetric functions of $\mathbf{q}$, whereas their value at $\mathbf{q} = 0$ vanishes, because a nonvanishing value would imply a nonphysical long-range dependence of the currents on density fluctuations, in contrast with our assumption of local thermodynamic equilibrium. The long-wavelength low-frequency limit of the coupling constants can thus be assumed to be $\tilde{\Lambda}^{ij}(\mathbf{q}, \omega) \sim q^2 \lambda^{ij}$, so that the macroscopic ($\mathbf{q} = 0$) stationary ($\omega = 0$) components of the currents, $\mathbf{J}^i = \frac{1}{\Omega} \int j^i(\mathbf{r}) d\mathbf{r}$, are related to the corresponding components of the density gradients, $\mathbf{D}^i = \frac{1}{\Omega} \int \nabla a^i(\mathbf{r}) d\mathbf{r}$, through the equation:

$$\mathbf{J}^i = \sum_j \lambda^{ij} \mathbf{D}^j. \tag{7}$$

In the following, the macroscopic component of a current will be indicated as a *flux*.

Let $x^i = \frac{\partial S}{\partial A^i}$ be the intensive variable conjugate to A^i, where S is the system's entropy and $\chi^{ij} = \frac{1}{\Omega} \frac{\partial A^i}{\partial x^j}$ the corresponding susceptibility. For instance, when A^i is the energy of the system, the corresponding conjugate variable is the inverse temperature, $x^i = 1/T$, while, when A^i represents the number of particles of a given species, one has $x^i = -\mu^i/T$, μ^i being the corresponding chemical potential. The hypothesis of local thermodynamic equilibrium allows defining local values of the intensive variables, and we define *thermodynamic forces* as their average gradients: $\mathbf{F}^i = \frac{1}{\Omega} \int \nabla x^i(\mathbf{r}) d\mathbf{r}$. The average density gradients are related to the thermodynamic forces through the susceptibility defined above, as:

$$\mathbf{D}^i = \sum_j \chi^{ij} \mathbf{F}^j. \tag{8}$$

By inserting this relation into Eq. (7), one gets:

$$\mathbf{J}^i = \sum_j L^{ij} \mathbf{F}^j, \tag{9}$$

where $L^{ij} = \sum_k \lambda^{ik} \chi^{kj}$. Equation (9) expresses the linear relation between fluxes, the $\mathbf{J}$'s, and thermodynamic affinities, the $\mathbf{F}$s, for which Onsager derived his celebrated reciprocity relations ($L^{ji} = L^{ij}$) from microscopic reversibility

(Onsager 1931a, b; Casimir 1945). Note that, according to our definition, both the **J**s and the **F**s in Eq. (9) do not depend on the size of the system.

2.2   Linear-Response Theory

In order to evaluate the L^{ij} phenomenological coefficients appearing in Eq. (9), we consider a classical system of N interacting atoms described by the Hamiltonian

$$H^\circ(\Gamma) = \sum_n \frac{1}{2M_n}(\mathbf{P}_n)^2 + V(\mathbf{R}_1, \mathbf{R}_2, \cdots \mathbf{R}_N),\qquad(10)$$

where M_n, $\mathbf{R}_n$, and $\mathbf{P}_n$ are the masses, coordinates, and momenta of the n-th particle, $\Gamma = \{\mathbf{R}_n, \mathbf{P}_n\}$ indicates the phase-space coordinates of the entire system, and V is a generic many-body potential. Let us now suppose that the system is subject to an external perturbation that can be described as a linear combination of the conserved densities, $\{a^i(\mathbf{r}; \Gamma)\}$, as:

$$V'(\Gamma, t) = \sum_i \int v^i(\mathbf{r}, t)a^i(\mathbf{r}; \Gamma)d\mathbf{r},\qquad(11)$$

where $a(\mathbf{r}; \Gamma)$ is a phase-space function whose ensemble average is the conserved density:

$$\begin{aligned} a(\mathbf{r}) &= \langle a(\mathbf{r}; \Gamma)\rangle \\ &= \int a(\mathbf{r}; \Gamma)\mathcal{P}^\circ(\Gamma)d\Gamma, \end{aligned}\qquad(12)$$

$\mathcal{P}^\circ(\Gamma) \propto e^{-\frac{H^\circ(\Gamma)}{k_B T}}$ is the equilibrium distribution, k_B the Boltzmann constant, and $\{v^i(\mathbf{r}, t)\}$ are time-dependent fields that couple to the conserved densities and vanish at $t = -\infty$, when the system is assumed to be in thermal equilibrium at some temperature T. Of course, conserved currents are also expected values of some phase-space functions, $\mathbf{j}(\mathbf{r}) = \langle \mathbf{j}(\mathbf{r}; \Gamma)\rangle$. The phase-space functions whose expected values are conserved densities/currents will be referred to as *phase-space samples* of the currents/densities. In the following, when the phase-space dependence of a conserved density/current is explicitly indicated, we will mean a phase-space sample; when it is not, a phase-space average will be implied. When a phase-space sample is evaluated along a dynamical trajectory, Γ_t, the sample function will depend on time and on the initial conditions of the trajectory. Averaging with respect to the initial conditions will result in a time-dependent expected value for the conserved densities (or currents):

$$\begin{aligned} a(\mathbf{r}, t) &= \langle a(\mathbf{r}; \Gamma'_t)\rangle_0 \\ &= \int a(\mathbf{r}; \Gamma'_t)\mathcal{P}^\circ(\Gamma_0)d\Gamma_0. \end{aligned}\qquad(13)$$

In Eq. (13), the notation Γ'_t denotes somewhat pedantically that the time evolution in phase space is driven by the perturbed Hamiltonian, $H° + V'$. If it were driven by $H°$, evidently the value of a would be time-independent. In the following, the notation Γ_t will indicate an unperturbed time evolution. As an example, the phase-space sample of the particle density can be assumed to be $n(\mathbf{r}; \Gamma) = \sum_n \delta(\mathbf{r} - \mathbf{R}_n)$, the corresponding current is $\mathbf{j}(\mathbf{r}, \Gamma) = \sum_n \delta(\mathbf{r} - \mathbf{R}_n)\mathbf{P}_n/M_n$, and a local external potential is described by $V'(\Gamma, t) = \sum_n v(\mathbf{R}_n, t) = \int v'(\mathbf{r}, t)n(\mathbf{r}; \Gamma)d\mathbf{r}$. Note that sample functions are not necessarily univocally defined. Different functions whose phase-space averages coincide in the long-wavelength limit sample the same hydrodynamical variable. More on this in Sect. 3.

According to Green (1954), Kubo (1957), and Kubo et al. (1957), the linear response of the i-th conserved current to the perturbation is:

$$j^i_\alpha(\mathbf{r}, t) = \frac{1}{k_B T} \sum_j \int_{-\infty}^t dt' \int d\mathbf{r}' \left\langle j^i_\alpha(\mathbf{r}, \Gamma_t)\dot{a}^j(\mathbf{r}', \Gamma_{t'})\right\rangle_0 v^j(\mathbf{r}', t'), \tag{14}$$

$$= \frac{-1}{k_B T} \sum_{j,\beta} \int_{-\infty}^t dt' \int d\mathbf{r}' \left\langle j^i_\alpha(\mathbf{r}, \Gamma_t)\partial'_\beta j^j_\beta(\mathbf{r}', \Gamma_{t'})\right\rangle_0 v^j(\mathbf{r}', t') \tag{15}$$

$$= \frac{1}{k_B T} \sum_{j,\beta} \int_{-\infty}^t dt' \int d\mathbf{r}' \left\langle j^i_\alpha(\mathbf{r}, \Gamma_t) j^j_\beta(\mathbf{r}', \Gamma_{t'})\right\rangle_0 \partial'_\beta v^j(\mathbf{r}', t'). \tag{16}$$

The second line follows from the first through the continuity equation, Eq. (3), while the third line follows after integrating by parts with respect to $\mathbf{r}'$. The notation $\partial'_\beta = \frac{\partial}{\partial r'_\beta}$ has been used.

By integrating Eq. (16) all over the space, and assuming space-time homogeneity as well as isotropy, one recovers Eq. (9) with:

$$J^i_\alpha(\Gamma) = \frac{1}{\Omega} \int j^i_\alpha(\mathbf{r}, \Gamma)d\mathbf{r} \tag{17}$$

$$F^i_\alpha(\Gamma) = \frac{1}{\Omega T} \int \partial_\alpha v^i(\mathbf{r}, \Gamma)d\mathbf{r} \tag{18}$$

$$L^{ij}_{\alpha\beta} = \frac{\Omega}{k_B} \int_0^\infty \left\langle J^i_\alpha(\Gamma_t) J^j_\beta(\Gamma_0)\right\rangle_0 dt. \tag{19}$$

This completes the derivation of the Green-Kubo formula for transport coefficients, Eq. (1), from classical linear-response theory. Onsager's reciprocity relations, $L^{ij} = L^{ji}$ (Onsager 1931a, b), follow from Eq. (19) leveraging time-translational invariance, $\langle J^i_\alpha(\Gamma_t)J^j_\beta(\Gamma_0)\rangle = \langle J^i_\alpha(\Gamma_0)J^j_\beta(\Gamma_{-t})\rangle$, and micro-reversibility, $\langle J^i_\alpha(\Gamma_t)J^j_\beta(\Gamma_0)\rangle = \langle J^i_\alpha(\Gamma_{-t})J^j_\beta(\Gamma_0)\rangle$.

2.2.1 Einstein-Helfand Expression for Transport Coefficients and the Wiener-Khintchine Theorem

The celebrated Einstein's relation between the mean-square displacement of a diffusing particle and its velocity autocorrelation function is easily generalized to an arbitrary stochastic process and has in fact been utilized by Helfand (1960) to provide an "Einstein-like" expression for transport coefficients.

Let X_t be a stationary stochastic process. One has:

$$\frac{1}{\mathcal{T}} \left\langle \left| \int_0^{\mathcal{T}} X_t dt \right|^2 \right\rangle = 2 \int_0^{\mathcal{T}} \langle X_t X_0 \rangle \, dt - \frac{2}{\mathcal{T}} \int_0^{\mathcal{T}} \langle X_t X_0 \rangle \, t \, dt. \tag{20}$$

In the large $\mathcal{T}$ limit, the second term on the right-hand side of Eq. (20) can be neglected.

When the stochastic process is the velocity of a Brownian particle, Eq. (20) allows one to establish a relation between the diffusion constant of the particle, temperature, and the autocorrelation time of the velocity. When X_t is the heat flux of a macroscopic body, Eq. (20) allows one to estimate the thermal conductivity, as given by Eq. (1), from the asymptotic behavior of the "energy displacement" $\mathcal{D}(\tau) = \int_0^{\tau} \mathbf{J}(\Gamma_t) dt$.

Equation (20) can be easily generalized to the finite-frequency regime, to get:

$$\begin{aligned} S_{\mathcal{T}}(\omega) &= \frac{1}{\mathcal{T}} \left\langle \left| \int_0^{\mathcal{T}} X_t e^{i\omega t} dt \right|^2 \right\rangle \\ &= 2\mathfrak{Re} \int_0^{\mathcal{T}} \langle X_t X_0 \rangle \, e^{i\omega t} dt + \mathcal{O}(\mathcal{T}^{-1}). \end{aligned} \tag{21}$$

This equation expresses the Wiener-Khintchine theorem (Wiener 1930; Khintchine 1934), which states that the expectation of the squared modulus of the Fourier transform of a stationary process is the Fourier transform of its time correlation function, which is usually referred to as the process *power spectral density*:

$$S(\omega) = \int_{-\infty}^{\infty} \langle X_t X_0 \rangle \, e^{i\omega t} dt, \tag{22}$$

aka the *power spectrum*. In the following, the subscript $\mathcal{T}$ will be neglected for simplicity and its value assumed to be sufficiently large as to be considered infinite. More generally, when several conserved currents interact with each other, one can define the *cross-spectrum* of the conserved fluxes as the Fourier transform of the cross-time correlation functions:

$$S^{kl}(\omega) = \int_{-\infty}^{\infty} \langle X_t^k X_0^l \rangle \, e^{i\omega t} dt,$$

$$= \frac{1}{\mathcal{T}} \Re\mathfrak{e} \left\langle \int_0^{\mathcal{T}} X_t^k e^{-i\omega t} dt \times \int_0^{\mathcal{T}} X_t^l e^{i\omega t} dt \right\rangle + \mathcal{O}(\mathcal{T}^{-1}) \tag{23}$$

Eqs. (20) and (21) indicate that the transport coefficients we are after essentially are the zero-frequency value of the (cross-)power spectrum of the corresponding current(s), a fact that will be instrumental in our approach to data analysis, as explained in Sect. 5. Therefore, Eq. (19) can be cast into the form:

$$L^{kl} = \frac{\Omega}{2k_B} S^{kl}(\omega = 0), \tag{24}$$

where the Cartesian indices have been omitted for clarity.

2.3 Heat Transport

The above treatment allows one to compute the linear response of a system at thermal equilibrium to a generic mechanical perturbation. Heat transport is determined by temperature gradients that cannot be described by any mechanical perturbation. The concept of temperature distribution implies that the system is locally at thermal equilibrium over lengths and times large with respect to atomic distances and relaxation times. Temperature affects the physical properties of a system through the Boltzmann distribution function. When the temperature is not constant, $T(\mathbf{r}) = T + \Delta T(\mathbf{r})$ ($|\Delta T| \ll T$), the effects of this inhomogeneity can be formally described by the distribution function:

$$\mathcal{P}(\Gamma) \propto e^{-\int \frac{e(\mathbf{r};\Gamma)}{k_B T(\mathbf{r})} d\mathbf{r}}, \tag{25}$$

$$= e^{-\frac{H^\circ(\Gamma)+V'(\Gamma)}{k_B T}}, \tag{26}$$

where $e(\mathbf{r}; \Gamma)$ is an energy (Hamiltonian) density, such that $\int e(\mathbf{r}; \Gamma) d\mathbf{r} = H^\circ(\Gamma)$. Equation (11) becomes:

$$V'(\Gamma) = -\frac{1}{T} \int \Delta T(\mathbf{r}) e(\mathbf{r}; \Gamma) d\mathbf{r} + \mathcal{O}(\Delta T^2). \tag{27}$$

Equation (27) shows that the effects of temperature inhomogeneities can be mimicked by a mechanical perturbation coupled to the temperature distribution. From Eqs. (9) and (17), (18), and (19), we conclude that in a system where the only nontrivial conserved quantity is the energy, the heat (energy) flow is coupled to temperature gradients through the constitutive equation:

$$\mathbf{J}^E = -\kappa \nabla T, \tag{28}$$

where the thermal conductivity $\kappa_{\alpha\beta} = L_{\alpha\beta}^{EE}/T^2$ (see Eq. (9)) can be expressed by a Green-Kubo relation in terms of the fluctuations of the energy flux as:

$$\kappa_{\alpha\beta} = \frac{\Omega}{k_B T^2} \int_0^\infty \left\langle J_\alpha^E(\Gamma_t) J_\beta^E(\Gamma_0) \right\rangle_0 dt, \tag{29}$$

and

$$\mathbf{J}^E(\Gamma) = \frac{1}{\Omega} \int \boldsymbol{j}^E(\mathbf{r}; \Gamma) d\mathbf{r}. \tag{30}$$

In order to obtain an explicit expression for the energy flux from a microscopic expression for the energy density, we multiply the continuity equation, Eq. (3), by $\mathbf{r}$ and integrate by parts to obtain:

$$\mathbf{J}^E(\Gamma_t) = \frac{1}{\Omega} \int \dot{e}(\mathbf{r}; \Gamma_t) \, \mathbf{r} \, d\mathbf{r} \tag{31}$$

$$= \frac{1}{\Omega} \int \left[\sum_n \left(\frac{\partial e(\mathbf{r}; \Gamma_t)}{\partial \mathbf{R}_n} \cdot \mathbf{V}_n + \frac{\partial e(\mathbf{r}; \Gamma_t)}{\partial \mathbf{P}_n} \cdot \mathbf{F}_n \right) \right] \mathbf{r} \, d\mathbf{r}, \tag{32}$$

where $\mathbf{F}_n$ is the force acting on the n-th atom and $\mathbf{V}_n = \frac{\mathbf{P}_n}{M_n}$ its velocity.

The manipulations leading from the continuity equation, Eq. (3), to Eq. (32) deserve some further comments, as they imply neglecting a boundary term, $\mathbf{J}_{\partial\Omega} = \frac{1}{\Omega} \int_{\partial\Omega} \left(\boldsymbol{j}(\mathbf{r}) \cdot \hat{\mathbf{n}} \right) \mathbf{r} \, d\mathbf{r}$ (where $\partial\Omega$ is the boundary of the integration volume and $\hat{\mathbf{n}}$ the normal to it), which in general does not vanish in the thermodynamic limit and is ill-defined in periodic boundary conditions (PBC). The correct way of addressing this problem is to work with the Taylor expansion of the space Fourier transform of the continuity equation, Eq. (4), and to perform the thermodynamic limit at finite wavelength. The leading nonvanishing term in the Taylor expansion yields Eq. (31) without any boundary term in the way.

2.3.1 Energy Flux from Classical Force Fields

When atoms interact through a classical force field, $V(\mathbf{R}_1, \mathbf{R}_2, \cdots \mathbf{R}_N)$, an energy density can be defined in terms of local atomic energies as:

$$e(\mathbf{r}, \Gamma) = \sum_n \delta(\mathbf{r} - \mathbf{R}_n) e_n(\Gamma), \tag{33}$$

$$e_n(\Gamma) = \frac{(\mathbf{P}_n)^2}{2M_n} + v_n(\{\mathbf{R}\}), \tag{34}$$

where the v_ns are a set of atomic potential energies whose sum is the total potential energy of the system, $\sum_n v_n = V$, with a short-range dependence on the coordinates of the other atoms. In the presence of long-range forces, this condition is effectively guaranteed by local charge neutrality, which we will assume throughout. By inserting Eq. (33) into Eq. (32), the energy flux can be cast into the form:

$$
\mathbf{J}^E(\Gamma) = \frac{1}{\Omega} \left[\sum_n \mathbf{V}_n e_n + \sum_n \mathbf{R}_n \left(\mathbf{F}_n \cdot \mathbf{V}_n + \sum_m \mathbf{V}_m \cdot \frac{\partial v_n}{\partial \mathbf{R}_m} \right) \right]
$$

$$
= \frac{1}{\Omega} \left[\sum_n \mathbf{V}_n e_n + \sum_{n,m} (\mathbf{R}_n - \mathbf{R}_m) \mathbf{F}_{nm} \cdot \mathbf{V}_n \right], \tag{35}
$$

where $\mathbf{F}_{nm} = -\frac{\partial v_m}{\partial \mathbf{R}_n}$ is the contribution of the m-th atom to the force acting on the n-th atom, $\sum_m \mathbf{F}_{nm} = \mathbf{F}_n$, and $\mathbf{F}_{nm} = -\mathbf{F}_{mn}$. When the interaction among atoms can be expressed in terms of two-body potentials, one has $v_m = \frac{1}{2} \sum_n v(\mathbf{R}_n - \mathbf{R}_m)$ and $\mathbf{F}_{nm} = -\frac{1}{2} \nabla_{\mathbf{R}_n} v(\mathbf{R}_n - \mathbf{R}_m)$. Here we implicitly assumed that the interaction energy is equally partitioned between atoms m and n. In Sect. 3, we shall see this is not the only possible choice, with far-reaching consequences on the theory of heat transport.

The first term on the right-hand side of Eq. (35) is often called *convective* and the second *virial*. We feel that the wording "convective" is somewhat misleading in this context, as the convective current, as well as its contribution to heat conductivity, may not vanish even in the absence of convection.

2.3.2   Multicomponent Fluids

In a multicomponent fluid, there is one conserved quantity (the particle number) per atomic species, plus the total energy and the three Cartesian components of the total momentum. The momentum densities are mass currents: the mass flux is therefore the total momentum, which vanishes in the center of mass reference frame. The transverse components of the momentum densities are decoupled from the other conserved densities (Foster 1975), while the longitudinal one can be assumed to coincide with the total momentum in the long-wavelength limit. Momentum conservation thus constrains the number of fluxes interacting with the energy flux in Eq. (9) to $Q - 1$, Q being the number of atomic species, so that the resulting dimension of the matrix of Onsager coefficients, L, is $Q \times Q$. The heat flux is defined as the non-convective component of the energy flux, i.e., the value of the latter in the absence of mass transport, that is to say when all the particle fluxes vanish. It is unfortunate, but inevitable due to common usage, that this definition of non-convective flux clashes with a different definition given above while commenting Eq. (35). By imposing this condition in Eq. (9), with $\mathbf{J}^1 \equiv \mathbf{J}^E$, and $\mathbf{J}^q$ ($q = 2, \dots Q$) being independent particle fluxes, the thermal conductivity, defined as the ratio of the heat flux over the temperature gradient, is given by:

$$\kappa = \frac{1}{T^2 \, (L^{-1})^{11}}. \tag{36}$$

This expression can be proved to be invariant under *any* non-singular linear transformation of the independent particle fluxes. For instance, in the case of a two-component liquid, energy and particle currents are coupled as in:

$$\mathbf{J}^E = L^{EE} \, \nabla \left(\frac{1}{T} \right) + L^{EQ} \, \nabla \left(\frac{\mu}{T} \right),$$

$$\mathbf{J}^Q = L^{EQ} \, \nabla \left(\frac{1}{T} \right) + L^{QQ} \, \nabla \left(\frac{\mu}{T} \right), \tag{37}$$

where $\mathbf{J}^Q$ is the particle current of one of the two species (say, the second) and μ the corresponding chemical potential (Sindzingre and Gillan 1990). By imposing that the particle current vanishes, the resulting thermal conductivity is:

$$\kappa = \frac{1}{T^2} \left(L^{EE} - \frac{(L^{EQ})^2}{L^{QQ}} \right). \tag{38}$$

3   Gauge Invariance of Heat Transport Coefficients

It is often implicitly assumed that the well-definiteness of thermal transport coefficients would stem from the uniqueness of the decomposition of the system's total energy into localized, atomic, contributions. This assumption is manifestly incorrect, as any decomposition leading to the same value for the total energy as Eq. (34) should be considered as legitimate. The difficulty of partitioning a system's energy into subsystems' contributions is illustrated in Fig. 1, which depicts a system made of two interacting subsystems. When defining the energy of each of the two subsystems, an arbitrary decision has to be made as to how the interaction energy is partitioned. In the case depicted in Fig. 1, for instance, the energy of each of the two subsystems can be defined as $\mathcal{E}(\Omega_i) = E(\Omega_i) + \frac{1}{2}(1 \pm \lambda) W_{12}$, where $E(\Omega_i)$ are the energies of the two isolated subsystems, W_{12} their interaction energy, and λ an arbitrary constant. In the thermodynamic limit, when all the subsystems' energies are much larger than the interaction between any pairs of them, the value of the λ constant is irrelevant. When it comes to defining energy densities (i.e., energies of infinitesimal portions of a system) or atomic energies, instead, the magnitude of the interaction between different subsystems is comparable to their energies, which become therefore intrinsically ill-defined.

Let us consider a mono-atomic fluid interacting through pair potentials, $v(|\mathbf{R}_n - \mathbf{R}_m|)$, and define the atomic energies as (Marcolongo 2014; Ercole et al. 2016):
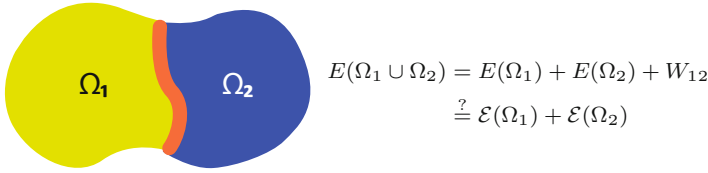
Fig. 1 The energy of an isolated system is the sum of the energies of its subsystems (as defined when they are isolated as well) plus the interaction among them, W_{12}, whose magnitude scales as the area of the interface, depicted in red. When defining the energies of individual subsystems, $\mathcal{E}$, W_{12} has to be arbitrarily partitioned among them

$$e_{\gamma,n}(\Gamma) = \frac{1}{2M_n}(\mathbf{P}_n)^2 + \frac{1}{2}\sum_{m\neq n} v(|\mathbf{R}_n - \mathbf{R}_m|)(1 + \gamma_{nm}), \tag{39}$$

where $\gamma_{nm} = -\gamma_{mn}$ is *any* antisymmetric matrix. As the interatomic potential appearing in Eq. (39) is symmetric with respect to the atomic indices, it is clear that the sum of all the atomic energies does not depend on γ, thus making any choice of γ equally permissible. This trivial observation has deep consequences on the theory of thermal fluctuations and transport, because the value of the macroscopic energy flux, instead, depends explicitly on γ, thus making one fear that the resulting transport coefficients would depend on γ as well. Using the same manipulations that lead from Eqs. (33) and (34) to Eq. (35), for any choice of the γ matrix in Eq. (39), a corresponding expression for the macroscopic energy flux can be found, reading (Marcolongo 2014; Ercole et al. 2016):

$$\mathbf{J}_\gamma^E = \mathbf{J}^E + \frac{1}{2\Omega}\sum_{n,m\neq n} \gamma_{nm}\Big(v_{nm}\mathbf{V}_n + \big(\mathbf{V}_n \cdot \nabla_{\mathbf{R}_n}v_{nm}\big)(\mathbf{R}_n - \mathbf{R}_m)\Big), \tag{40}$$

where $v_{nm} = v(|\mathbf{R}_n - \mathbf{R}_m|)$.

As a specific example, Ercole et al. (2016) ran MD simulations for a Lennard-Jones monoatomic fluid described by the interatomic potential $v(r) = \epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right]$ at temperature $T = 1.86\frac{\epsilon}{k_B}$ and density $\rho = 0.925\sigma^{-3}$. In Fig. 2a, we display the resulting macroscopic energy-flux autocorrelation function corresponding to different choices of the γ matrix in Eqs. (39) and (40). Figure 2a clearly shows that the $\langle \mathbf{J}_\gamma^E(t) \cdot \mathbf{J}_\gamma^E(0)\rangle$ correlation functions dramatically depend on the γ matrices in Eqs. (39) and (40). Notwithstanding, the integrals of all these time correlation functions tend to the same limit at large integration times, as shown in Fig. 2b.

In order to get insight into this remarkable invariance property, let us inspect the difference between the generalized flux in Eq. (40) and the standard expression of Eq. (35):

$$\Delta\mathbf{J}_\gamma^E = \mathbf{J}_\gamma^E - \mathbf{J}^E = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{4\Omega}\sum_{n,m\neq n} \gamma_{nm}\, v(|\mathbf{R}_n - \mathbf{R}_m|)(\mathbf{R}_n - \mathbf{R}_m). \tag{41}$$
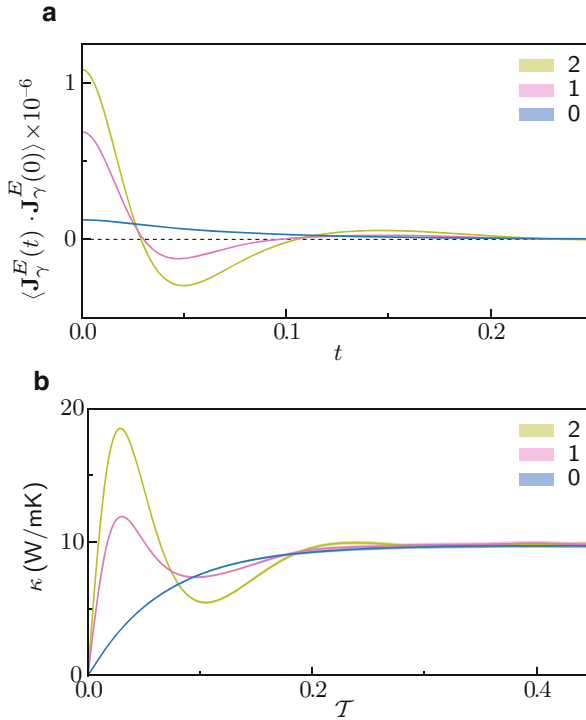
Fig. 2 (**a**) Time correlation functions of the modified macroscopic energy flux of a Lennard-Jones fluid, at the conditions described in the text, as defined in Eq. (40), for different definitions of the γ matrix. The "0" line refers to the standard definition ($\gamma = 0$), whereas the labels "1" and "2" correspond to two other (arbitrary) definitions of γ as described in Ercole et al. (2016). (**b**) Integral of the time correlation functions displayed in Fig. 2a, multiplied by the prefactor appearing in the GK relation, Eq. (29), as a function of the upper limit of integration. The barely visible shaded area surrounding each line is an indication of the error bars, as estimated by standard block analysis. Units are Lennard-Jones units ($M = \sigma = \varepsilon = 1$)

We see that the two different expressions for the macroscopic energy flux differ by a total time derivative of a bounded phase-space vector function. In the following, we show that this is a consequence of energy conservation and extensivity and a sufficient condition for the corresponding thermal conductivities to coincide.

The very possibility of defining an energy current density, from which the energy fluxes of Eqs. (35) and (40) ultimately depend, stems from energy extensivity. The considerations illustrated in Fig. 1 indicate that any two densities, $e'(\mathbf{r}, t)$ and $e(\mathbf{r}, t)$, whose integrals over a macroscopic volume differ by a quantity that scales as the volume boundary, should be considered as equivalent. This equivalence can be expressed by the condition that two equivalent densities differ by the divergence of a (bounded) vector field:

$$e'(\mathbf{r}, t) = e(\mathbf{r}, t) - \nabla \cdot \boldsymbol{p}(\mathbf{r}, t). \tag{42}$$

In a sense, two equivalent energy densities can be thought of as different *gauges* of the same scalar field. Energy is also conserved: because of this, for any given gauge of the energy density, $e(\mathbf{r}, t)$, an energy current density can be defined, $\boldsymbol{j}(\mathbf{r}, t)$, so as to satisfy the continuity equation, Eq. (3). By combining Eqs. (42) and (3), we see that energy current densities and macroscopic fluxes transform under a gauge transformation as:

$$\boldsymbol{j}'(\mathbf{r}, t) = \boldsymbol{j}(\mathbf{r}, t) + \dot{\boldsymbol{p}}(\mathbf{r}, t), \tag{43}$$

$$\mathbf{J}'(t) = \mathbf{J}(t) + \dot{\mathbf{P}}(t), \tag{44}$$

where $\mathbf{P}(t) = \frac{1}{\Omega} \int \boldsymbol{p}(\mathbf{r}, t) d\mathbf{r}$. We conclude that the macroscopic energy fluxes in two different energy gauges differ by the total time derivative of a bounded phase-space vector function.

We now show that the energy fluxes of the same system in two different energy gauges, e and e', differing by a bounded total time derivative, as in Eq. (44), result in the same heat conductivity, as given by the Green-Kubo formula, Eq. (29). More generally, the Onsager coefficients coupling two fluxes, $\mathbf{J}^1$ and $\mathbf{J}^2$, do not depend on the gauge of either one of them. In fact, let $(\mathbf{J}^1)' = \mathbf{J}^1 + \dot{\mathbf{P}}$; one has:

$$
\begin{aligned}
\left(L^{11}\right)' &= \frac{\Omega}{2k_B} \int_{-\infty}^{+\infty} \left\langle \left(\mathbf{J}^1(t) + \dot{\mathbf{P}}(t)\right) \cdot \left(\mathbf{J}^1(0) + \dot{\mathbf{P}}(0)\right) \right\rangle dt \\
&= L^{11} + \frac{\Omega}{2k_B} \left[\left\langle \mathbf{P}(t) \cdot \dot{\mathbf{P}}(0) \right\rangle \Big|_{-\infty}^{+\infty} + 2 \left\langle \mathbf{P}(t) \cdot \mathbf{J}^1(0) \right\rangle \Big|_{-\infty}^{+\infty} \right].
\end{aligned}
\tag{45}
$$

The expectation of the time-lagged products in Eq. (45) is equal to the products of two expectations at large time lag. As the equilibrium expectations of both a total time derivative and a current vanish, we conclude that $(L^{11})' = L^{11}$. A slight generalization of this argument, also using microscopic reversibility as in Onsager (1931a, b), allows us to conclude that $(L^{12})' = L^{12}$ and that, in general, $\kappa' = \kappa$.

3.1 Molecular Fluids

In a one-component molecular fluid such as liquid water or, say, ethanol, there are in general Q fluxes interacting with each other through Onsagers' Eq. (9), where Q is the number of atomic species in a molecule. The requirement that atoms are bound in molecules of fixed composition, however, sets a number of constraints that substantially simplify the treatment of heat transport, making the molecular case similar to the one-component one.

Let us consider a molecule of chemical formula $A_{N_A} B_{N_B} \cdots$, where $A, B, \cdots$ indicate atomic species and $N_A, N_B, \cdots$ the corresponding atomic stoichiometric indices. For each atomic species, we define the normalized number flux as:

$$\mathbf{J}^X = \frac{1}{N_X} \sum_{n \in X} \mathbf{V}_n. \tag{46}$$

If we indicate by M_X the atomic mass of species X, momentum conservation requires that $\sum_X M_X N_X \mathbf{J}^X = 0$ in the center-of-mass reference frame. The flux $\mathbf{J}^{XY} = \mathbf{J}^X - \mathbf{J}^Y$ is the total time derivative of a bounded vector, because its integral is the sum over all the molecules of the difference between the average atomic positions of either species within a same molecule, which is obviously bounded if molecules do not dissociate. As any number flux $\mathbf{J}^X$ can be expressed as a linear combination of the total momentum and of several $\mathbf{J}^{XY}$ fluxes, each of them is the total time derivative of a bounded vector. Therefore, the Onsager coefficient coupling any of these atomic fluxes with any other, or with the energy flux, vanishes. We conclude that energy is the only conserved quantity relevant for heat transport in a molecular fluid and that the energy-flux autocorrelation function directly yields the thermal conductivity, as in Eq. (1).

4   Density-Functional Theory of Adiabatic Heat Transport

Quantum simulation methods based on density-functional theory (DFT) have long been thought to be incompatible with the GK theory of thermal transport *because in first-principles calculations, it is impossible to uniquely decompose the total energy into individual contributions from each atom* (Stackhouse et al. 2010). For this reason, ab initio simulations of heat transport have often been performed using non-equilibrium approaches.

Stackhouse et al. (2010), for instance, computed the thermal conductivity of periclase MgO using a method devised by Müller-Plathe (1997). In this approach, a net heat flux, rather than a temperature gradient, is imposed to the simulated system, and the thermal conductivity is evaluated as the ratio between the heat flux and the resulting temperature gradient.

In the so-called *approach to equilibrium* methodology of Lampin et al. (2013), the system is first prepared in an out-of-equilibrium state characterized by an inhomogeneous temperature distribution, and the thermal conductivity is evaluated from the time it takes for the system to relax to equilibrium. This technique has been combined with AIMD to simulate thermal transport in a GeTe$_4$ glass by Bouzid et al. (2017) and further generalized and applied to crystalline and nanostructured MgO by Puligheddu et al. (2017).

Recently, there have been several attempts to combine the GK approach to heat transport with ab initio techniques based on electronic-structure theory, by adopting some ad hoc definitions for the energy flux. Kang and Wang (2017), for

instance, derived an expression for the energy flux from a (rather arbitrary) quantum-mechanical definition of the atomic energies and used a modified MD integration algorithm to cope with the difficulties ensuing from the implementation of their expression in PBC. Carbogno et al. (2017) gave a different expression for the energy flux, based on a normal-mode decomposition of the atomic coordinates and forces, which, while allowing to reduce the effects of thermal fluctuations, can only be applied to crystalline solids.

In spite of the undoubted ingenuity of these proposals, the problem still remains as of how it is possible that a rather arbitrary definition of the heat flux results in an allegedly well-defined value for the thermal conductivity. The gauge-invariance principle introduced in Sect. 3 not only provides a solution to this conundrum, but it also gives a rigorous way of deriving an expression for the energy flux directly from DFT, without introducing any ad hoc ingredients.

In order to derive such an expression for the adiabatic energy flux, we start with the standard DFT expression of the total energy in terms of the Kohn-Sham (KS) eigenvalues ε_v, eigenfunctions $\phi_v(\mathbf{r})$, and density $n(\mathbf{r}) = \sum_v |\phi_v(\mathbf{r})|^2$ (Martin 2008):

$$E_{DFT} = \frac{1}{2}\sum_n M_n V_n^2 + \frac{e^2}{2}\sum_{n,m\neq n}\frac{Z_n Z_m}{|\mathbf{R}_n - \mathbf{R}_m|}$$

$$+ \sum_v \varepsilon_v - \frac{e^2}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + \int (\epsilon_{XC}[n](\mathbf{r}) - \mu_{XC}[n](\mathbf{r}))\,n(\mathbf{r})d\mathbf{r}, \qquad (47)$$

where e is the electron charge; $\epsilon_{XC}[n](\mathbf{r})$ is a local exchange-correlation (XC) energy per particle defined by the relation $\int \epsilon_{XC}[n](\mathbf{r})n(\mathbf{r})d\mathbf{r} = E_{XC}[n]$, the latter being the total XC energy of the system; and $\mu_{XC}(\mathbf{r}) = \frac{\delta E_{XC}}{\delta n(\mathbf{r})}$ is the XC potential. The DFT total energy can be readily written as the integral of a DFT energy density (Chetty and Martin 1992):

$$E_{DFT} = \int e_{DFT}(\mathbf{r})d\mathbf{r},$$

$$e_{DFT}(\mathbf{r}) = e_{el}(\mathbf{r}) + e_Z(\mathbf{r}), \qquad (48)$$

where:

$$e_{el}(\mathbf{r}) = \mathfrak{Re}\sum_v \phi_v^*(\mathbf{r})\big(H_{KS}\phi_n(\mathbf{r})\big)$$

$$- \frac{1}{2}n(\mathbf{r})v_H(\mathbf{r}) + (\epsilon_{XC}(\mathbf{r}) - \mu_{XC}(\mathbf{r}))\,n(\mathbf{r}), \qquad (49)$$

$$e_Z(\mathbf{r}) = \sum_n \delta(\mathbf{r} - \mathbf{R}_n)\left(\frac{1}{2}M_n V_n^2 + w_n\right), \qquad (50)$$

$$w_n = \frac{e^2}{2} \sum_{m \neq n} \frac{Z_n Z_m}{|\mathbf{R}_n - \mathbf{R}_m|}, \tag{51}$$

H_{KS} is the instantaneous self-consistent Kohn-Sham Hamiltonian and $v_H = e^2 \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ is the Hartree potential. An explicit expression for the DFT energy flux is obtained by computing the first moment of the time derivative of the energy density, Eqs. (48), (49), (50), and (51), as indicated in Eq. (31), resulting in a number of terms, some of which are either infinite or ill-defined in PBC. Casting the result in a regular, boundary-insensitive, expression requires a careful breakup and refactoring of the various harmful terms, as explained by Marcolongo (2014) and in the online version of Marcolongo et al. (2016). The final result reads:

$$\mathbf{J}_{DFT}^E = \mathbf{J}^H + \mathbf{J}^Z + \mathbf{J}^0 + \mathbf{J}^{KS} + \mathbf{J}^{XC}, \tag{52}$$

$$\mathbf{J}^H = \frac{1}{4\pi \Omega e^2} \int \nabla v_H(\mathbf{r}) \dot{v}_H(\mathbf{r}) d\mathbf{r}, \tag{53}$$

$$\mathbf{J}^Z = \frac{1}{\Omega} \sum_n \left[\mathbf{V}_n \left(\frac{1}{2} M_n V_n^2 + w_n \right) + \sum_{m \neq n} (\mathbf{R}_n - \mathbf{R}_m) \left(\mathbf{V}_m \cdot \frac{\partial w_n}{\partial \mathbf{R}_m} \right) \right], \tag{54}$$

$$\mathbf{J}^0 = \frac{1}{\Omega} \sum_n \sum_v \left\langle \phi_v \left| (\mathbf{r} - \mathbf{R}_n) \left(\mathbf{V}_n \cdot \frac{\partial \hat{v}_0}{\partial \mathbf{R}_n} \right) \right| \phi_v \right\rangle, \tag{55}$$

$$\mathbf{J}^{KS} = \frac{1}{\Omega} \mathfrak{Re} \sum_v \langle \bar{\boldsymbol{\phi}}_v^c | H_{KS} + \varepsilon_v | \dot{\phi}_v^c \rangle, \tag{56}$$

$$J_\alpha^{XC} = \begin{cases} 0 & \text{(LDA)} \\ -\frac{1}{\Omega} \int n(\mathbf{r}) \dot{n}(\mathbf{r}) \frac{\partial \epsilon^{GGA}(\mathbf{r})}{\partial (\partial_\alpha n)} d\mathbf{r} & \text{(GGA)}, \end{cases} \tag{57}$$

where $\hat{v}_0$ is the bare, possibly nonlocal, (pseudo-)potential acting on the electrons and:

$$|\bar{\boldsymbol{\phi}}_v^c\rangle = \hat{P}_c \, \mathbf{r} \, |\phi_v\rangle, \tag{58}$$

$$|\dot{\phi}_v^c\rangle = \hat{\dot{P}}_v \, |\phi_v\rangle, \tag{59}$$

are the projections over the empty-state manifold of the action of the position operator over the v-th occupied orbital, Eq. (58), and of its adiabatic time derivative (Giannozzi et al. 2017), Eq. (59), $\hat{P}_v$ and $\hat{P}_c = 1 - \hat{P}_v$ being the projector operators over the occupied- and empty-states manifolds, respectively. Both these functions are well defined in PBC and can be computed, explicitly or implicitly, using standard density-functional perturbation theory (Baroni et al. 2001).

5 Data Analysis

The MD evaluation of the GK integral, Eq. (1), usually proceeds in two steps. One first evaluates the integrand as a running average of the time-lagged current products, $\langle J^i(\tau)J^j(0)\rangle \sim \frac{1}{\mathcal{T}-\tau}\int_0^{\mathcal{T}-\tau} J^i(t+\tau)J^j(t)dt$, where $\mathcal{T}$ is the length of the MD trajectory. The matrix defined in Eq. (19) is then estimated as a function of the upper limit of integration: $L^{ij}(\mathcal{T}) \propto \frac{\Omega}{k_B}\int_0^{\mathcal{T}}\langle J^i(\tau)J^j(0)\rangle\, d\tau$. One then recovers, via Eq. (36), an estimate for the thermal conductivity depending on $\mathcal{T}$: $\kappa(\mathcal{T}) \propto \frac{1}{(L^{-1}(\mathcal{T}))^{11}}$. This function is usually very noisy: in fact, at times greater than the correlation time between J^i and J^j, the correlation function $\langle J^i(\tau)J^j(0)\rangle$ approaches zero; hence $L^{ij}(\mathcal{T})$ starts integrating noise and behaves like the distance traveled by a random walk, whose variance grows linearly with the upper integration limit. The evaluation of transport coefficients thus requires averaging over multiple trajectories (possibly multiple segments of a same long trajectory) and estimating the resulting uncertainty as a function of both the length of each trajectory and the upper limit of integration. This is a cumbersome task that often leads to a poor estimate of the statistical and systematic errors on the computed conductivity, all the more so when the signal is inherently oscillatory, due to the existence of high-frequency features in the power spectrum of the energy flux, possibly due to intramolecular oscillations that meddle with the noise. Some authors try to overcome these problems by either fitting the autocorrelation function or the GK integral with a multi-exponential function (Schelling et al. 2002; Zhang et al. 2015) or by extrapolating the power spectrum of the energy flux to the zero-frequency limit (Volz and Chen 2000). Others have attempted an error analysis of the MD estimate of the GK integral, based on either heuristic or rigorous arguments (Jones and Mandadapu 2012; Wang et al. 2017; Oliveira and Greaney 2017), but they all require an estimate of an optimal value for the upper limit of integration, which determines a bias in the estimate and which is in general difficult to obtain. Different classes of systems require different approaches to error analysis, but it is widely believed that all of them always require so long simulation times as to be unaffordable with accurate but expensive AIMD techniques (Carbogno et al. 2017). In order to solve this problem, Ercole et al. (2017) considered it in the light of the statistical theory of stationary time series.

5.1 Solids and One-Component Fluids

In practice, MD gives access to a discrete sample of the flux process (a *time series*), $J_n = J(n\epsilon)$, $0 \leq n \leq N-1$, where ϵ is the sampling period of the flux and N the length of the time series, that we assume to be even. As was shown in Sect. 2.2.1, the Wiener-Khintchine theorem allows one to express the heat conductivity in terms of the zero-frequency value of the power spectrum of the energy flux (see Eqs. (21), (22), (23), and (24)):

$$\kappa = \frac{\Omega}{2k_B T^2} S(\omega = 0). \tag{60}$$

Let us define the discrete Fourier transform of the flux time series as:

$$\tilde{J}_k = \sum_{n=0}^{N-1} e^{2\pi i \frac{kn}{N}} J_n, \tag{61}$$

for $0 \le k \le N - 1$. Here, the convention for the sign in the exponential of the time-to-frequency Fourier transform is opposite to what is adopted in Ercole et al. (2017) and in most of the signal analysis literature, in order to comply with the convention for the space-time Fourier transforms usually adopted in the Physics literature and in Eqs. (4) and (5). The *sample spectrum* $\hat{S}_k$, aka *periodogram*, is defined as:

$$\hat{S}_k = \frac{\epsilon}{N} \left| \tilde{J}_k \right|^2, \tag{62}$$

and, for large N, it is an unbiased estimator of the power spectrum of the process, as defined in Eq. (21), evaluated at $\omega_k = 2\pi \frac{k}{N\epsilon}$, namely: $\langle \hat{S}_k \rangle = S(\omega_k)$. The reality of the $\hat{J}$s implies that $\tilde{J}_k = \tilde{J}_{N-k}^*$ and $\hat{S}_k = \hat{S}_{N-k}$, so that periodograms are usually reported for $0 \le k \le \frac{N}{2}$ and their Fourier transforms evaluated as discrete cosine transforms.

The space autocorrelations of conserved currents are usually short-ranged. Therefore, in the thermodynamic limit, the corresponding fluxes can be seen as sums of (almost) independent identically distributed stochastic variables, so that, according to the central-limit theorem, their equilibrium distribution is Gaussian. A slight generalization of this argument allows us to conclude that any conserved-flux process is Gaussian as well. The flux time series is in fact a multivariate stochastic variable that, in the thermodynamic limit, results from the sum of (almost) independent variables, thus tending to a multivariate normal deviate. This implies that at equilibrium, the real and imaginary parts of the $\tilde{J}_k$'s defined in Eq. (61) are zero-mean normal deviates that, in the large-N limit, are uncorrelated among themselves and have variances proportional to the power spectrum evaluated at ω_k. For $k = 0$ or $k = \frac{N}{2}$, $\tilde{J}_k$ is real and $\sim \mathcal{N}\left(0, \frac{N}{\epsilon} S(\omega_k)\right)$; for $k \notin \left\{0, \frac{N}{2}\right\}$, $\mathfrak{Re}\tilde{J}_k$ and $\mathfrak{Im}\tilde{J}_k$ are independent and both $\sim \mathcal{N}\left(0, \frac{N}{2\epsilon} S(\omega_k)\right)$, where $\mathcal{N}(\mu, \sigma^2)$ indicates a normal deviate with mean μ and variance σ^2. We conclude that in the large-N limit the sample spectrum of the heat flux time series reads:

$$\hat{S}_k = S(\omega_k) \xi_k, \tag{63}$$

where the ξs are independent random variables distributed as a χ_1^2 variate for $k = 0$ or $k = \frac{N}{2}$ and as one half a χ_2^2 variate, otherwise. Here and in the following χ_ν^2 indicates the chi-square distribution with ν degrees of freedom. For the sake of simplicity, we make as though all the ξs were identically distributed, $\xi_k \sim \frac{1}{2}\chi_2^2$

for all values of k, thus making an error of order $\mathcal{O}(1/N)$, which vanishes in the longtime limit that is being assumed throughout this section.

In many cases of practical interest, multiple time series are available to estimate the power spectrum of a same process, $\{^{p}J_{n}\}$, $p = 1, \cdots \ell$. For instance, in equilibrium MD, a same trajectory delivers one independent time series per Cartesian component of the heat flux, all of which are obviously equivalent in isotropic systems. In these cases, it is expedient to define a mean sample spectrum by averaging over the ℓ different realizations:

$$
\begin{aligned}
{}^{\ell}\hat{S}_{k} &= \frac{\epsilon}{\ell N} \sum_{p=1}^{\ell} \left| {}^{p}\tilde{J}_{k} \right|^{2} , \\
&= S\left(\omega_{k} \right) {}^{\ell}\xi_{k},
\end{aligned}
\tag{64}
$$

where the ${}^{\ell}\xi$s are $\chi_{2\ell}^{2}$ variates, divided by the number of degrees of freedom: ${}^{\ell}\xi_{k} \sim \frac{1}{2\ell}\chi_{2\ell}^{2}$ (for $k \notin \{0, \frac{N}{2}\}$).

Equations (63) and (64) show that ${}^{\ell}\hat{S}_{0}$ is an unbiased estimator of the zero-frequency value of the power spectrum, $\langle {}^{\ell}\hat{S}_{0} \rangle = S(0)$, and through Eq. (60), of the transport coefficients we are after. However, this estimator is not consistent, i.e., its variance does not vanish in the large-N limit. This is so because a longer time series increases the number of discrete frequencies at which the power spectrum is sampled, rather than its accuracy at any one of them.

Figure 3 displays the periodogram of water at ambient conditions, obtained from a 100 ps classical MD trajectory, showing the extremely noisy behavior of
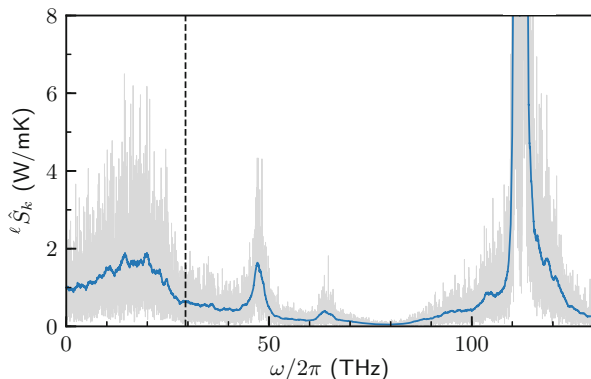
Fig. 3 Periodogram of a classical flexible model of water obtained from a 100 ps MD trajectory. Gray: periodogram obtained directly from Eq. (64), with $\ell = 3$. Blue: periodogram filtered with a moving average window of width 1 THz, useful to reveal the main features of the spectrum (see text). The vertical dashed line delimits the low-frequency region used in the subsequent cepstral analysis

the periodogram as an estimator of the spectrum. Averaging over the values of the periodogram within a frequency window of given width (Weisstein MovingAverage) would consistently reduce the statistical noise, but the multiplicative nature of the latter in Eq. (63) makes it difficult to disentangle the noise from the signal and may introduce a bias. In order to cope with this problem, we had better transform the multiplicative noise into an additive one by defining the log-periodogram, ${}^{\ell}\hat{L}_k$, as:

$$
\begin{aligned}
{}^{\ell}\hat{L}_k &= \log\left({}^{\ell}\hat{S}_k\right) \\
&= \log\left(S(\omega_k)\right) + \log\left({}^{\ell}\hat{\xi}_k\right) \\
&= \log\left(S(\omega_k)\right) + {}^{\ell}\Lambda + {}^{\ell}\lambda_k,
\end{aligned}
\tag{65}
$$

where ${}^{\ell}\lambda_k = \log\left({}^{\ell}\hat{\xi}_k\right) - {}^{\ell}\Lambda$ are zero-mean identically distributed independent stochastic variables, ${}^{\ell}\Lambda = \left\langle\log\left({}^{\ell}\hat{\xi}\right)\right\rangle = \psi(\ell) - \log(\ell)$, and $\psi(z)$ is the digamma function (Weisstein PolyGamma). The variance of the ${}^{\ell}\lambda$ variables is $\sigma_{\ell}^2 = \psi'(\ell)$, where $\psi'(z)$ is the tri-gamma function (Weisstein PolyGamma).

Whenever the number of (inverse) Fourier components of the logarithm of the power spectrum is much smaller than the length of the time series, applying a low-pass filter to Eq. (65) would result in a reduction of the power of the noise, without affecting the signal. In order to exploit this idea, we define the "*cepstrum*" of the time series as the inverse Fourier transform of its sample log-spectrum (Childers et al. 1977):

$$
{}^{\ell}\hat{C}_n = \frac{1}{N}\sum_{k=0}^{N-1} {}^{\ell}\hat{L}_k e^{-2\pi i\frac{kn}{N}}.
\tag{66}
$$

A generalized central-limit theorem for Fourier transforms of stationary time series ensures that, in the large-N limit, these coefficients are a set of independent (almost) identically distributed zero-mean normal deviates (Anderson 1994; Peligrad and Wu 2010). It follows that:

$$
\begin{aligned}
{}^{\ell}\hat{C}_n &= \lambda_{\ell}\delta_{n0} + C_n + {}^{\ell}\mu_n, \\
C_n &= \frac{1}{N}\sum_{k=0}^{N-1}\log\left(S(\omega_k)\right)e^{-2\pi i\frac{kn}{N}},
\end{aligned}
\tag{67}
$$

where ${}^{\ell}\mu_n$ are independent zero-mean *normal* deviates with variances $\left\langle{}^{\ell}\mu_n^2\right\rangle = \frac{1}{N}\sigma_{\ell}$ for $n \notin \left\{0, \frac{N}{2}\right\}$ and $\left\langle{}^{\ell}\mu_n^2\right\rangle = \frac{2}{N}\sigma_{\ell}^2$ otherwise. Figure 4 displays the cepstral coefficients of the low-frequency region of the spectrum of water (marked in Fig. 3), showing that only the first few coefficients are substantially different from zero.
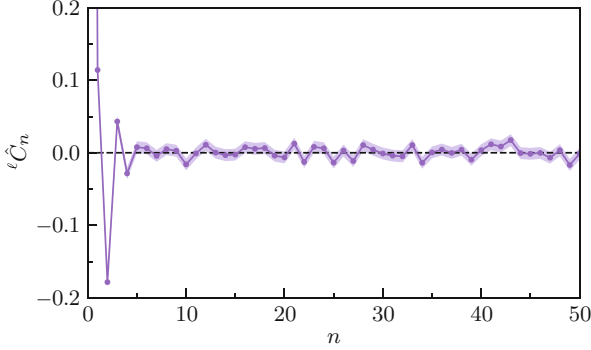
Fig. 4 Cepstral coefficients of water computed analyzing the low-frequency region of the periodogram (see Fig. 3), defined in Eq. (66)

Let us indicate by P^* the smallest integer such that $C_n \approx 0$ for $P^* \leq n \leq N - P^*$. By limiting the Fourier transform of the sample cepstrum, Eq. (66), to P^* coefficients, we obtain an efficient estimator of the zero-frequency component of the log-spectrum as:

$$
\begin{aligned}
{}^{\ell}\hat{L}_0^* &= {}^{\ell}\hat{C}_0 + 2 \sum_{n=1}^{P^*-1} {}^{\ell}\hat{C}_n \\
&= {}^{\ell}\Lambda + \log(S_0) + {}^{\ell}\mu_0 + 2 \sum_{n=1}^{P^*-1} {}^{\ell}\mu_n.
\end{aligned}
\tag{68}
$$

Inspection of Eq. (68) shows that ${}^{\ell}\hat{L}_0^*$ is a normal estimator whose expectation and variance are:

$$
\langle {}^{\ell}\hat{L}_0^* \rangle = \log(S_0) + {}^{\ell}\Lambda,
\tag{69}
$$

$$
\sigma_\ell^*(P^*, N)^2 = \sigma_\ell^2 \frac{4P^* - 2}{N}.
\tag{70}
$$

Using Eq. (60), we see that the logarithm of the conductivity can be estimated from the cepstral coefficients of the flux time series through Eqs. (68), (69), and (70) and that the resulting estimator is always normal with a variance that depends on the specific system *only* through the number of these coefficients, P^*. Notice that the absolute error on the logarithm of the conductivity directly and nicely yields the relative error on the conductivity itself.

The efficacy of this approach obviously depends on our ability to estimate the number of coefficients necessary to keep the bias introduced by the truncation to

a value smaller than the statistical error while maintaining the magnitude of the latter at a prescribed acceptable level. Ercole et al. (2017) proposed to estimate P^* using the Akaike's information criterion (Akaike 1974), but other more advanced *model selection* approaches (Claeskens and Hjort 2008) may be more effective. This method consists in choosing P^* as the one that minimizes the function:

$$\text{AIC}(P) = \frac{N}{\sigma_\ell^2} \sum_{n=P}^{\frac{N}{2}} \hat{C}_n^2 + 2P. \qquad (71)$$

In Fig. 5a, we report the low-frequency region of the spectrum of water obtained by limiting the number of cepstral coefficients to P^*:

$$^\ell \hat{S}_k^* = \exp\left[2 \sum_{n=1}^{P^*-1} {}^\ell \hat{C}_n e^{2\pi i \frac{kn}{N}} + {}^\ell \hat{C}_0 - {}^\ell \Lambda \right], \qquad (72)$$

thus showing the filtering effect of this choice. Finally, Fig. 5b shows the value of thermal conductivity of water obtained through Eqs. (68), (69), and (70).

5.2 Multicomponent Fluids

In Sect. 2.3.2, we have seen that in a fluid made of Q atomic species, there are in general Q macroscopic fluxes interacting with each other through Onsager's phenomenological equations, Eq. (9), not counting the different Cartesian components that do not interact among themselves because of space isotropy. A MD simulation thus samples Q stochastic processes, one for each interacting flux, that we suppose to be stationary. These processes can be thought of as different components of a same multivariate process (Bertossa et al. 2019). As in Sect. 5.1, for the sake of generality we suppose to have ℓ independent samples of such a process, described by a multivariate time series of length N: $\{{}^p J_n^i\}$; $p = 1, \ldots \ell$; $i = 1, \ldots Q$; $n = 0, \ldots N-1$. Stationarity implies that $\langle J_n^i \rangle$ does not depend on n and that $\langle J_n^i J_m^j \rangle$ only depends on $n - m$. We will further assume that $\langle J_n^i \rangle = 0$ and that $\langle J_n^i J_0^j \rangle$ is an even function of n, which is the case when J^i and J^j have the same signature under time-reversal. By combining Eq. (36) with Eq. (24), we see that in order to evaluate the thermal conductivity in the multicomponent case, we need an efficient estimator for $\left(S_0^{-1}\right)^{11}$, where $S_0^{kl} = S^{kl}(\omega = 0)$ is the zero-frequency cross-spectrum of the relevant fluxes, ordered in such a way that the energy one is the first.

Similarly to the one-component case, we define a mean sample cross-spectrum (or *cross-periodogram*) as:
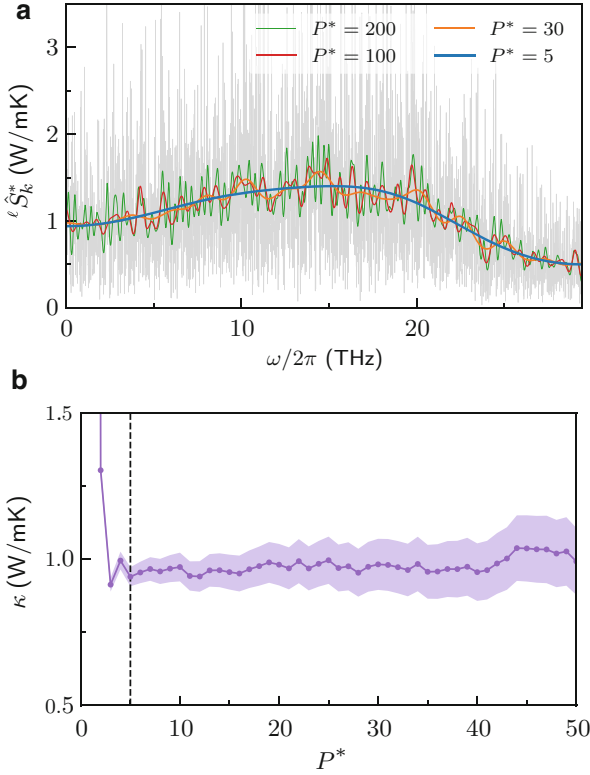
Fig. 5 (**a**) Filtered low-frequency region of the power spectrum of water obtained by limiting the number of cepstral coefficients to various values of P^*, Eq. (72). $P^* = 7$ is the cutoff value suggested by the Akaike's information criterion, Eq. (71). Gray: the unfiltered periodogram obtained from Eq. (62). (**b**) Thermal conductivity of water estimated from Eqs. (68), (69), and (70) as a function of the cutoff, P^*. The colored bands indicate one standard deviation as estimated from theory. The vertical dashed line indicates the value suggested by the Akaike's information criterion, Eq. (71)

$$^{(\ell Q)}\hat{S}_k^{ij} = \frac{1}{\ell} \sum_{p=1}^{\ell} \frac{\epsilon}{N} \left(^{p}\tilde{J}_k^i\right)^* {}^{p}\tilde{J}_k^j. \tag{73}$$

By discretizing Eq. (23), we see that $^{(\ell Q)}\hat{S}_k^{ij}$ is an unbiased estimator of the cross-spectrum, $\left\langle ^{(\ell Q)}\hat{S}_k^{ij}\right\rangle = S^{ij}\left(\omega_k = \frac{2\pi k}{N\epsilon}\right)$. As it was the case for univariate processes, in the large N limit the real and imaginary parts of $\tilde{J}_k^i$ are normal deviates that are uncorrelated for $k \neq k'$. We conclude that the cross-periodogram is a random matrix distributed as a complex Wishart deviate (Goodman 1963a, b):

$$^{(\ell Q)}\hat{S}_k \sim \mathcal{CW}_Q\left(S(\omega_k), \ell\right). \tag{74}$$

The notation $\mathcal{CW}_Q(S, \ell)$ in Eq. (74) indicates the distribution of the $Q \times Q$ Hermitian matrix ${}^{(\ell Q)}\hat{S}^{ij} = \frac{1}{\ell} \sum_{p=1}^{\ell} {}^{p}X^i \, {}^{p}X^{j*}$, where $\{{}^{p}X^i\}$ ($p = 1, \cdots \ell$, $i = 1, \cdots Q$) are ℓ samples of an Q-dimensional zero-mean normal variate whose covariance is $S^{ij} = \langle X^i X^{j*} \rangle$.

Similarly to the real case, a Bartlett decomposition (Kshirsagar 1959) holds for complex Wishart matrices (Nagar and Gupta 2011), reading:

$$^{(\ell Q)}\hat{S} = \frac{1}{\ell} \mathcal{S} R R^\top \mathcal{S}^\dagger, \tag{75}$$

where "$\top$" and "$\dagger$" indicate the transpose and the adjoint of a real and complex matrix, respectively; $\mathcal{S}$ is the Cholesky factor of the covariance matrix, $S = \mathcal{S}\mathcal{S}^\dagger$; and R is a real lower triangular random matrix of the form:

$$R = \begin{pmatrix} c_1 & 0 & 0 & \cdots & 0 \\ n_{21} & c_2 & 0 & \cdots & 0 \\ n_{31} & n_{32} & c_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_{Q1} & n_{Q2} & n_{Q3} & \cdots & c_Q \end{pmatrix}, \tag{76}$$

where $c_i^2 \sim \chi^2_{2(\ell-i+1)}$ and $n_{ij} \sim \mathcal{N}(0, 1)$. We stress that R is independent of the specific covariance matrix and only depends upon ℓ and Q. In particular, it is independent of the ordering of the fluxes J^i. By expressing the QQ matrix element of the inverse of ${}^{(\ell Q)}\hat{S}$ in Eq. (75) as the ratio between the corresponding minor and the full determinant, and using some obvious properties of the determinants and of triangular matrices, we find that:

$$\frac{\ell}{\left({}^{(\ell Q)}\hat{S}_k^{-1}\right)^{QQ}} = \frac{1}{\left(S_k^{-1}\right)^{QQ}} c_Q^2, \tag{77}$$

As the ordering of the fluxes is arbitrary, a similar relation holds for all the diagonal elements of the inverse of the cross-periodogram. We conclude that the generalization of Eq. (64) for the multicomponent case is:

$$^{\ell}\underline{\hat{S}}_k \equiv \frac{\ell}{2(\ell - Q + 1)} \frac{1}{\left({}^{(\ell Q)}\hat{S}_k^{-1}\right)^{11}} = \frac{1}{\left(S_k^{-1}\right)^{11}} \xi_k, \tag{78}$$

where ξ_k are independent random (with respect to k) random variables, distributed as:

$$\xi_k \sim \begin{cases} \frac{1}{\ell-Q+1}\,\chi^2_{\ell-Q+1} & \text{for } k \in \{0, \frac{N}{2}\}, \\[2em] \frac{1}{2(\ell-Q+1)}\,\chi^2_{2(\ell-Q+1)} & \text{otherwise.} \end{cases} \qquad (79)$$

Starting from here we can apply the cepstral analysis as in the one-component case. The only difference is the number of degrees of freedom of the χ^2 distribution, that becomes $2(\ell - Q + 1)$, and a different factor in front of the result. Figure 6 shows an example of multi-component power spectrum for a solution of water and ethanol.

The method discussed so far shows a fundamental advantage with respect to a naïve implementation of direct time-integration approach. Figure 7 shows the two-component conductivity κ, obtained via Eq. (38), in the case of a water-ethanol solution, as a function of the upper time-integration limit $\mathcal{T}$ (Bertossa et al. 2019). Both the Green-Kubo and the Einstein-Helfand definitions of the finite-time expression of Onsager's coefficients (see Eq. (20)) are displayed. Due to thermal fluctuations, the integral of the correlation function becomes a random walk as soon as the latter vanishes, eventually assuming any value. Therefore, there will be a set of times (see Fig. 7) where the term L^{QQ} at the denominator in Eq. (38) vanishes, leading to divergences in the evaluation of κ; an issue not affecting the one-component case. Hence, in such a formulation of the multicomponent case, the mean value of the thermal conductivity estimator *in the time domain* does not exist. On the contrary, the multicomponent frequency-domain approach presented in this section, and built on sound statistical basis, provides a well-defined expression for the estimator of κ and its statistical error.
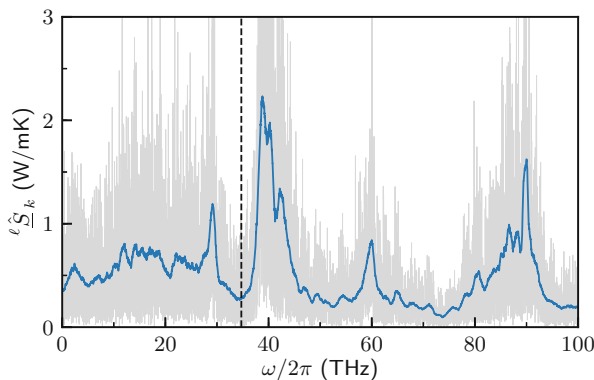
Fig. 6 Multicomponent power spectrum, as defined in Eq. (78), for a classical flexible model of a solution of water and ethanol 50 mol%, obtained from a 100 ps trajectory. Gray: $^\ell \hat{\underline{S}}_k$ obtained directly from Eq. (78), with $\ell = 3$ and $Q = 2$. Blue: $^\ell \hat{\underline{S}}_k$ filtered with a moving average window of width 1 THz in order to reveal its main features. The vertical dashed line delimits the low-frequency region used in the subsequent cepstral analysis (Bertossa et al. 2019)
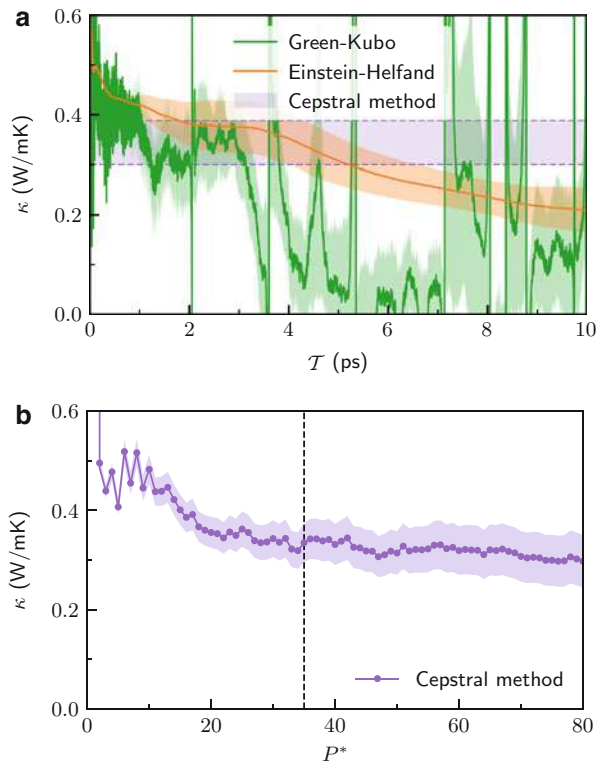
Fig. 7 Convergence of the multicomponent thermal conductivity estimator κ using the direct time-integration approach and the cepstral method, for a classical flexible model of a solution of water and ethanol 50 mol%, obtained from a 100 ps trajectory. (**a**) Direct time-integration approach in its Green-Kubo (green, as obtained from the matrix $L^{ij}(\mathcal{T}) \propto \int_0^{\mathcal{T}} \langle J^i(t) J^j(0) \rangle dt$) and Einstein-Helfand (orange – obtained from the matrix $(L^{ij})'(\mathcal{T}) \propto \int_0^{\mathcal{T}} (1 - \frac{t}{\mathcal{T}}) \langle J^i(t) J^j(0) \rangle dt$) formulations. The horizontal purple band indicates the value obtained by the cepstral method. (**b**) Estimate of κ with the cepstral method as a function of the number of cepstral coefficients, P^*, see Eqs. (68), (69), and (70). The dashed vertical line indicates the value of P^* selected by the AIC, Eq. (71) (Bertossa et al. 2019)

5.3   Data Analysis Workflow

We summarize the steps leading to the estimation of thermal conductivity by the *cepstral analysis* method, in order to highlight the simplicity of its practical implementation.

1. From a MD simulation compute the heat flux time series J_n^1 and the independent particle fluxes J_n^q, $q = 2, \ldots, Q$.
2. Compute the discrete Fourier transform of the fluxes, $\tilde{J}_k^i$, and the element $1/(\hat{S}^{-1})^{11}$. In practice, only a selected low-frequency region shall be used (see

Ercole et al. (2017) for a detailed discussion). To lighten the notation, we drop the left superscripts of the variables in this subsection.

3. Calculate $\log\left[1/(\hat{S}^{-1})^{11}\right]$.
4. Compute the inverse discrete Fourier transform of the result to obtain the cepstral coefficients $\hat{C}_n$.
5. Apply the Akaike Information Criterion, Eq. (71), to estimate the number of cepstral coefficients to retain, P^*.
6. Finally apply Eq. (68) to obtain $\hat{L}_0^*$, and evaluate the thermal conductivity as

$$\kappa = \frac{\Omega}{2k_B T^2}\exp\left[\hat{L}_0^* - \psi(\ell - Q + 1) + \log(\ell - Q + 1)\right], \tag{80}$$

and its statistical error as:

$$\frac{\Delta\kappa}{\kappa} = \sqrt{\psi'(\ell - Q + 1)\frac{4P^* - 2}{N}}. \tag{81}$$

6   A Few Representative Results

Calculations of the thermal conductivity based on the Green-Kubo formalism combined with first-principles molecular dynamics are quite recent. The first benchmarks from Marcolongo et al. (2016) have been performed on liquid Argon and heavy water at ambient conditions, as reported below.

6.1   A Benchmark on a Model Mono-atomic Fluid

As a first test, liquid Argon was simulated by Marcolongo et al. (2016) using a local LDA functional neglecting dispersion forces. The resulting fictitious system, dubbed LDA-Argon, is a hard-core weakly interacting fluid whose dynamics is expected to be easily mimicked by a simple two-body potential, which can be engineered by standard force-matching techniques. This observation allows one to effectively test the ideas developed in Sect. 4: even if the energy density in LDA-Argon and in its fitted classical counterpart will likely be different, the resulting thermal conductivity is expected to coincide within the quality of the classical fit.

Simulations were performed in a cubic supercell of 108 atoms with an edge of 17.5 Å, corresponding to a density of 1.34 g cm^{-3}. Trajectories were sampled in the NVE ensemble for 100 ps, and the classical model was fitted with a pair potential of the form $V(r) = P_2(r)e^{-\alpha r}$, $P_2(r)$ being a second-order polynomial. In Fig. 8, we show the resulting autocorrelation functions at a representative temperature of 400 K. The DFT autocorrelation function shows a more structured behavior than that of the classical potential. Nevertheless, when considering the long time limit of the

a

b
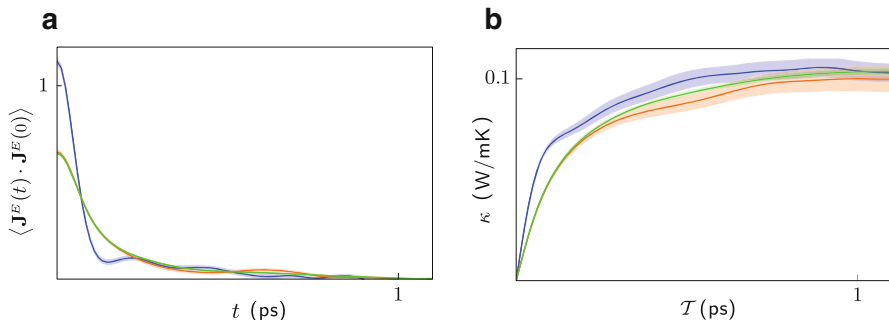
Fig. 8 Benchmark of LDA-Argon, reproduced from Marcolongo et al. (2016). (**a**) The heat-flux autocorrelation function. (**b**) The GK integral as a function of the upper limit of integration. Color coding common to the two panels. Blue: ab initio molecular dynamics (100 ps). Orange: classical molecular dynamics (100 ps). Green: classical molecular dynamics (1000 ps)

Green-Kubo integral, the thermal conductivities indeed coincide within statistical uncertainty, as predicted by theory.

6.2   Heavy Water at Ambient Conditions

Marcolongo et al. (2016) also computed the thermal conductivity of heavy water at ambient conditions. Simulations were performed using the PBE exchange correlation energy functional (Perdew et al. 1996), which is known to predict qualitatively the self-diffusion coefficient of water at ambient conditions only for a simulation temperature of about 400 K (Sit and Marzari 2005): the latter condition was imposed accordingly. A system of 64 heavy water molecules was sampled for 90 ps in the NVE ensemble using a cubic cell corresponding to the experimental density of $1.11 \, \mathrm{g\,cm}^{-3}$.

The ab initio treatment of heat transport in molecular fluids requires some care in order to eliminate non-diffusive components of the energy flux that, while not contributing to the conductivity, do increase the noise of the flux time series to a level that may compromise its analysis. To see where the problem comes from, let us split the potential energy of the system into the sum of non-interacting atomic energies plus an interaction energy as: $V(\{\mathbf{R}_n\}) = \sum_n \epsilon_n^\circ + V_{int}(\{\mathbf{R}_n\})$, where ϵ_n° is the energy of the n-th atom when it is isolated from the rest. In classical simulations, the energy of isolated atoms never enters the description of the system, and the ϵ_n°'s can be simply set to zero. In quantum simulations, instead, atomic and interaction energies enter on a same footing, and the former give a large and fluctuating contribution to the total energy flux, $\mathbf{J}^\circ = \sum_X \epsilon_X^\circ \mathbf{J}^X$, where $\mathbf{J}^X$ is the flux defined in Eq. (46). In a mono-atomic fluid, $\mathbf{J}^X$ is constant because of momentum conservation, and it is actually equal to zero in the center-of-mass reference frame. In molecular fluids, the $\mathbf{J}^X$ do not vanish, but, as we have seen in Sect. 3.1, they are non-diffusive and hence do not contribute to the heat conductivity while adding considerable noise

to the energy-flux time series. In order to remove them, instead of estimating $\mathbf{J}°$ from the noninteracting atomic energies, we prefer to implement a *decorrelation technique*, as described below.

Current decorrelation builds on a general inequality whose proof can be found in Marcolongo et al. (2016). Let $\mathbf{J}^1$ and $\mathbf{J}^2$ be two macroscopic fluxes and $\mathbf{J}^{12} = \mathbf{J}^1 + \mathbf{J}^2$ their sum. The corresponding conductivities κ^1, κ^2, and κ^{12} then satisfy $|\kappa^{12} - \kappa^1 - \kappa^2| \leq 2\sqrt{\kappa^1 \kappa^2}$. As a consequence, when κ^2 vanishes, κ^{12} coincides with κ^1. Let us now suppose that a set of fluxes $\{\mathbf{Y}^u\}$, $u = 1, \ldots U$ is known to exhibit a non-diffusive behavior. The above argument shows that the auxiliary flux defined as:

$$\mathbf{J}' \equiv \mathbf{J} - \sum_w \lambda^w \mathbf{Y}^w \tag{82}$$

will yield the same thermal conductivity as $\mathbf{J}$. Optimal values of the $\{\lambda^u\}$ coefficients can then be determined by imposing that the new time series $\mathbf{J}'$ is uncorrelated with respect to the non-diffusive ones, namely:

$$\langle \mathbf{J} \mathbf{Y}^u \rangle - \sum_w \lambda^w \langle \mathbf{Y}^w \mathbf{Y}^u \rangle = 0, \quad u = 1, \ldots U. \tag{83}$$

This procedure is particularly useful when the $\mathbf{Y}^u$ fluxes give a slowly converging contribution to the Green-Kubo integral, which is thus difficult to evaluate numerically.

The decorrelation technique has been applied to heavy water considering two non-diffusive number fluxes: $\mathbf{Y}^1 = \mathbf{J}^H + \mathbf{J}^O$, i.e., the sum of hydrogen and oxygen average velocities and $\mathbf{Y}^2 = \mathbf{J}^{el}$, the adiabatic electronic current. Note that the two time series $\mathbf{J}^H$ and $\mathbf{J}^O$ are trivially related, because of momentum conservation. Therefore $\mathbf{J}^H$, $\mathbf{J}^O$, or $\mathbf{J}^H + \mathbf{J}^O$ would all be equivalent choices. The latter is defined, following the same notation of Sect. 4, as:

$$\mathbf{J}^{el} = \frac{2}{\Omega} \mathfrak{Re} \sum_v \langle \bar{\boldsymbol{\phi}}_v^c | \dot{\boldsymbol{\phi}}_v^c \rangle, \tag{84}$$

as can be derived from the continuity equation for the density: $\nabla \cdot \boldsymbol{j}^{el}(\boldsymbol{r}, t) = -\dot{n}^{el}(\boldsymbol{r}, t)$. In insulators $\mathbf{J}^{el}$ is non-diffusive and can thus be used to decorrelate the heat current. In the original paper the thermal conductivity was evaluated from the slope of the energy displacement $\mathcal{D}(\tau) = \int_0^\tau \mathbf{J}(t) dt$ (see Eq. (20)) and the corresponding error obtained from a standard block analysis, resulting in a nonoptimal estimate of both (see Fig. 9a). The same data have been reanalyzed with the novel cepstral technique presented in Sect. 5. We denote by $(\mathbf{J}^1)'$ and $(\mathbf{J}^{12})'$ the currents decorrelated with respect to $\mathbf{Y}^1$ alone and with respect to both $\mathbf{Y}^1$ and $\mathbf{Y}^2$, respectively. The power spectra of the two currents are plotted in Fig. 9b. Only the low-frequency region of the spectra (up to ~ 9.0 THz) was used for the cepstral analysis (see Ercole et al. (2017) for the technical details). The minimization
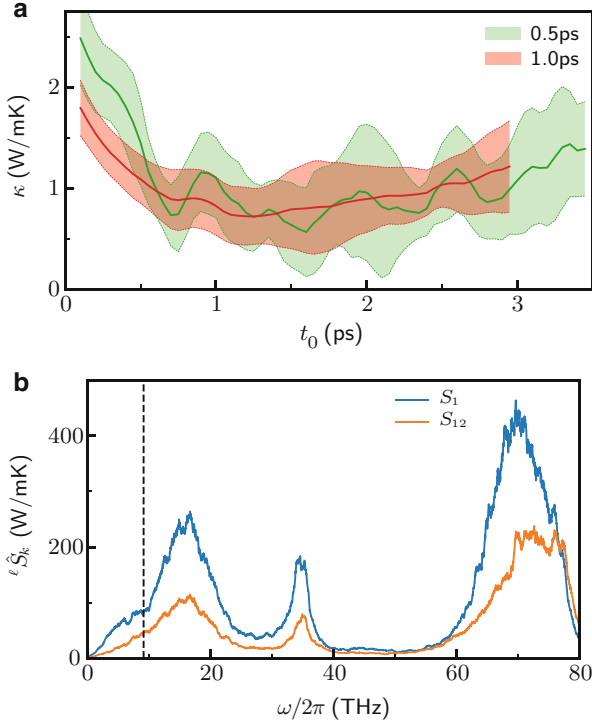
Fig. 9 (**a**) Value of thermal conductivity of heavy water obtained from a linear fitting of the energy displacement of $(\mathbf{J}^1)'$, $\mathcal{D}(\tau) = \int_0^\tau (\mathbf{J}^1)'(t)dt$. The two curves refer to different window widths used for the linear fit, of length 0.5 and 1.0 ps. The abscissa corresponds to the origin of the fitting window. (**b**) Periodogram of the $(\mathbf{J}^1)'$ and $(\mathbf{J}^{12})'$ currents, filtered with a moving average window in order to reveal the prominent features. The vertical dashed line delimits the low-frequency region used for cepstral analysis

of Eq. (71) suggests a cutoff $P_1^* = 17$ and $P_{12}^* = 15$, but a value 1.5 times larger was actually been used to reduce the bias possibly due to the fast variation of the spectrum at frequency close to zero. The resulting thermal conductivities are $\kappa^1 = 0.80 \pm 0.12$ W/mK for the $(\mathbf{J}^1)'$ flux and $\kappa^{12} = 0.93 \pm 0.14$ W/mK for $(\mathbf{J}^{12})'$, compatible with each other. By comparison, experiments give a value $\kappa \approx 0.6$ W/mK (Matsunaga and Nagashima 1983; Ramires et al. 1995). In this case, data analysis would not have yielded any meaningful results failing a proper decorrelation of the heat flux time series.

7 Outlook

We believe that the ideas presented in this chapter will pave the way to new developments and applications in the field of heat transport, particularly for strongly

anharmonic and/or disordered systems, where approaches based on the Boltzmann transport equation do not apply or are bound to fail. The general concept of gauge invariance of heat conductivity will likely apply to other transport properties as well, such as ionic conduction, viscosity, and many others, and/or simulation methodologies, such as those based on a neural-network representation of interatomic potentials, which hold the promise of a strong and long-lasting impact on molecular simulations. The applicability of this concept would not be as broad if not assisted by the powerful data-analysis methods which have also been described in this chapter. Here again, we believe that there is ample room for improvement, leveraging more general (possibly non-Fourier) representations of the log-spectrum of the currents to be analyzed and more advanced statistical-inference techniques to estimate the parameters of the spectral models resulting from these representations. From the applicative point of view, we expect that these methodological advances will have a strong impact in all those cases where heat, mass, and charge transport occurs in conditions that cannot be adequately described at the atomistic level by simple force fields, e.g., in complex materials; systems at extreme external conditions, such as those occurring in the planetary interiors; complex or reactive fluids; and many others.

8   Cross-References

▶ On the Kinetic Theory of Thermal Transport in Crystals
▶ Thermal Transport by First-Principles Anharmonic Lattice Dynamics
▶ Thermal Transport: Overview

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723. https://doi.org/10.1109/TAC.1974.1100705

Anderson TW (1994) The statistical analysis of time series. Wiley-Interscience, New York

Baroni S, Giannozzi P, Testa A (1987) Green's-function approach to linear response in solids. Phys Rev Lett 58:1861–1864. https://doi.org/10.1103/PhysRevLett.58.1861

Bertossa R, Grasselli F, Ercole L, Baroni S (2019) Transport coefficients in multi-component fluids from equilibrium molecular dynamics. Phys Rev Lett 122:255901. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.122.255901

Baroni S, de Gironcoli S, Dal Corso A, Giannozzi P (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73:515–562. https://doi.org/10.1103/RevModPhys.73.515

Bouzid A, Zaoui H, Palla PL, Ori G, Boero M, Massobrio C, Cleri F, Lampin E (2017) Thermal conductivity of glassy GeTe$_4$ by first-principles molecular dynamics. Phys Chem Chem Phys 19:9729–9732. https://doi.org/10.1039/C7CP01063J

Broido DA, Malorny M, Birner G, Mingo N, Stewart DA (2007) Intrinsic lattice thermal conductivity of semiconductors from first principles. Appl Phys Lett 91:231922. https://doi.org/10.1063/1.2822891

Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. Phys Rev Lett 55:2471–2474. https://doi.org/10.1103/PhysRevLett.55.2471

Carbogno C, Ramprasad R, Scheffler M (2017) Ab initio Green-Kubo approach for the thermal conductivity of solids. Phys Rev Lett 118:175901. https://doi.org/10.1103/PhysRevLett.118.175901

Casimir HBG (1945) On Onsager's principle of microscopic reversibility. Rev Mod Phys 17:343–350. https://doi.org/10.1103/RevModPhys.17.343

Chetty N, Martin R (1992) First-principles energy density and its applications to selected polar surfaces. Phys Rev B 45:6074–6088. https://doi.org/10.1103/PhysRevB.45.6074

Childers DG, Skinner DP, Kemerait RC (1977) The cepstrum: a guide to processing. Proc IEEE 65:1428–1443. https://doi.org/10.1109/PROC.1977.10747

Claeskens G, Hjort NL (2008) Model selection and model averaging. Cambridge University Press, Cambridge

Debernardi A, Baroni S, Molinari E (1995) Anharmonic phonon lifetimes in semiconductors from density-functional perturbation theory. Phys Rev Lett 75:1819–1822. https://doi.org/10.1103/PhysRevLett.75.1819

Ercole L, Marcolongo A, Umari P, Baroni S (2016) Gauge invariance of thermal transport coefficients. J Low Temp Phys 185:79–86. https://doi.org/10.1007/s10909-016-1617-6

Ercole L, Marcolongo A, Baroni S (2017) Accurate thermal conductivities from optimally short molecular dynamics simulations. Sci Rep 7:15835. https://doi.org/10.1038/s41598-017-15843-2

Foster D (1975) Hydrodynamic fluctuations, broken symmetry, and correlation functions. Benjamin, Reading

Giannozzi P, De Gironcoli S, Pavone P, Baroni S (1991) Ab initio calculation of phonon dispersions in semiconductors. Phys Rev B 43:7231–7242. https://doi.org/10.1103/PhysRevB.43.7231

Giannozzi P, Andreussi O, Brumme T, Bunau O, Nardelli MB, Calandra M, Car R, Cavazzoni C, Ceresoli D, Cococcioni M, Colonna N, Carnimeo I, Corso AD, de Gironcoli S, Delugas P, Jr RAD, Ferretti A, Floris A, Fratesi G, Fugallo G, Gebauer R, Gerstmann U, Giustino F, Gorni T, Jia J, Kawamura M, Ko HY, Kokalj A, Küçükbenli E, Lazzeri M, Marsili M, Marzari N, Mauri F, Nguyen NL, Nguyen HV, de-la Roza AO, Paulatto L, Poncé S, Rocca D, Sabatini R, Santra B, Schlipf M, Seitsonen AP, Smogunov A, Timrov I, Thonhauser T, Umari P, Vast N, Wu X, Baroni S (2017) Advanced capabilities for materials modelling with quantum espresso. J Phys Condens Matter 29:465901. https://doi.org/10.1088/1361-648X/aa8f79

Goodman NR (1963a) The distribution of the determinant of a complex Wishart distributed matrix. Ann Math Stat 34:178–180. https://doi.org/10.1214/aoms/1177704251

Goodman NR (1963b) Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). Ann Math Stat 34:152–177

Green MS (1952) Markoff random processes and the statistical mechanics of time-dependent phenomena. J Chem Phys 20:1281–1295. https://doi.org/10.1063/1.1700722

Green MS (1954) Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids. J Chem Phys 22:398–413. https://doi.org/10.1063/1.1740082

Helfand E (1960) Transport coefficients from dissipation in a canonical ensemble. Phys Rev 119:1–9. https://doi.org/10.1103/PhysRev.119.1

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871. https://doi.org/10.1103/PhysRev.136.B864

Irving JH, Kirkwood JG (1950) The statistical mechanical theory of transport processes. IV. The equations of hydrodynamics. J Chem Phys 18:817. https://doi.org/10.1063/1.1747782

Jones RE, Mandadapu KK (2012) Adaptive Green-Kubo estimates of transport coefficients from molecular dynamics based on robust error analysis. J Chem Phys 136:154102. https://doi.org/10.1063/1.3700344

Kadanoff LP, Martin PC (1963) Hydrodynamic equations and correlation functions. Ann Phys 24:419–469. https://doi.org/10.1016/0003-4916(63)90078-2

Kang J, Wang LW (2017) First-principles Green-Kubo method for thermal conductivity calculations. Phys Rev B 96:20302. https://doi.org/10.1103/PhysRevB.96.020302

Khintchine A (1934) Korrelationstheorie der stationären stochastischen Prozesse. Math Ann 109:604–615. https://doi.org/10.1007/BF01449156

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133

Kshirsagar AM (1959) Bartlett decomposition and wishart distribution. Ann Math Statist 30:239–241. https://doi.org/10.1214/aoms/1177706379

Kubo R (1957) Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. J Phys Soc Jpn 12:570–586. https://doi.org/10.1143/JPSJ.12.570

Kubo R, Yokota M, Nakajima S (1957) Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance. J Phys Soc Jpn 12:1203–1211. https://doi.org/10.1143/JPSJ.12.1203

Lampin E, Palla PL, Francioso PA, Cleri F (2013) Thermal conductivity from approach-to-equilibrium molecular dynamics. J Appl Phys 114:033525. https://doi.org/10.1063/1.4815945

Marcolongo A (2014) Theory and ab initio simulation of atomic heat transport. PhD thesis, Scuola Internazionale Superiore di Studi Avanzati, Trieste, an optional note

Marcolongo A, Umari P, Baroni S (2016) Microscopic theory and ab initio simulation of atomic heat transport. Nat Phys 12:80–84. https://doi.org/10.1038/nphys3509

Martin RM (2008) Electronic structure: basic theory and practical methods. Cambridge University Press, Cambridge

Marx D, Hutter J (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Cambridge

Matsunaga N, Nagashima A (1983) Transport properties of liquid and gaseous D2O over a wide range of temperature and pressure. J Phys Chem Ref Data 12:933–966. https://doi.org/10.1063/1.555694

Müller-Plathe F (1997) A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. J Chem Phys 106:6082–6085. https://doi.org/10.1063/1.473271

Nagar DK, Gupta AK (2011) Expectations of functions of complex wishart matrix. Acta Appl Math 113:265–288. https://doi.org/10.1007/s10440-010-9599-x

Oliveira LDS, Greaney PA (2017) Method to manage integration error in the Green-Kubo method. Phys Rev E 95:023308. https://doi.org/10.1103/PhysRevE.95.023308

Onsager L (1931a) Reciprocal relations in irreversible processes. I. Phys Rev 37:405–426. https://doi.org/10.1103/PhysRev.37.405

Onsager L (1931b) Reciprocal relations in irreversible processes. II. Phys Rev 38:2265. https://doi.org/10.1103/PhysRev.38.2265

Peierls R (1929) Zur kinetischen theorie der wärmeleitung in kristallen. Ann Phys (Berlin) 395:1055–1101. https://doi.org/10.1002/andp.19293950803

Peligrad M, Wu WB (2010) Central limit theorem for Fourier transforms of stationary processes. Ann Probab 38:2009–2022. https://doi.org/10.1214/10-AOP530

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865

Puligheddu M, Gygi F, Galli G (2017) First-principles simulations of heat transport. Phys Rev Mater 1:060802. https://doi.org/10.1103/PhysRevMaterials.1.060802

Ramires MLV, de Castro CAN, Nagasaka Y, Nagashima A, Assael MJ, Wakeham WA (1995) Standard reference data for the thermal conductivity of water. J Phys Chem Ref Data 24:1377–1381. https://doi.org/10.1063/1.555963

Schelling PK, Phillpot SR, Keblinski P (2002) Comparison of atomic-level simulation methods for computing thermal conductivity. Phys Rev B 65:144306. https://doi.org/10.1103/PhysRevB.65.144306

Sindzingre P, Gillan MJ (1990) A computer simulation study of transport coefficients in alkali halides. J Phys Condens Matter 2:7033

Sit PHL, Marzari N (2005) Static and dynamical properties of heavy water at ambient conditions from first-principles molecular dynamics. J Chem Phys 122:204510. https://doi.org/10.1063/1.1908913

Stackhouse S, Stixrude L, Karki BB (2010) Thermal conductivity of periclase (MgO) from first principles. Phys Rev Lett 104:208501. https://doi.org/10.1103/PhysRevLett.104.208501

Turney JE, Landry ES, McGaughey AJH, Amon CH (2009) Predicting phonon properties and thermal conductivity from anharmonic lattice dynamics calculations and molecular dynamics simulations. Phys Rev B 79:064301. https://doi.org/10.1103/PhysRevB.79.064301

Volz SG, Chen G (2000) Molecular-dynamics simulation of thermal conductivity of silicon crystals. Phys Rev B 61:2651–2656. https://doi.org/10.1103/PhysRevB.61.2651

Wang Z, Safarkhani S, Lin G, Ruan X (2017) Uncertainty quantification of thermal conductivities from equilibrium molecular dynamics simulations. Int J Heat Mass Trans 112:267–278. https://doi.org/10.1016/j.ijheatmasstransfer.2017.04.077

Weisstein EW (MovingAverage) Moving average. From MathWorld – a Wolfram Web Resource. http://mathworld.wolfram.com/MovingAverage.html

Weisstein EW (PolyGamma) Polygamma functions. From MathWorld – a Wolfram Web Resource. http://mathworld.wolfram.com/PolygammaFunction.html

Wiener N (1930) Generalized harmonic analysis. Acta Math 55:117–258. https://doi.org/10.1007/BF02546511

Zhang Y, Otani A, Maginn EJ (2015) Reliable viscosity calculation from equilibrium molecular dynamics simulations: a time decomposition method. J Chem Theory Comput 11:3537–3546. https://doi.org/10.1021/acs.jctc.5b00351

Zhou J, Liao B, Chen G (2016) First-principles calculations of thermal, electrical, and thermoelectric transport properties of semiconductors. Semicond Sci Technol 31:043001. https://doi.org/10.1088/0268-1242/31/4/043001

Lattice Thermal Boundary Resistance

36

Aleandro Antidormi and Luciano Colombo

Contents

Abstract

Control and modelling of thermal transport at the nanoscale has emerged either as a key issue in modern nanoscience or as a compelling demand for the ever-increasing miniaturization process in information technologies: here the thermal budget of nanodevices is basically ruled over by heat exchanges across interfaces, and the relevant physics is cast in terms of a thermal boundary resistance. In this chapter, we present a unified discussion about the fundamental knowledge developed in this framework. Starting from the most general thermodynamical description of an interface, where the driving force for heat transport is identified together with the actual location and thickness of the interface itself, we define what the thermal boundary resistance is in fact. We then delve into the most successful modelling approaches, based either on the phonon picture or on

A. Antidormi · L. Colombo (✉)
Department of Physics, University of Cagliari, Cittadella Universitaria, Monserrato(Ca), Italy
e-mail: aleandro.antidormi@dsf.unica.it; luciano.colombo@dsf.unica.it

the atomistic picture. Adopting different assumptions and employing different implementation strategies, they offer a complementary description of the physical mechanisms underlying thermal boundary resistance, and they provide useful computational protocols to predict its value in realistic systems.

1   Introduction

Since the discovery of a thermal resistance at the interface between liquid helium and a solid (Kapitza 1941), the problem of thermal transport across interfaces has gradually gained interest, eventually reaching a key role in nanomaterials physics: robust advancements in the understanding of the mechanisms ruling over heat transport across boundaries have been made, with possible impact on many intriguing technological applications.

The study of thermal boundary resistance has naturally adopted the theoretical background of the most general field of thermal transport in bulk solid-state materials as a starting point: as a matter of fact, evaluation of a thermal boundary resistance requires, at least in principle, methodologies to calculate heat fluxes in non-equilibrium systems, similarly to the widely discussed problem of thermal conductivity. In particular, from the modelling point of view, the phonon picture (including the explicit treatment of the interface scattering) has been largely employed, although at different levels of sophistication, resulting in differently accurate models. Similarly, atomistic simulations have been fruitfully exploited to analyze transport in realistic systems.

It has been recognized that the key issue in the solution of the conundrum is represented by the interface, which ultimately represents the very unique and characterizing feature of thermal boundary resistance. A rigorous definition of what an interface is in fact (together with its corresponding description and modelling) has then become crucially important in order to properly understand its effects on the thermal transport across it. The most general definition of interface is provided by non-equilibrium thermodynamics which is able to identify the essential ingredients. Moreover, challenging questions exist about the role played by the interface structure, geometry, and bonding in the overall thermal resistance. Some traits of the effect of the interface have been clearly identified in the presence of interfacial scattering, ballistic transport, and wave effects (Cahill et al. 2003; Chen 2005; Hopkins 2013). These phenomena are believed to derive from new phonon-electron interaction mechanisms at the interface which are particularly hard to model and control.

The current interest in the thermal boundary resistance is driven not only by pure scientific curiosity but also by technological needs. In particular, as the ongoing miniaturization process taking place in the information technologies has led to the manufacturing of device at the nanoscale, new challenges in the management of thermal power have emerged (Balandin 2002; Maasilta and Minnich 2014). For example, power dissipation limits the performance of electronic systems from individual microprocessors to data centers. Silicon-based microprocessors with

clock frequencies larger than a few GHz cannot even work due to on-chip power densities exceeding $100\,\mathrm{W/cm}^{-2}$. Also in radio-frequency (RF) amplifier designs using III-V group compound semiconductors, thermal high power densities are sufficient to prevent a proper device operation. These huge power densities lead to device failure and to the need for better strategies of heat dissipation and control at the nanoscale.

Similar challenges are currently encountered in the field of thermoelectric applications: today solutions for thermal waste recovery suffer an extremely low efficiency, and the quest for good thermoelectric materials is consequently particularly intense. The most promising strategies adopted to date exploit materials nanostructuring to reduce the overall thermal conductivity without jeopardizing electron transport properties (Minnich et al. 2009; Dresselhaus et al. 2007).

In addition, the junction of two materials with different thermal properties represents the fundamental building block of the so-called phononic devices (Volz et al. 2016). In analogy with the electronic diodes, joining two materials with different thermal properties allows for the realization of thermal diodes (i.e., devices acting as thermal conductors or insulators depending on the sign of the temperature difference applied). This behavior, generally called "thermal rectification" (Dames 2009), is clearly due to the interface asymmetry which reflects into a different amount of thermal energy exchanged through it when the temperature difference is reversed. Starting from these simple elementary blocks, an entire information technology based on heat could be envisioned and developed.

It is then clear how an improved ability to control thermal transport at the nanoscale would represent a fundamental progress toward new smaller and more energy-efficient nanodevices. Such considerations are compelling motivations to address thermal transport across an interface by computer modelling. Two modelling strategies have been mostly exploited in the investigation on thermal boundary resistance, corresponding to fundamentally different physical pictures, namely, the phonon picture and the atomistic picture. The former represents a valid approach inspired by the theory of heat transport in crystalline materials, and it relies on the definition of phonons as energy carriers. Being collective lattice vibrations, the reciprocal space is the natural space in which thermal transport is more intuitively described and interpreted. However, due to some oversimplifying assumptions adopted (e.g., the absence of imperfections in the materials structures), its accuracy and predictivity is rather poor. A more sophisticated and accurate calculation of thermal boundary resistances is, on the contrary, provided by atomistic simulations. Within this picture, the thermal transport properties are calculated by simulating the trajectories of individual atoms for sufficiently long times. Hence, calculations are performed in the direct space. The extremely accurate atomic-scale description of the interface and the intrinsic properties of the materials allow to obtain values of thermal boundary resistance in excellent agreement with experiments. More importantly, the possibility of introducing realistic features in the system (e.g., atom vacancies, interface roughness, interdiffusion, etc.) offers the advantage of exploring the effect of non-idealities on the heat transport, which are generally ignored in the phonon picture. Both the modelling strategies will be separately discussed below.

The chapter is organized as follows. After outlining the basic thermodynamics of thermal transport in Sect. 2, the concept of thermal boundary resistance will be properly defined in Sect. 3. Here we express it as the ratio between the temperature drop across the interface and the corresponding heat flux passing through. The rest of the chapter will be devoted to the modelling efforts of thermal boundary resistance. In particular, Sect. 4 delves into the phonon picture of transport across interfaces: fundamental observables are identified as well as the assumptions necessary to get to the results. In Sect. 5, we finally describe the atomistic picture. Questions on how to correctly identify the interface and how to compute the heat flux passing through will be answered, hence offering a complete guide to a numerical evaluation of thermal boundary resistance in realistic systems. General conclusions will close this chapter.

2 Lattice Thermal Transport: Basics

A net flux of heat (e.g., the thermal energy amount crossing a unit of surface in a unit of time) in a condensed matter system can only be observed in non-equilibrium conditions, i.e., when a thermal bias – acting as a driving force – is applied across a specimen: the resulting current of thermal energy is in fact the response of the system to such a perturbation, in the attempt to recover the initial equilibrium configuration (Lienhard IV and Lienhard V 2017). In principle, there exist two kinds of microscopic heat carriers, namely, electrons and lattice vibrations. In case the system of interest is a crystalline material, these latter ones are referred to as phonons. When dealing with nonmetallic materials, as it is assumed throughout this chapter, the electron contribution is negligibly small with respect to the lattice one.

Any elementary theory of lattice heat transport is ruled over by two basic ingredients, namely, (i) a constitutive equation characterizing the thermal force-flux relationship and (ii) some transport coefficients, which depend on the actual material system and quantify its ability to transmit thermal energy. In order to elaborate a theory suitable for computer simulations, some simplifying assumptions must be adopted in order to proceed. First of all, we assume that no mass or charge transport phenomena, as well as no chemical reactions, occur (this can be done without any loss of generality since we are focussing on pure heat transport in solid-state materials). Next, we only consider a steady-state transport configuration, where the system response to the perturbation is stationary; this implies that either the applied thermal bias does not change in time and we follow the system after a long enough transient of time, i.e., after that the system has set up its own stationary response. This implies that, although a thermal bias is present, local equilibrium holds everywhere: accordingly, provided that all relevant quantities are given as the corresponding local density, the equations of equilibrium thermodynamics still hold in each selected material portion. Finally, and most importantly, we assume that the perturbation is small: the system will activate a linear response.

If we now consider a volume region of a homogeneous system subjected to a thermal bias under the above assumptions, we can write the entropy continuity equation as (Kjelstrup and Bedeaux 2008)

$$\frac{\partial s}{\partial t} = -\frac{\partial J_s}{\partial z} + \sigma_s \tag{1}$$

where $\partial s / \partial t$ is the change of entropy density within the selected volume element, $\partial J_s / \partial z$ is the net flow of entropy through it, and σ_s is the entropy density production term provided by any possible entropy source within the same volume element Hereafter z will be the direction of the heat current. By using the local Gibbs equation $du = T ds$ (where u is the internal energy density and T is the local temperature) and the energy balance equation $\partial u / \partial t = -\partial J_h / \partial z$ (we are assuming mechanical equilibrium, and, therefore, no mechanical work must be considered), Eq. 1 is easily transformed into

$$\frac{\partial s}{\partial t} = -\frac{\partial}{\partial z}\left(\frac{J_h}{T}\right) + J_h \frac{\partial}{\partial z}\left(\frac{1}{T}\right) \tag{2}$$

In this equation, J_h is the heat flux, i.e., the amount of thermal energy crossing a unit area in a unit of time. By comparing Eqs. 1 and 2, we can immediately identify the σ_s term which, in the present case of pure heat transport, can be expressed by the Onsager relation $\sigma_s = J_h X_h$ where X_h is the generalized force associated to J_h. This allows for identifying the actual force driving thermal transport as

$$X_h = \frac{\partial}{\partial z}\left(\frac{1}{T}\right) \tag{3}$$

Within the linear response regime we have assumed, we can further state $X_h = r J_h$, thus introducing the *Onsager resistivity coefficient r*: this is exactly the material-specific parameter that has been indicated as a basic ingredient of the present elementary transport theory. We can now identify the second ingredient by inserting the Onsager resistivity in Eq. 3 and getting the following force-flux equation

$$\frac{\partial}{\partial z}\left(\frac{1}{T}\right) = r J_h \tag{4}$$

corresponding to the *constitutive equation for heat transport*. We remark that according to Eq. 4 the actual *thermodynamic driving force for thermal transport is the gradient of an inverse temperature* (Kittel and Kroemer 1980). By developing the z-derivative of the inverse temperature, we get $\partial T / \partial z = -T^2 r J_h$ which can be recast in the form of the well-known *Fourier law*

$$J_h = -\kappa \frac{\partial T}{\partial z} \tag{5}$$

where the thermal conductivity κ (measured in $\mathrm{Wm^{-1}K^{-1}}$ units) is related to the inverse of the Onsager resistivity $\kappa = (T^2 r)^{-1}$. We remark that this latter relation establishes the link between the Onsager and the Fourier pictures: the former being rooted in robust non-equilibrium thermodynamics and the latter being fully phenomenological.

3   The Concept of Thermal Boundary Resistance

Let us now move to the case of two materials, with unlike thermal transport properties (i.e., with different thermal conductivity), meeting at an interface across which a thermal bias is applied. While for a homogeneous system a linear variation of temperature throughout the whole specimen occurs in steady-state condition, a sudden temperature drop ΔT is experimentally observed at the heterojunction in a similar condition: the situation is conceptualized in Fig. 1 where, at least for the moment, we do not need to specify the actual temperature profile in the interface region but, rather, we simply assign its average value hereafter referred to as *interface temperature* T_{int}. Furthermore, T_{left} and T_{right} are the local temperature just at the left and right boundary of the interface region.

In this situation Eq. 1 is recast in the form (Kjelstrup and Bedeaux 2008)

$$\frac{\partial s^{\mathrm{int}}}{\partial t} = -\left(J_{s,\mathrm{left}} - J_{s,\mathrm{right}}\right) + \sigma_s^{\mathrm{int}} \tag{6}$$
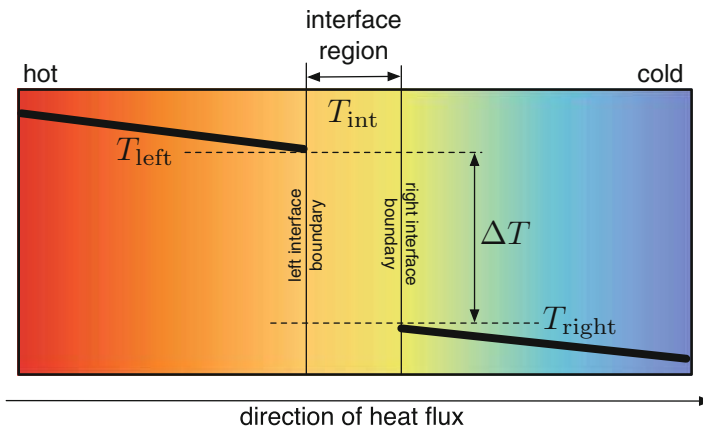
Fig. 1 Idealized scheme for the definition of the thermal boundary resistance concept. The bold black line represents the local temperature profile along the direction of heat flux. The interface region is marked by vertical light lines, while ΔT indicates the temperature drop occurring between the left and the right boundaries of the interface, respectively, at temperatures T_{left} and T_{right}

where the superscript "int" indicates an interface quantity, while $J_{s,\text{left}}$ and $J_{s,\text{left}}$ are the entropy density flux through the indicated interface boundaries. Similarly to the case treated in the previous section, we easily obtain σ_s^{int} as

$$\sigma_s^{\text{int}} = J_q \left(\frac{1}{T_{\text{int}}} - \frac{1}{T_{\text{left}}} \right) + J_q \left(\frac{1}{T_{\text{right}}} - \frac{1}{T_{\text{int}}} \right) \tag{7}$$

proving that *two flux-force equations are actually needed for the interface problem*, one for each boundary

$$\left(\frac{1}{T_{\text{int}}} - \frac{1}{T_{\text{left}}} \right) = r_{\text{left}} J_h \qquad \left(\frac{1}{T_{\text{right}}} - \frac{1}{T_{\text{int}}} \right) = r_{\text{right}} J_h \tag{8}$$

and each one defining its own Onsager resistivity coefficient.

This analysis allows to treat the *series of two Onsager resistances*

$$R_{\text{TBR}} = T_{\text{int}}^2 \left(r_{\text{left}} + r_{\text{right}} \right) = T_{\text{int}}^2 \frac{1}{J_h} \left(\frac{1}{T_{\text{right}}} - \frac{1}{T_{\text{left}}} \right) = \frac{T_{\text{int}}^2}{T_{\text{right}} T_{\text{left}}} \frac{T_{\text{left}} - T_{\text{right}}}{J_h} \tag{9}$$

as the net thermal boundary resistance (TBR) the heat current must pass through. The $(T_{\text{left}} - T_{\text{right}})/J_h$ factor on the right-hand side of the above equation is commonly called *Kapitza resistance* R_{Kapitza} (Pollack 1969; Swartz and Pohl 1989; Cahill et al. 2003, 2014; Kapitza 1941)

$$R_{\text{kapitza}} = \frac{T_{\text{left}} - T_{\text{right}}}{J_h} = \frac{\Delta T}{J_h} \tag{10}$$

and it is typically used to characterize interface thermal transport.

In summary, non-equilibrium thermodynamics provides the proper framework to define the Kapitza resistance, leaving us with problem of calculating the ratio between the temperature drop at the interface and the corresponding steady-state thermal current J_h. There exist basically two strategies to accomplish this task either based on calculations of lattice vibrational properties or on atomistic simulations. The former corresponds to a phonon picture of thermal boundary resistance; the latter offers an atomistic perspective for transport across the interface. The mentioned modelling pictures will be treated separately in the two following sections.

4 The Phonon Picture for the Thermal Boundary Resistance

In this section we present the phonon-based semiclassical formalism of thermal transport across an interface and retrace the main steps to predict the thermal boundary resistance. As commented above, we shall only consider the interfaces

between nonmetals, for which the electron contribution to heat flux can be neglected with respect to the lattice contribution.

The prototypical system on which we will focus is shown in Fig. 1. In the following, the materials involved will be assumed to be perfectly crystalline and the interface region small enough to be well approximated by the plane $z = 0$. The heat flux can be expressed in terms of quanta of vibrational energy, the phonons, on the left and right parts of the interface. The most simple and successful picture consists in attributing the onset of thermal resistance to phonon scattering at the interface: a phonon incident on the interface has a certain probability of being transmitted to the other side. Such a probability of transmission generally depends on the phonon mode, its polarization p, its wave vector $\mathbf{k}$, and its frequency $\omega_p(\mathbf{k})$. The net heat transfer across the interface is due to the net balance of incoming, transmitted, and reflected energy. Since energy is carried by phonons, this reflects into a balance for the incoming, transmitted, and reflected phonon flux (Swartz and Pohl 1989; Chen 2005). The mechanism of phonon transmission and reflection is illustrated in Fig. 2.

Under the simplifying assumptions of negligible inelastic scattering at the interface and no phonon-phonon interactions, the heat flux incident from the side α ($\alpha = 1, 2$) on the interface can be written as

$$J_\alpha = \sum_p \int_{\mathbf{k}'} \frac{1}{4\pi} \hbar \omega_p(\mathbf{k}) f_\alpha(\omega_p(\mathbf{k}), T) \, \mathbf{v}_p(\mathbf{k}) \cdot \mathbf{n} \, d\mathbf{k} \qquad (11)$$

where $\mathbf{n}$ is the unit vector normal to the interface, $\mathbf{v}_p(\mathbf{k})$ is the group velocity of the phonon with wavevector $\mathbf{k}$, and $f_i(\omega, T)$ is the Bose-Einstein distribution
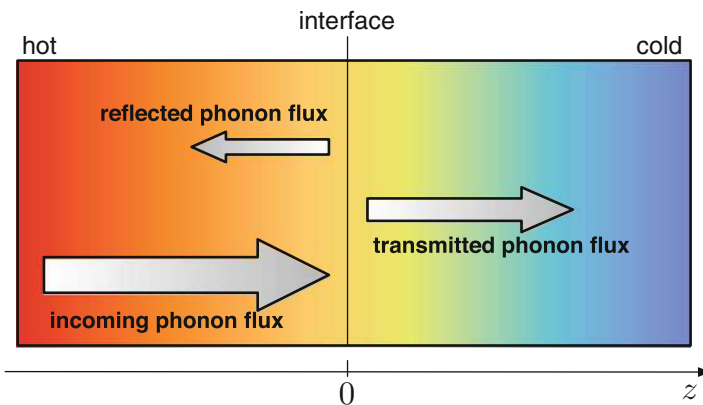
Fig. 2 Illustration of the physical mechanisms of thermal boundary resistance. TBR is explained in terms of heat fluxes impinging on the interface: part of the flux coming from hot side is effectively transmitted to the cold one; the remaining is back-reflected. Analogously, a similar flux comes from cold side (not shown here), part of which will be transmitted across the interface and the rest back-reflected

function. The integral extends over all the phonon wavevectors $\mathbf{k}'$ in the Brillouin zone actually incident on the separation plane, i.e., such that the corresponding group velocity $\mathbf{v}$ has a positive (negative) component along the z-axis for $\alpha = 1$ (2). Finally, $\sum_p$ indicates the sum over all possible phonon polarizations. Changing the integration variable to the frequency ω and introducing angular coordinates (the azimuth angle φ is defined by the vector $\mathbf{n}$ and a direction normal to z, while the θ angle is defined in the plane normal to the azimuthal direction), the formula above can be recast in the form

$$J_\alpha = \sum_p \int_0^{\omega_{max}} \int_{\varphi=0}^{2\pi} \int_{\theta=-\pi/2}^{\pi/2} \hbar\omega_p \, f_\alpha(\omega_p, T) \, v_p(\omega_p) \cos(\theta) \sin(\theta) D_i(\omega_p) d\theta d\varphi d\omega_p$$

(12)

where the phononic density $D_\alpha(\omega_p)$ of the side α has been introduced.

Any given phonon has a certain probability $t_{\alpha \to \beta}(\mathbf{k})$ of being transmitted from side α to side β. Multiplying the previous expression by this factor, we obtain that the heat flux from one side to the other is given by

$$J_{\alpha \to \beta} = \sum_p \int_{\mathbf{k}'} \frac{1}{4\pi} \hbar\omega_p(\mathbf{k}) \, f_\alpha(\omega_p(\mathbf{k}), T) \, t_{\alpha \to \beta}(\mathbf{k}) \, \mathbf{v}_p(\mathbf{k}) \cdot \mathbf{n} \, d\mathbf{k}$$

(13)

The resulting net heat flux across the surface is then readily computed as

$$J_h = J_{1 \to 2} - J_{2 \to 1}$$

(14)

As we have already pointed out, the transmission probability depends on the microscopic details of the interface, and its calculation is a major theoretical task. In order to proceed, an average transmission probability can be identified as $\bar{t}_{\alpha \to \beta} = J_{\alpha \to \beta}/J_\alpha$. How to compute its value will be discussed in the next section: here we assume that $\bar{t}_{\alpha \to \beta}$ is known. Then, the overall heat flux, when a temperature difference $\Delta T = T_2 - T_1$ is applied, can be equivalently expressed by

$$J_h = [J_1 \cdot \bar{t}_{1 \to 2}]_{T_1} - [J_2 \cdot \bar{t}_{2 \to 1}]_{T_2}$$

(15)

Recalling that at equilibrium ($\Delta T = 0$), the above quantity is zero, i.e.

$$[J_1 \cdot \bar{t}_{1 \to 2}]_{T_1 = T} = [J_2 \cdot \bar{t}_{2 \to 1}]_{T_2 = T}.$$

(16)

one can finally obtain an expression for the thermal boundary resistance $R_{\text{kapitza}} = \frac{\Delta T}{J_h}$ by linearly expanding the J_h formula for a small temperature offset $\Delta T \to 0$, yielding

$$R_{\text{Kapitza}} = \lim_{\Delta T \to 0} \left[\frac{J_h|_{T+\Delta T} - J_h|_T}{\Delta T} \right]^{-1} = \left[\frac{\partial}{\partial T} [J_{1\to 2} \cdot \bar{t}_{1\to 2}] \right]^{-1}$$

$$= \left[\frac{\partial}{\partial T} [J_{2\to 1} \cdot \bar{t}_{2\to 1}] \right]^{-1} \tag{17}$$

This is the formula for the thermal boundary resistance in the phonon picture that we were looking for. Under the adopted assumptions, R_{Kapitza} can in principle be calculated by only making reference to one side of the interface; the other side still enters the game through the transmission probability $\bar{t}$.

According to Eq. 17 and the underlying concepts, a thermal boundary resistance between two materials arises whenever a difference in the phonon dispersion curves is found. This creates, indeed, an imbalance of heat carriers between the two interacting materials and, hence, a non-negligible heat flux when a temperature gradient is applied. In order to get a deeper insight, it is convenient to assume a linear dispersion relation with sound speed v and transmission probabilities not depending on the temperature. Under these assumptions, we obtain

$$R \propto \left[\sum_p \bar{t}_{1\to 2} \, C_{1,p} \, v_{1,p} \right]^{-1} \tag{18}$$

where $C_{1,p}$ is the single-mode contribution to the heat capacity (per unit volume) of material 1. It is now clear how materials whose dispersion relation gives large values for Cv are generally characterized by smaller thermal boundary resistance. Given that the heat capacity depends on the temperature, a variation of its value with T is expected, in addition to the T-dependence imposed in Eq. 11 by the Bose-Einstein distribution. As reported in Swartz and Pohl (1989), one finds $R \propto T^{-3}$ for T smaller than the Debye temperature, while R reaches a plateau for higher temperatures.

Another key factor displayed in Eq. 17 is represented by the transmission probabilities $t(\mathbf{k})$, expressing the capability of each single phonon mode to transmit its energy across the interface. Materials with strong interfacial coupling have generally large values of probabilities and a consequently smaller thermal boundary resistance. The calculation of the transmission probabilities represents a formidable task, and several models have been proposed employing different approximations. The most used and successful modelling attempts are the acoustic mismatch model and diffuse mismatch model that will be presented in the next sections.

4.1   The Acoustic Mismatch Model

The acoustic mismatch model is based on a three-fold assumption, namely, (i) phonons are treated as plane waves, (ii) the medium in which they propagate

is a continuum, and (iii) the interface is as negligibly thick as a plane. Such approximations are expected to be valid only in the regime of low temperatures, where the vibrational spectrum is dominated by phonons with wavelength larger than the interatomic spacing. Equations of continuum acoustics are then applied to phonon waves: a wave incident on the interface generates a reflected and a refracted wave, each of them being the superposition of one longitudinal and two transverse phonon waves. The angles of incidence and refraction θ_{in} and θ_{trans} are governed by Snell's law

$$v_\alpha^{L/T} \sin \theta_{in} = v_\beta^{L/T} \sin \theta_{trans} \tag{19}$$

where $v_\alpha^{L/T}$ is the acoustic longitudinal (transverse) phonon velocity in medium α (see Fig. 3).

Under the continuum ansatz, the transmission probabilities are calculated by the acoustic analog of the Fresnel equations, which translate into a set of boundary conditions. The normal and tangent components of the displacement must be continuous at the interface; moreover, the normal component of the stress tensor must also be continuous. In this way, the transmission probabilities can be computed for each phonon mode and incidence angle. It is noteworthy that, in complete analogy with electromagnetic waves, no guess on the interface energy budget is required (i.e., nor elastic nor inelastic scattering is described by the acoustic mismatch model). Although simple and clear in the underlying physics, the continuum model just described is known to have poor accuracy, generally predicting values of thermal boundary resistance smaller than experimental values at room temperature (Chen 2005).
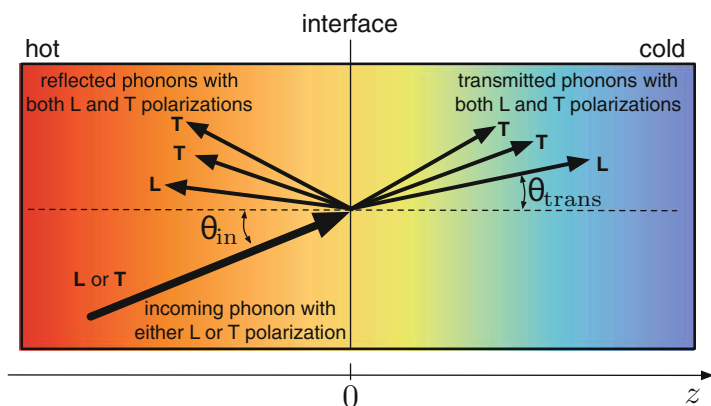
Fig. 3 Schematic of the many possibilities within the acoustic mismatch model for phonon incident on the interface. When a longitudinal phonon (L) reaches the interface, a superposition of one longitudinal (L) and two transverse waves (T) are transmitted and reflected

4.2 The Diffuse Mismatch Model

The diffuse mismatch model is mainly addressed to describe the effect of scattering of phonons at the interface. The fundamental assumption is that all phonons incident on the interface are diffusively scattered and acoustic correlations between incoming and transmitted phonons are completely destroyed by diffuse scattering. Differently stated, when a phonon impinges on the interface, it is scattered into another phonon which is almost independent from the incoming mode. Hence, any info about the incoming phonon (like its direction and polarization) is lost during the interface scattering event. The only retained information is about the incoming single-mode frequency ω, which is conserved since elastic scattering is assumed.

The probability that the phonon will scatter into a given side of the system is then chosen to be proportional to the vibrational density of states in that side $D_\alpha(\omega)$. The diffuse mismatch model is then more appropriate to describe the transmission probabilities at high temperatures and rough interfaces. At room temperature, indeed, the phonons involved in the heat flux wavelengths which are small compared to the interatomic spacing and surface roughness and can thus hardly see interfaces as flat.

Specifically, the transmission probability for a phonon from side α to side β does not depend on the phonon $\mathbf{k}$, $t_{\alpha\to\beta}(\mathbf{k}, \omega) = t_{\alpha\to\beta}(\omega)$; moreover, the probability for a phonon to be transmitted from side α to side β must be equal to the probability of a phonon in side β to be reflected at the interface, i.e.

$$t_{\alpha\to\beta}(\omega) = 1 - t_{\beta\to\alpha}(\omega) \tag{20}$$

Under these assumptions, the actual values of transmission probabilities can be determined inserting Eq. 20 into Eq. 16. For Debye solids at sufficiently low temperatures, one finds

$$t_{\alpha\to\beta}(\omega) = \frac{\sum_p v_{\beta,p}^{-2}}{\sum_p v_{\alpha,p}^{-2} + \sum_p v_{\beta,p}^{-2}} \tag{21}$$

where $v_{\beta,p}$ is the frequency-dependent group velocity of phonons with polarization p. The diffuse model just presented generally predicts higher thermal boundary resistance than the acoustic model for quite similar materials, smaller resistance for very dissimilar ones (Hopkins 2013).

Variations of the diffuse mismatch model have been proposed to include also interface features: in particular, the effect of interface roughening and the effect of interfacial inelastic scattering.

It is important to recognize that both the acoustic and diffuse models show a rather scarce predictive capability (Stoner and Maris 1993; Swartz and Pohl 1987; Hopkins 2013); notwithstanding, their simplicity and conceptual clarity is efficient enough to highlight the basic physical processes at the core of the thermal boundary resistance at crystalline solid interfaces.

5   The Atomistic Picture for the Thermal Boundary Resistance

The phonon picture developed at different levels in the previous section is based on the treatment of crystalline atomic motions as collective vibrational modes. This picture is no longer valid when translational invariance is lost, as found in many situations of practical interest, including (but not limited to) an interface between noncrystalline (i.e., amorphous or nanocrystalline) materials, diffused or graded interfaces, nonperiodic distributions of defects (like engineered alloying profiles), and nonperiodic multiple interfaces. When the phonon picture is unfeasible (or not accurate enough due to the underlying approximations), we can calculate the thermal boundary resistance by a more direct approach based on molecular dynamics (MD) simulations (Frenkel and Smit 1996).

MD simulations neither assume crystal periodicity nor require the calculation of any phonon property. They make no assumption about the actual scattering mechanisms affecting heat carriers at the interface, while vibrational anharmonicity is fully taken into account as well as the actual atomistic structure of the interface. Rather, MD simulations are based on the non-equilibrium thermodynamics picture described above, and, therefore, the determination of the Kapitza resistance is reduced to separately evaluate the temperature drop ΔT across the interface and the steady-state heat flux J_h passing through. This twofold task requires different simulation strategies, which are addressed below. The overall validity of any MD calculation depends on the reliability, accuracy, and transferability of the adopted force field, an issue which is not address here (Finnis 2003).

5.1   Calculating the Interface Temperature Drop ΔT

In order to attempt a calculation of ΔT, we need at first to unambiguously define the location and the thickness of the interface: to this aim we can follow the *Gibbs construction* (Kjelstrup and Bedeaux 2008).

By selecting a material property showing different values in the two bulk-like regions far away from the interface, we can evaluate its profile $P(z)$ along the same z-direction selected for applying the thermal bias (we understand that such a direction crosses the interface). This is typically accomplished by computing the local value of that property over a set of suitably thin material slabs at sequential z-positions, resulting in a plot like Fig. 4 (top) where $P(z)$ is shown as a thick black line. For instance, in the case an interface between two lattice mismatched semiconductors $P(z)$ could be the interplanar lattice constant along z. Alternatively, $P(z)$ could represent the actual content of a dopant, or a given chemical species, or any other structural defect. Finally, for a crystalline/amorphous interface, $P(z)$ could represent the local average atomic coordination.

We agree in defining the interface as that region where $P(z)$ differs from the pure material values (hereafter referred to as P_{left} and P_{right}, respectively) by some
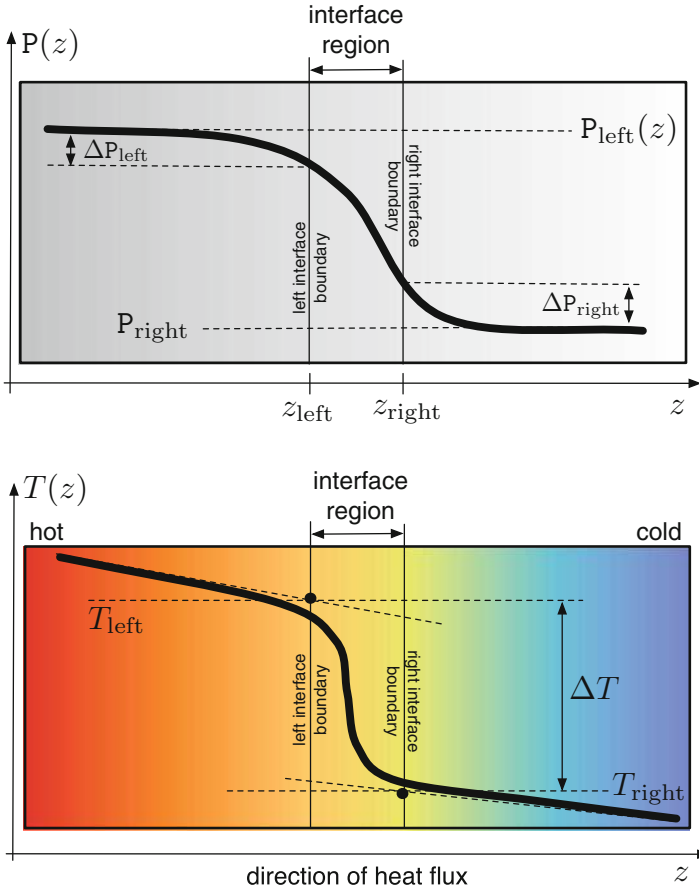
Fig. 4 Top: Gibbs construction: a suitable material property $P(z)$ is calculated across the heterojunction, and the interface boundaries are defined according to the criteria described in the text. The gray color shading is a pictorial representation of the local value of the $P(z)$ property. Bottom: determination of the T_{left} and T_{righ} temperatures under a steady-state thermal bias conduction. The temperature field is represented by colors

arbitrary amount. In practice, the position of the left interface boundary is set at z_{left} where the offset $\Delta P_{left} = P_{left} - P(z_{left})$ is typically 2–5% of P_{left}, and similarly for the right boundary. This procedure also defines the actual thickness of the interface region. When applied to real solid-state systems, the Gibbs construction provides evidence that, while $P(z)$ is normally a well-behaved function across the interface region (i.e., it shows a monotonic trend with no abrupt discontinuities), such a material portion is neither symmetrically extended over the two facing leads nor is centered at their nominal boundary: in other words, we have learned that the thermodynamical interface does not coincide with the nominal interface. The

intriguing point is that the Kapitza resistance is determined by a temperature drop across the thermodynamical interface, as detailed below. It has been also shown that the interface region can be treated as an autonomous thermodynamical system (Rurali et al. 2016) with its own temperature T_{int}.

The Gibbs construction provides a robust definition of the interface suitably tailored for the calculations we are interested in. If we apply a temperature bias along z and plot the temperature profile $T(z)$ once the steady-state regime has been reached, we typically obtain what pictured in Fig. 4, bottom: while $T(z)$ displays a linear variation in the two bulk-like regions (with a different slope, the two materials differing in thermal conductivity), it shows a typical sigmoidal behavior over the interface region. This figure directly defines two relevant temperatures in our problem, namely, the left T_{left} and right T_{right} temperatures, just at the boundaries of the interface region: they are straightforwardly calculated at the intercept between the extrapolation of linear trends and the interface boundaries previously defined. We also get a clear indication on how to evaluate the *interface temperature* T_{int}: it is easily evaluated within a MD simulation by simply computing the average kinetic energy among all atoms placed in the interface region.

5.2   Calculating the Heat Flux

5.2.1   Setting up the Thermal Bias Across an Interface

The situation we are interested in is a non-equilibrium one, and, therefore, the first task to accomplish is twofold, namely, (i) applying a thermal bias to the simulation cell and (ii) ageing the system until a steady-state regime of thermal conduction is eventually reached. To these aims, once that the z-direction is selected for heat conduction, a simulation cell as long as L_z is assembled and thermalized through conventional thermostatting (Frenkel and Smit 1996) at a suitable temperature T_0. Next, the opposite terminal ends of the simulation cell are coupled to two thermostats at temperature T_{hot} and T_{cold}, respectively, in such a way that $T_{cold} < T_0 < T_{hot}$. They act, respectively, as a hot and cold reservoir. The system is then aged by keeping both ends at constant temperature while integrating the equations of motions of all atoms in the core region without any thermostatting. This simulation protocol is commonly referred to as the "direct method" (Schelling et al. 2002; He et al. 2012; Sellan et al. 2010; McGaughey and Kaviany 2006).

In order to assess the steady-state condition, the local temperature profile $T(z, t)$ is calculated during the system ageing. This is done by recording the average kinetic temperature over a suitable set of thin material slabs distributed throughout the whole sample: the steady-state condition is proclaimed as soon as the $T(z, t)$ profile does not further evolve in time. At this stage the first needed input to evaluate the Kapitza resistance, namely, the temperature drop ΔT, is straightforwardly obtained by applying the Gibbs construction: T_{left} and T_{right} are easily obtained from the resulting steady-state temperature profile.

5.2.2 Evaluating Heat Flux Through Atomic Trajectories

The evaluation of the heat flux is a very subtle problem. Its microscopic definition is provided by the virial formulation (Fan et al. 2015)

$$\mathbf{J}_h = \frac{d}{dt} \sum_i \mathbf{r}_i E_i = \sum_i \left[\mathbf{v}_i E_i + \mathbf{r}_i \frac{d}{dt} E_i \right] \tag{22}$$

where $\mathbf{r}_i$ and $\mathbf{v}_i$ are the atomic positions and velocities (i labels atoms), while $E_i = 1/2 m_i v_i^2 + U_i$ are the corresponding single-atom energies (m_i are atomic masses and U_i is the on-site potential). The actual calculation of $\mathbf{J}_h$ is made difficult by the fact that, while atomic trajectories and velocities are directly provided on-the-fly during the MD run, single-atom energies are difficult to calculate: as a matter of fact, this task is straightforward only in the case of two-body interaction potentials which, however, are mostly inadequate to describe real materials.

The two contributions appearing on the last right-hand side of Eq. 22 are, respectively, known as the *convective contribution* $\mathbf{J}_h^{\mathrm{conv}} = \sum_i \mathbf{v}_i E_i$ and the *potential contribution*

$$\mathbf{J}_h^{pot} = \sum_i \mathbf{r}_i \frac{d}{dt} E_i = \sum_i \mathbf{r}_i \left[\mathbf{F}_i \cdot \mathbf{v}_i \right] + \sum_i \mathbf{r}_i \frac{dU_i}{dt} \tag{23}$$

where $\mathbf{F}_i(t)$ is the net force on i-th atom at time t. They correspond to similar terms in the Einstein formalism for transport coefficients (Kinaci et al. 2012), and it is similarly found that the convective term is negligibly small in solid-state materials: this remark is consistent with the approximations here adopted to describe pure thermal transport.

The key idea (Fan et al. 2015) to calculate the potential contribution $\mathbf{J}_h^{\mathrm{pot}}$ is that for any particle pair (i, j) we can define a pairwise force $\mathbf{F}_{ij} = -\mathbf{F}_{ji} = \partial U_i/\partial \mathbf{r}_{ij} - \partial U_j/\partial \mathbf{r}_{ji}$ based on the natural choice $U_i = 1/2 \sum_{j \neq i} U_{ij}$, where U_{ij} only depends on the relative distance vectors $\mathbf{r}_{ij}$ originating from particle i. This allows to cast in the most general form the potential contribution to the heat flux as

$$\mathbf{J}_h^{\mathrm{pot}}(t) = \sum_i \sum_{j \neq i} \mathbf{r}_{ij}(t) \left[\frac{\partial U_j(t)}{\partial \mathbf{r}_{ji}(t)} \cdot \mathbf{v}_i(t) \right] \tag{24}$$

where all quantities can be straightforwardly calculated during a MD run, including the instantaneous value of the relative (i, j) distance $\mathbf{r}_{ij}(t) \equiv \mathbf{r}_j(t) - \mathbf{r}_i(t)$. This result is sometimes referred to as the Hardy formula (Hardy 1963).

5.2.3 Evaluating Heat Flux Through Work Rates

An alternative evaluation of the heat flux J_h consists in calculating the rate of thermostatting work separately done by the hot and cold thermostats placed at the terminal ends of the simulated sample. Of course these two quantities must be the same in a steady-state regime, but work is more directly calculated in

MD for any specific kind of adopted force field and implemented thermostat, without the ambiguities reported above for the calculation of the heat flux. In practice, the instantaneous energies injected/extracted by the hot/cold reservoir into/from the system must be computed during the ageing of the system. Their time derivatives have opposite sign and are directly related to the injected J_h^{in} and extracted J_h^{ex} thermal current, respectively. The steady state can be proclaimed when $|J_h^{\text{in}}| = |J_h^{\text{ex}}|$, and it is corresponding set $J_h = |J_h^{\text{in,ex}}|$ to be used in the formula for the Kapitza resistance. The deviation from the asymptotic relative difference $|J_h^{\text{in}} - J_h^{\text{ex}}|/|J_h^{\text{in}}|$ can be usefully adopted for error estimate of the heat current.

Finally, it is possible to design an alternative scheme where it is neither needed to compute microscopically the heat flux nor use thermostats. This second feature is interesting since concerns have been raised about the dependence of the resulting temperature profile upon the kind of heat reservoirs implemented in the MD run (Chen et al. 2012). We can in fact generate the non-equilibrium configuration we are interested in simply by imposing an energy current through the system which generates a nonconstant temperature profile: this simulation protocol is commonly referred to as the "inverse method" (Schelling et al. 2002; Müller-Plathe 1997).

The first way to implement the method consists in starting from a periodic system previously equilibrated at a suitable temperature T^0. Then, by defining two different space regions, hereafter referred to as "hot" and "cold" reservoir, the kinetic energy of any atom there contained is rescaled by the same amount $\pm \Delta e_{\text{kin}}$ during a time interval Δt. The resulting energy (heat) flux imposed to the system is $J_h = \Delta e_{\text{kin}}/2\Sigma \Delta t$, where Σ is the cross section of the simulated system. Eventually, a temperature profile is established, from which the temperature drop at the interface is easily obtained following the methods discussed above. Alternatively, the coldest atom with velocity $\mathbf{v}_{\text{coldest}}$ and hottest atom with velocity $\mathbf{v}_{\text{hottest}}$ are selected, respectively, in the hot and cold reservoir. Their velocities are then swapped. By iterating such a $\mathbf{v}_{\text{coldest}} \leftrightarrow \mathbf{v}_{\text{hottest}}$ swap over a suitable time interval Δt, we effectively impose an energy (heat) flux $J_h = m/2 \sum_{\text{swaps}}(v_{\text{hottest}}^2 - v_{\text{coldest}}^2)/2\Sigma \Delta t$, where m is the atomic mass.

In both cases the steady state can be proclaimed by detecting a stationary temperature profile or by evaluating the work done in the reservoir regions. The Kapitza resistance is eventually calculated by identifying J_h with the energy flux due to rescaling or swapping, while ΔT is evaluated through the Gibbs construction.

6 Conclusion

In this chapter we provided a unified description of the thermal boundary resistance. The most general definition of the problem is realized in the thermodynamics framework, which allows to identify the generalized forces responsible for heat transport and to derive a definition of the thermal boundary resistance as the ratio of temperature gradient and the heat flux passing through the interface. The actual calculation of these quantities can be accomplished adopting two different physical

pictures: the phonon-based picture and the atomistic one. The former, applicable mostly to interfaces between crystalline materials, provides a description of the relevant physical mechanisms in terms of phonons impinging onto the interface and transmission probabilities of their energy. Such a model, although simple and clear, has limited validity (it applies only to crystalline materials and requires quite strong assumptions on the form of the transmission probabilities). As a consequence, it poorly reproduces experimental values. The second approach, based on atomistic simulations, overcomes these limitations by offering a reliable strategy to define the thermal interface in real systems and to properly evaluate heat fluxes and temperature gradients from molecular-dynamics simulations. This strategy has several advantages including the fact that it can be used for all kinds of interfaces, with no constraints imposed on the particular structure of the materials. Although it lacks a clear physical interpretation, it is well suited to study realistic systems and to analyze the effects of imperfections and defects, which are always present in real devices, on the overall thermal transport.

References

Balandin AA (2002) Nanoscale thermal management. IEEE Potentials 21(1):11–15. https://doi.org/10.1109/45.985321

Cahill DG, Ford WK, Goodson KE, Mahan GD, Majumdar A, Maris HJ, Merlin R, Phillpot SR (2003) Nanoscale thermal transport. J Appl Phys 93(2):793–818. https://doi.org/10.1063/1.1524305

Cahill DG, Braun PV, Chen G, Clarke DR, Fan S, Goodson KE, Keblinski P, King WP, Mahan GD, Majumdar A, Maris HJ, Phillpot SR, Pop E, Shi L (2014) Nanoscale thermal transport. II. 2003–2012. Appl Phys Rev 1(1):011305. https://doi.org/10.1063/1.4832615

Chen G (2005) Nanoscale energy transport and conversion: a parallel treatment of electrons, molecules, phonons, and photons. MIT-Pappalardo series in mechanical engineering. Oxford University Press. https://books.google.it/books?id=M3n3lUJpYDYC

Chen J, Zhang G, Li B (2012) Thermal contact resistance across nanoscale silicon dioxide and silicon interface. J Appl Phys 112(6):064319. https://doi.org/10.1063/1.4754513

Dames C (2009) Solid-state thermal rectification with existing bulk materials. J Heat Trans 131(6):061301

Dresselhaus M, Chen G, Tang M, Yang R, Lee H, Wang D, Ren Z, Fleurial JP, Gogna P (2007) New directions for low-dimensional thermoelectric materials. Adv Mater 19(8):1043–1053. https://doi.org/10.1002/adma.200600527

Fan Z, Pereira LFC, Wang HQ, Zheng JC, Donadio D, Harju A (2015) Force and heat current formulas for many-body potentials in molecular dynamics simulations with applications to thermal conductivity calculations. Phys Rev B 92:094301. https://link.aps.org/doi/10.1103/PhysRevB.92.094301

Finnis M (2003) Interatomic forces in condensed matter. Oxford University Press, Oxford. https://books.google.it/books?id=RNFmvsWQmZoC

Frenkel D, Smit B (1996) Understanding molecular simulation: from algorithms to applications, 1st edn. Academic Press, Orlando

Hardy RJ (1963) Energy-flux operator for a lattice. Phys Rev 132:168–177. https://link.aps.org/doi/10.1103/PhysRev.132.168

He Y, Savic I, Donadio D, Galli G (2012) Lattice thermal conductivity of semiconducting bulk materials: atomistic simulations. Phys Chem Chem Phys 14:16209–16222. https://doi.org/10.1039/C2CP42394D

Hopkins PE (2013) Thermal transport across solid interfaces with nanoscale imperfections: effects of roughness, disorder, dislocations, and bonding on thermal boundary conductance. ISRN Mech Eng 2013, 682586. https://doi.org/10.1155/2/682586

Kapitza PL (1941) Heat transfer and superfluidity of helium II. Phys Rev 60:354–355. https://link.aps.org/doi/10.1103/PhysRev.60.354

Kinaci A, Haskins JB, Cagin T (2012) On calculation of thermal conductivity from Einstein relation in equilibrium molecular dynamics. J Chem Phys 137(1):014106. https://doi.org/10.1063/1.4731450

Kittel C, Kroemer H (1980) Thermal physics. Freeman, W.H. and Co., New York. https://books.google.it/books?id=c0R79nyOoNMC

Kjelstrup S, Bedeaux D (2008) Non-equilibrium thermodynamics of heterogeneous systems. Series on advances in statistical mechanics vol 16. World Scientific. https://books.google.it/books?id=yAUxz6tQKl0C

Lienhard IV JH, Lienhard V JH (2017) A heat transfer textbook, 4th edn. Phlogiston Press, Cambridge, MA. http://ahtt.mit.edu, version 2.11

Maasilta I, Minnich AJ (2014) Heat under the microscope. Phys Today 67(8):27–32. https://doi.org/10.1063/PT.3.2479

McGaughey A, Kaviany M (2006) Phonon transport in molecular dynamics simulations: formulation and thermal conductivity prediction. Adv Heat Tran 39(C):169–255. https://doi.org/10.1016/S0065-2717(06)39002-8

Minnich AJ, Dresselhaus MS, Ren ZF, Chen G (2009) Bulk nanostructured thermoelectric materials: current research and future prospects. Energy Environ Sci 2:466–479. https://doi.org/10.1039/B822664B

Müller-Plathe F (1997) A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. J Chem Phys 106(14):6082–6085. https://doi.org/10.1063/1.473271

Pollack G (1969) Kapitza resistance. Rev Mod Phys 41:48–81. https://link.aps.org/doi/10.1103/RevModPhys.41.48

Rurali R, Colombo L, Cartoixa X, Wilhelmsen O, Trinh TT, Bedeaux D, Kjelstrup S (2016) Heat transport through a solid-solid junction: the interface as an autonomous thermodynamic system. Phys Chem Chem Phys 18:13741–13745. https://doi.org/10.1039/C6CP01872F

Schelling PK, Phillpot SR, Keblinski P (2002) Comparison of atomic-level simulation methods for computing thermal conductivity. Phys Rev B 65:144306. https://link.aps.org/doi/10.1103/PhysRevB.65.144306

Sellan DP, Landry ES, Turney JE, McGaughey AJH, Amon CH (2010) Size effects in molecular dynamics thermal conductivity predictions. Phys Rev B 81:214305. https://link.aps.org/doi/10.1103/PhysRevB.81.214305

Stoner RJ, Maris HJ (1993) Kapitza conductance and heat flow between solids at temperatures from 50 to 300 k. Phys Rev B 48:16373–16387. https://link.aps.org/doi/10.1103/PhysRevB.48.16373

Swartz ET, Pohl RO (1987) Thermal resistance at interfaces. Appl Phys Lett 51(26):2200–2202. https://doi.org/10.1063/1.98939

Swartz ET, Pohl RO (1989) Thermal boundary resistance. Rev Mod Phys 61:605–668. https://link.aps.org/doi/10.1103/RevModPhys.61.605

Volz S, Ordonez-Miranda J, Shchepetov A, Prunnila M, Ahopelto J, Pezeril T, Vaudel G, Gusev V, Ruello P, Weig EM, Schubert M, Hettich M, Grossman M, Dekorsy T, Alzina F, Graczykowski B, Chavez-Angel E, Sebastian Reparaz J, Wagner MR, Sotomayor-Torres CM, Xiong S, Neogi S, Donadio D (2016) Nanophononics: state of the art and perspectives. European Phys J B 89(1):15. https://doi.org/10.1140/epjb/e2015-60727-7

Energy Relaxation and Thermal Transport in Molecules

37

David M. Leitner

Contents

Abstract

The nature of energy relaxation in molecules and its role in mediating thermal conduction through a molecular junction are discussed. In addition to an overview of thermalization in molecules and quantum mechanical effects that influence this process, illustrative examples for alkane chain, perfluoroalkane, and polyethylene glycol (PEG) oligomer junctions are provided over a wide range of temperature. The calculated energy relaxation rates help clarify the scope of applicability of approaches that can be used to predict thermal conduction, e.g., where Fourier's law breaks down and where an approach that neglects energy relaxation and thermalization is suitable, such as one based on a Landauer-like model. Comparison with experimental data, where available, is provided and discussed.

D. M. Leitner (✉)
Department of Chemistry, University of Nevada, Reno, NV, USA
e-mail: dml@unr.edu

1   Introduction

Energy relaxation and ergodicity in molecules have long fascinated molecular scientists (Uzer 1991; Leitner 2015), in part because of their central role in chemical kinetics (Kuharski et al. 1988; Leitner and Wolynes 2006; Berrios et al. 2013) and molecular spectroscopy (Lehmann et al. 1994; Leitner and Wolynes 1996a; Leitner and Wolynes 1997a; Sibert III and Gruebele 2006) but also due to the many properties dynamics of molecules share with dynamics of other complex systems (Giannoni et al. 1991). In recent years, connections between energy relaxation and thermal transport in molecules have been the focus of much interest. Measurements and molecular simulations reveal that energy relaxation mediates thermal transport across molecular interfaces (Foley et al. 2014; Majumdar et al. 2015). Molecular interfaces have received considerable attention (Cahill et al. 2003, 2014; Segal et al. 2003; Galperin et al. 2007; Kuang and Gezelter 2011; Losego et al. 2012; O'Brien et al. 2012; Lee et al. 2013; Leitner 2013; Luo and Chen 2013; Stocker and Gezelter 2013; Gaskins et al. 2015; Majumdar et al. 2015; Craven and Nitzan 2016; Giri et al. 2016; Segal and Agarwalla 2016; Stocker et al. 2016), due in part to their many potential applications, ranging from nanoscale devices comprising composite materials, to photothermal therapies in the cell (Qin and Bischof 2012), to charge transfer with potential thermoelectric applications (Craven and Nitzan 2016, 2017; Chen et al. 2017).

In this perspective article, we discuss the role of energy relaxation on thermal transport across molecular interfaces, as in Fig. 1, where two substrates bridged by a monolayer junction are depicted. One object, for instance, a gold substrate or nanoparticle, is held at temperature, T_1, and is separated from the thermal bath at T_2, which could be another solid substrate, or a liquid or membrane, by a monolayer at the interface. If local temperature were well defined at the scale of atoms or a few atoms, we could sketch a temperature profile between T_1 and T_2, where for $T_1 > T_2$ we would observe a drop in temperature from the gold to the molecule, across the molecule, and again from the molecule to the outer layer. The boundary resistance between the gold and the outer layer is the sum of these resistances in series. Such a temperature profile is typically produced from the data provided by classical molecular dynamics (MD) simulations, since temperature can always be defined as proportional to the mean square atomic velocity in a classical system. However, local temperature in the molecules between the substrates may be less clear. For local temperature to be well defined, energy relaxation must be rapid, hence the need to clarify the nature and rate of energy relaxation in molecules. Local temperature can be established by inelastic processes, resulting from anharmonic interactions, which redistribute vibrational energy among the vibrational states (Allen and Feldman 1993; Cahill et al. 2003). We focus here on energy relaxation in molecules and its effect on thermal conduction.

In one limit energy relaxation is very rapid, which allows local temperature to be defined on essentially the scale of the atoms in the molecule, a condition for Fourier's heat law to apply. In the other limit, in the absence of thermalization, a
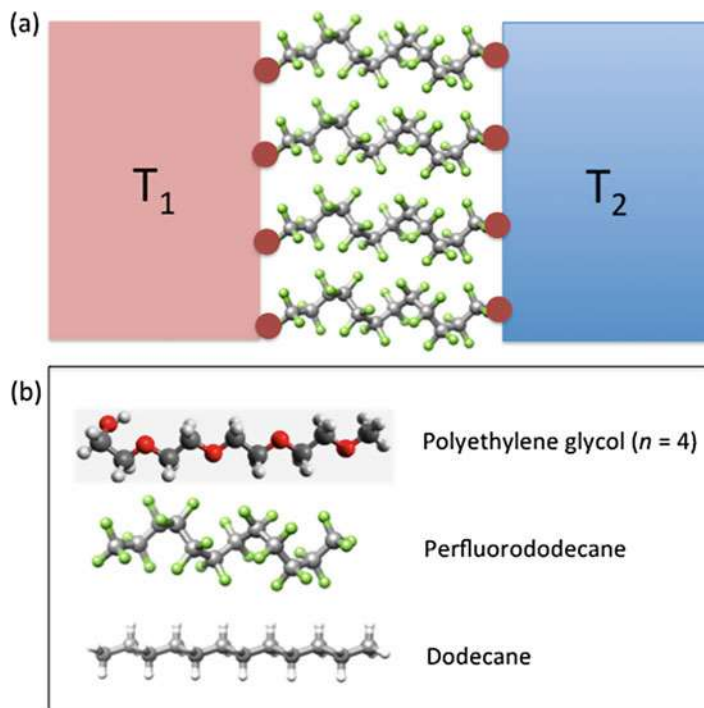
Fig. 1 (**a**) Two thermal baths at different temperature, T_1 and T_2, with perfluorododecane molecules bridging the baths. Red circles indicate bonding between molecules and substrates. (**b**) Structures representative of a PEG oligomer (top), perfluoroalkane (middle), and alkane chain (bottom). Red is oxygen, black is carbon, white is hydrogen, and green is fluorine

quantum mechanical approach based on a Landauer model, which neglects effects of inelastic scattering (Hopkins 2013; Li et al. 2015; Majumdar et al. 2015), may be appropriate for modeling thermal conduction. The Landauer formalism, introduced to quantify electrical conductance in mesoscopic systems (Datta 1995; Li et al. 2017), can be applied to thermal transport through molecules between two leads at different temperature and can accommodate elastic scattering of vibrational quasiparticles, i.e., the vibrational excitations of the molecule akin to phonons in a solid, which carry heat. By examining the extent of thermalization within molecules, we aim to clarify the scope of applicability of Landauer and fully thermalized pictures. The differences in thermal conduction from these two perspectives when one of the objects in Fig. 1 is gold can be quite striking, due to the low Debye temperature of gold relative to the vibrational frequencies of the surrounding layers.

Classical molecular dynamics (MD) simulations can be used to predict thermal conductance across molecular interfaces and the mechanisms that control it. However, in some cases energy relaxation, and its role in thermal conduction, may be exaggerated. This possibility was found in an experimental-computational study of

thermal transport through a molecular interface between two metal leads (Majumdar et al. 2015). The measured interface conduction between two leads, at least one gold, bridged by alkane chain monolayer junctions decreases with increasing mismatch of the vibrational density of states of the metal leads, whereas results of MD simulations indicate an increase over the same range of vibrational density mismatch. The authors suggest (Majumdar et al. 2015) scattering based on anharmonic interactions within the junction is likely exaggerated in the classical simulations, allowing vibrational energy entering from Au (Debye temperature 170 K) to up-convert and transfer to Pd (Debye temperature 275 K), channels apparently not available in the molecular interface probed experimentally (Majumdar et al. 2015).

Energy transport in molecules and effects of energy relaxation on transport are measured by a variety of time-resolved spectroscopies (Botan et al. 2007; Backus et al. 2008a, b, 2009; Rubtsov 2009; Fujii et al. 2014; Hassan et al. 2014; Rubtsova and Rubtsov 2015; Kondoh et al. 2016) and determined by molecular simulations and other computational approaches (Buldum et al. 1999; Leitner 2001a, b, 2002, 2012; Yu and Leitner 2003a, b, 2005; Enright et al. 2006; Enright and Leitner 2005; Ishikura and Yamato 2006; Leitner and Straub 2009; Xu et al. 2013; Agbo et al. 2014; Xu and Leitner 2014; Ishikura et al. 2015; Leitner et al. 2015; Leitner and Pandey 2015a, b; Buchenberg et al. 2016). Thermal conductance across molecular interfaces can be measured by thermo-reflectance (Hopkins 2013) and scanning thermal microscopy (Meier et al. 2014) techniques. For molecules attached to substrates, as in Fig. 1, bonding to the leads has been found to control thermal transport at the interface (Losego et al. 2012; O'Brien et al. 2012). In studies where the molecular length dependence has been examined, resistance within the molecule itself has also been observed to influence the boundary conductance (Gaskins et al. 2015).

Discussion of energy dynamics within a molecular junction is often framed within the limits of ballistic and diffusive, but it is important to clarify processes that give rise to non-ballistic dynamics and thermal resistance. While non-ballistic dynamics and thermal resistance can emerge from elastic and inelastic processes in the molecule, the response of the resistance to change in temperature depends on its origin. Inelastic scattering arises from the anharmonic coupling of the vibrational modes of the molecule, as well as from interactions between the molecule and its environment. Elastic scattering within the molecule arises from its aperiodicity, analogous to transport by diffusion modes of aperiodic materials (Fabian and Allen 1996). Eigenmodes that carry energy diffusively appear to share (Fabian and Allen 1996; Pandey and Leitner 2017a) properties predicted by random matrix theory (Feingold et al. 1989, 1991; Giannoni et al. 1991; Mehta 1991; Wilkinson et al. 1991; Leitner 1993; Leitner et al. 1994). The influence on the resistance of inelastic scattering, which gives rise to thermalization and local temperature and is necessary for Fourier's law, is quite different from effects of elastic scattering, which does not yield thermalization but nevertheless introduces resistance and diffusive vibrational quasiparticle dynamics. We consider the bonding or interaction between the molecule and the leads to which it is coupled to be sufficiently strong that thermal resistance in the junction arises from processes that occur within the

molecule itself, the origin of which may be elastic, inelastic, or both. A major aim of our work is clarifying the origin of thermal resistance.

In this perspective we discuss three classes of molecules: (i) alkane chains, (ii) perfluoroalkane chains, and (iii) polyethylene glycol (PEG) oligomers, which are depicted in Fig. 1. The PEG oligomers we examine here are $HO\text{-}(CH_2CH_2O)_n\text{-}CH_3$, $n = 2$ and 4, which are about 0.8 nm and 1.6 nm in length, respectively. We discuss for these systems thermal conduction and energy relaxation, which gives rise to thermalization in these molecules over a variety of lengths and temperature. At room temperature, and at a modest length of about 2 nm, we find no resistance or thermalization in alkane chains, some resistance but little thermalization in perfluoroalkanes, and significant thermalization and resistance in a PEG oligomer. These molecules thus exhibit significant diversity in thermalization and thermal resistance.

In Sect. 2 we discuss the nature and rate of energy relaxation in molecules, including the possibility of many-body localization (MBL) in the vibrational state space of a molecule, and effects of coupling between the molecule and thermal baths on MBL. In Sect. 3 we summarize two methods for estimating thermal conductance through a molecular junction, applicable in the limits of no energy relaxation and very rapid relaxation. In Sect. 4 we present calculations for alkane chains, perfluoroalkanes, and PEG oligomer junctions to illustrate effects of energy relaxation in molecules on thermal transport across a molecular junction. Section 5 summarizes this perspective.

2 Energy Relaxation in Molecules

Thermalization, the relaxation of vibrational energy in the molecule, occurs by inelastic scattering via anharmonic interactions. One approach to estimate the rate of thermalization quantum mechanically is to adopt first-order time-dependent perturbation theory, or Fermi's golden rule (GR). Because of restrictions on the applicability of the GR approach, however, the GR may break down when applied to the problem of energy relaxation in molecules. In particular, if the product of the local density of vibrational states and the matrix elements that couple them is smaller than order 1, we cannot use the GR to reliably estimate the energy relaxation rate in a molecule. In that case a more general approach is required, and we present such an approach below.

We address first energy relaxation in a molecule that is decoupled from its environment. We write the vibrational Hamiltonian of the molecule as $H = H_0 + V$, where the zero-order Hamiltonian, H_0, is expressed as a sum of the energy in each of the N vibrational modes, i.e., $H_0 = \sum_{\alpha=1}^{N} \varepsilon_\alpha \left(\hat{n}_\alpha\right)$, where α labels a mode, the number operator is defined by $\hat{n}_\alpha = b_\alpha^+ b_\alpha$, b_α^+ and b_α are raising and lowering operators, and ε_α is the energy in mode α. The vibrational modes of the molecule are coupled by anharmonic interactions, V. We take V

to include coupling to third order in the anharmonicity; terms of higher order can be included but typically decrease exponentially in magnitude (Gruebele and Bigwood 1998), and we thus neglect them. The zero-order states are product states labeled by the number of vibrational quanta in each mode. If the molecule, H, is prepared initially in one of the eigenstates of H_0, then energy may be redistributed via the anharmonic interactions, rearranging mode populations until the molecule has thermalized. However, for the isolated molecule, H, regardless of size, energy relaxation may not occur (Logan and Wolynes 1990; Leitner and Wolynes 1996a, b, c; Leitner 2015), even if we were to account for all anharmonic interactions, if the product of the magnitude of the interaction and the density of states coupled by the anharmonic terms is less than order 1. This is an example of a many-body localized (MBL) system (Lev and Reichman 2014; Burin 2015, 2017; Nandkishore and Huse 2015), where an isolated, many-body system cannot thermalize under its own dynamics. MBL has been observed in numerous vibrationally excited polyatomic molecules in the gas phase (Bigwood et al. 1998; Leitner 2015). Effects of MBL in molecules influence rates of chemical reactions (Leitner and Wolynes 1997; Leitner 1999, 2005, 2015; Leitner et al. 2003; Leitner and Gruebele 2008; Keshavamurthy 2013a, b).

As noted, localization occurs if the product of the average anharmonic matrix element and the local density of states coupled by anharmonicity is less than order 1. Detailed analysis leads to the criterion (Leitner 2015):

$$T(E) \equiv \frac{2\pi}{3} \langle |V_{\alpha\beta\gamma}| \rangle^2 \rho^2(E) < 1, \tag{1}$$

where ρ is the local density of states and $\langle |V_{\alpha\beta\gamma}| \rangle$ is the average size of the matrix element coupling the triple of modes α, β, and γ. This criterion emerges from a self-consistent analysis of the distribution of rates at which states on the energy shell relax toward equilibrium. The most probable rate in a large molecule is always 0 when $T < 1$, despite finite coupling between vibrational states due to anharmonic interactions.

It is convenient to separate the interactions between vibrational states into two relaxation processes that can occur while conserving energy. In one process, decay, a quasiparticle in a mode of frequency ω_α, i.e., a vibrational excitation in that mode, decays into two quasiparticles of frequency ω_β and ω_γ. In the other process, collision, a quasiparticle in a mode of frequency ω_α, combines with another of frequency ω_β to create a quasiparticle of frequency ω_γ. The matrix elements are of the form $|V_{\alpha\beta\gamma}| = |\Phi_{\alpha\beta\gamma}| \sqrt{n_\alpha (n_\beta + 1)(n_\gamma + 1)}$ for interactions in a decay process and $|V_{\alpha\beta\gamma}| = |\Phi_{\alpha\beta\gamma}| \sqrt{n_\alpha n_\beta (n_\gamma + 1)}$ for interactions in a collision. $\Phi_{\alpha\beta\gamma}$ are the coefficients of the cubic terms in the expansion of the interatomic potential in normal coordinates. To determine whether an isolated molecule can thermalize under its own dynamics, we need to calculate $T(E)$ by averaging over all states on the energy shell. For a large molecule, most states are occupied very

nearly according to $\langle n_a \rangle = \left(e^{\hbar\omega/k_B T} - 1 \right)^{-1}$, so for n_α we use $\langle n_a \rangle$. Averaging $\left| \Phi_{\alpha\beta\gamma} \right| \sqrt{n_\alpha \left(n_\beta + 1 \right) \left(n_\gamma + 1 \right)}$ and $\left| \Phi_{\alpha\beta\gamma} \right| \sqrt{n_\alpha n_\beta \left(n_\gamma + 1 \right)}$ for resonant coupling, we define $T = T_d + T_c$, where:

$$T_d \left(\omega_\alpha \right) = \frac{2\pi}{3} \langle \left| \Phi_{\alpha\beta\gamma} \right| n_\alpha \left(n_\beta + 1 \right) \left(n_\gamma + 1 \right) \rangle^2 \rho_{res}^2 \left(\omega_\alpha \right) \tag{2a}$$

$$T_c \left(\omega_\alpha \right) = \frac{2\pi}{3} \langle \left| \Phi_{\alpha\beta\gamma} \right| n_\alpha n_\beta \left(n_\gamma + 1 \right) \rangle^2 \rho_{res}^2 \left(\omega_\alpha \right). \tag{2b}$$

Finally, we average over the modes at a given temperature (internal energy):

$$T \left(\mathrm{T} \right) = Q^{-1} \sum_\alpha T \left(\omega_\alpha \right) e^{-\hbar\omega/k_B \mathrm{T}}, \tag{3}$$

where Q is the partition function.

We calculate the rate of thermalization as the thermal average of the most probable energy relaxation rate for each of the modes of the molecule. To determine the most probable rate, we begin with the average relaxation rate for excess energy in a mode, which can be estimated with the GR to third order in anharmonicity. Relaxation then has decay and collision contributions with rates W_d and W_c, respectively. For a vibrational mode α with frequency ω_α and excess vibrational excitation, the average rate can be written as $W_{ave} = W_d + W_c$, where (Maradudin and Fein 1962):

$$W_d \left(\omega_\alpha \right) = \frac{\hbar}{16\omega_\alpha} \sum_{\beta,\gamma} \frac{\left| \Phi_{\alpha\beta\gamma} \right|^2 \left(1 + n_\beta + n_\gamma \right)}{\omega_\beta \omega_\gamma} \frac{\left(\Gamma_\alpha + \Gamma_\beta + \Gamma_\gamma \right)}{\left(\omega_\alpha - \omega_\beta - \omega_\gamma \right)^2 + \left(\Gamma_\alpha + \Gamma_\beta + \Gamma_\gamma \right)^2 / 4}, \tag{4a}$$

$$W_d \left(\omega_\alpha \right) = \frac{\hbar}{8\omega_\alpha} \sum_{\beta,\gamma} \frac{\left| \Phi_{\alpha\beta\gamma} \right|^2 \left(n_\beta - n_\gamma \right)}{\omega_\beta \omega_\gamma} \frac{\left(\Gamma_\alpha + \Gamma_\beta + \Gamma_\gamma \right)}{\left(\omega_\alpha + \omega_\beta - \omega_\gamma \right)^2 + \left(\Gamma_\alpha + \Gamma_\beta + \Gamma_\gamma \right)^2 / 4}. \tag{4b}$$

Γ_β is the rate at which a vibrational excitation in mode β relaxes due to coupling to other modes and to the environment, n_α is the population, and $\Phi_{\alpha\beta\gamma}$ the coefficients of the cubic terms in the expansion of the interatomic potential in normal coordinates. For most vibrational modes of large molecules, Γ_β is of the order 1 ps^{-1} (Leitner 2008). For a large molecule often $(\omega_\alpha - \omega_\beta - \omega_\gamma)^2 < (\Gamma_\alpha + \Gamma_\beta + \Gamma_\gamma)^2/4$, i.e., the spacing between resonances is typically much smaller than the line width, and in practice the GR rate does not depend much on the value of Γ_β. We have used 1 ps^{-1} for each Γ_β in calculations, as in previous calculations for large

molecules (Stuchebrukhov 1986; Stuchebrukhov and Marcus 1993; Leitner 2012), and checked that the GR value remains essentially unchanged upon varying this value. However, for some molecules this condition may not be met, as we now consider.

If T is significantly larger than 1, the average relaxation rate calculated with the GR goes over to the most probable relaxation rate (Leitner 2015). However, near the MBL transition, the most probable rate is smaller than the average rate, meeting it only at large T. Correcting for the effects of a finite local density of states in the molecule, the most probable relaxation rate becomes (Leitner 2015):

$$W_d^{mp}\left(\omega_\alpha\right) = W_d\left(\omega_\alpha\right)\left(1 - T^{-1}\left(\omega_\alpha\right)\right)^{1/2}, T \geq 1 \tag{5a}$$

$$W_c^{mp}\left(\omega_\alpha\right) = W_c\left(\omega_\alpha\right)\left(1 - T^{-1}\left(\omega_\alpha\right)\right)^{1/2}, T \geq 1 \tag{5b}$$

We take a thermal average of the rates:

$$W\left(\mathrm{T}\right) = Q^{-1}\sum_\alpha \left(W_c^{mp}\left(\omega_\alpha\right) + W_d^{mp}\left(\omega_\alpha\right)\right)e^{-\hbar\omega/k_B\mathrm{T}}. \tag{6}$$

We use the transition rates given by Eqs. (5) and (6) to estimate the relaxation rate in molecules that form a thermal bridge between two substrates when $T > 1$.

A molecule at the interface between two substrates, as depicted in Fig. 1, is of course not isolated but is coupled to thermal baths. The molecular eigenstates are thus not localized, but the rate of thermalization is nevertheless reduced if the isolated molecule exhibits MBL. Even when the molecule is not isolated, the most probable rate at which excess vibrational energy in a mode relaxes toward equilibrium remains in practice slow if $T < 1$. We account for energy transfer to and from the thermal baths, assuming vibrational quasiparticles enter and leave the molecule in steady state.

MBL is lost as a result of dephasing (Leitner 2015) due to coupling between the molecule at the junction and its environment. We calculate the dephasing rate, η, using a ballistic transport time as an upper estimate; if there is elastic scattering, the time for a quasiparticle that has entered the molecule to leave it will be longer and η smaller. Using the larger, ballistic estimate has very little practical effect on the rate of thermalization, which is generally small if $T < 1$ (Pandey and Leitner 2016). The ballistic transport time is estimated as the ratio of the speed of sound in the molecule at the junction and the length of the molecule. For the three classes of molecules we consider here, alkanes, perfluoroalkanes, and PEG oligomers, the speed of sound has been calculated to be 1768 ms^{-1}, 641 ms^{-1} and 833 ms^{-1}, respectively (Pandey and Leitner 2016, 2017b).

Calculation of the effects of dephasing on thermalization within molecules is carried out as follows: For convenience we define dimensionless variables, $\overline{W}_{mp}(E)$ and $\overline{\eta}$, as

$$\overline{W}_{mp}(E) = W_{mp}(E)/W_{ave}(E) \tag{7a}$$

$$\overline{\eta}(E) = \eta(E)/W_{ave}(E) \tag{7b}$$

Solving for $\overline{\eta}$ we have (Leitner 2015):

$$\overline{\eta}(E) = \frac{\left(1 \pm \left[1 - 4T^{-1}(E)\overline{W}_{mp}^2(E)\right]^{1/2}\right)}{2\overline{W}_{mp}(E)} - \overline{W}_{mp}(E) \tag{8}$$

Since both $\Delta'_{mp}(E)$ and $\overline{\eta}$ are real and nonnegative, when $0 \le T \le 2$ and if $0 \le \overline{\eta} \le \eta_{max} = [2-T]/2T^{1/2}$, we choose the negative sign, while for $\overline{\eta} \ge \eta_{max}$ we choose the positive sign. For $T \ge 2$ only the positive sign is correct.

3   Thermal Conductance in Limits of Slow and Rapid Energy Relaxation

We illustrate the impact of energy relaxation in molecules on thermal conduction through a molecular interface by comparing two limits, one in which energy relaxation is severely restricted or so slow as to have no effect on thermal conduction and the other where it is so rapid that local temperature is well defined within the molecule. We consider the former case first and then turn to the latter.

3.1   No Thermalization: Landauer Model

Consider the thermal boundary conductance, h_{Bd}, between two materials, the inverse of the thermal boundary resistance or Kapitza resistance. (Swartz and Pohl 1989) is expressed in terms of the heat flow, $\dot{Q}$, driven by the difference in temperature, ΔT, between the two sides of the interface, and the area of the interface, A:

$$h_{Bd} = \frac{\dot{Q}}{A\Delta T} \tag{9}$$

Let the temperature on one side of the molecule be T_1 and the other T_2. The boundary conductance can be expressed (Swartz and Pohl 1989) in terms of the gross rate of heat flow from one side, say the left, L, to the other, R:

$$h_{Bd} = \frac{\dot{Q}_{L \to R}^{gross}(T_2) - \dot{Q}_{L \to R}^{gross}(T_1)}{A(T_2 - T_1)} \tag{10}$$

Thermal transport through the molecules at the interface can be expressed in terms of the quasiparticles, e.g., phonons in a solid, with energy $\hbar\omega$. The vibrational mode density per unit volume on side j is $\overline{\rho}_j(\omega)$; the mode occupation number is $n(\omega, T) = (\exp(\beta\hbar\omega) - 1)^{-1}$, where $\beta = 1/k_B T$; and the phonon speed on side j is $v_j(\omega)$. In the limit $T_2 = T_1 + dT$:

$$h_{Bd} = \frac{1}{A}\frac{d}{dT}\dot{Q}_{L \to R}^{gross} = \frac{1}{4}\frac{d}{dT}\int d\omega \, v_L \overline{\rho}_L(\omega) \; \hbar\omega \, n(\omega) \; \tau(\omega), \tag{11}$$

where τ is the transmission coefficient. In Eq. (11), the factor ¼ accounts for all incident angles of vibrational quasiparticles striking the interface (Swartz and Pohl 1989). Equation (11) is the thermal boundary conductance predicted by the diffuse mismatch model (DMM) (Swartz and Pohl 1989).

One contribution to the transmission coefficient, τ_0, accounts for detailed balance between the left and right sides of the interface:

$$\tau_0(\omega) = \frac{v_R \overline{\rho}_R(\omega)}{v_L \overline{\rho}_L(\omega) + v_R \overline{\rho}_R(\omega)} \tag{12}$$

Another contribution accounts for scattering within the molecules. Let l be the mean free path and L the length of the molecule. If upon scattering the vibrational excitation reverses direction 50% of the time and otherwise continues in the same direction, and if we neglect interference effects, the transmission coefficient, τ, is given by:

$$\tau(\omega) = \frac{\tau_0(\omega)}{1 + (L/l)\,\tau_0(\omega)}. \tag{13}$$

In the absence of inelastic processes, the integral in Eq. (11) is taken from 0 to the lower Debye frequency, ω_D, of the materials at each side of the interface. In general the mean free path, l, depends on temperature, but if we neglect that dependence, then only n in Eq. (11) is temperature-dependent and $\hbar\omega\frac{dn}{dT}$ is the heat capacity, $C(\omega)$. We write the thermal boundary conductance as

$$h_{Bd} = B \int d\omega \, v_L \overline{\rho}_L(\omega) \; C(\omega) \; \tau(\omega). \tag{14}$$

In Eq. (14) an additional factor, B, appears, which includes the factor of $1/4$ in Eq. (11) and allows, e.g., for incomplete surface contact, though for the examples illustrated here, we simply fit B to experimental data. It is convenient to adopt a Debye model for the phonon densities in Eqs. (11), (12), and (14). In that case we use for the mode density $\overline{\rho}(\omega) = \frac{3\omega^2}{2\pi^2 v^3}$, where we take one representative value for the speed of sound, v, for simplicity.

While we started with the DMM, we refer to Eq. (14) as a Landauer model to emphasize the absence of inelastic scattering through the junction and to permit elastic scattering, with scattering length, l, within each molecule, analogous to the treatment of the mean free path of an electron in a conductor (Datta 1995). The more standard form for the Landauer expression for thermal conductance is $\frac{1}{4\pi} \int d\omega \; \hbar\omega \frac{\partial n(\omega,T)}{\partial T} \overline{\tau}(\omega)$ (Dhar 2008; Wang et al. 2008; Duchemin and Donadio 2011), where the transmission coefficient is $\overline{\tau}(\omega)$, and can be related to Eq. (14) by $\overline{\tau}(\omega) = 4\pi B v_L \overline{\rho}_L(\omega) \; C(\omega) \; \tau(\omega)$. Thermal conduction using a Landauer expression may be calculated more directly, without separately estimating the mean free path in harmonic approximation, using methods based on the nonequilibrium Green's function formalism described by Dhar (2008), Wang et al. (2008), Duchemin and Donadio (2011), and Buerkle et al. (2015).

3.2   Rapid Thermalization: Series Model

If the Landauer approach neglects energy relaxation altogether, the other extreme is energy relaxation that is so rapid that temperature is well defined on the scale of an atom or a few atoms. If local temperature is well defined, we can calculate the resistance across the junction using a series model:

$$1/h_{Bd} = 1/h_{Bd}^{L-mol} + 1/h_{Bd}^{R-mol} + 1/h_{mol}, \tag{15}$$

where $1/h_{mol}$ is the thermal resistance within the molecule and $1/h_{Bd}^{L-mol}$ and $1/h_{Bd}^{R-mol}$ are, respectively, thermal resistance between the left substrate and the molecule and right substrate and molecule.

For the purpose of comparing results of a series model with those of a Landauer-like model, we simply estimate the conductance of the molecule in Eq. (15) in the limit where local temperature is well defined as

$$h_{mol} = \frac{1}{VL} D \sum_{\alpha} C_{\alpha}, \tag{16}$$

where α labels a vibrational mode of the molecule; the sum is taken over all modes; D is the coefficient for energy diffusion, taken as independent of mode frequency for simplicity; and V is the volume of the molecule. While the volume is somewhat arbitrary (Wu et al. 2017), we estimate it in a way that matches the Landauer formalism in the limit of low temperature, as it must.

The resistance through the molecule, $1/h_{mol}$, dominates the resistance in the junction when the junction is long. In general, $1/h_{Bd}^{L-mol}$ and $1/h_{Bd}^{R-mol}$ need to be evaluated, which we have done using Eq. (14) for convenience, though of course a more precise approach would account for the discrete modes of the molecule. The contributions of $1/h_{Bd}^{L-mol}$ and $1/h_{Bd}^{R-mol}$ to the thermal conductance are relatively small in the illustrative calculations we present in the following section.

4   Thermalization Rates in Molecules: Alkanes, Perfluoroalkanes, and PEG Oligomers

We consider first whether or not the molecules exhibit MBL when they are separated from the substrates. The vibrational states of a molecular junction cannot be localized because of contact with heat baths. However, if the states of the same molecule in isolation are localized, the rate of thermalization is predicted by Eq. (8), rather than the GR, and is in practice very small. We thus begin by evaluating the transition parameter, T, defined by Eqs. (2) and (3).

Values of T for hexane, dodecane, perfluorohexane, perfluorododecane, PEG_2, and PEG_4 are plotted in Fig. 2a at internal energies corresponding to temperatures from 100 to 400 K. The indicated temperature is the value used in the thermal averaging in Eq. (3). At 100 K all lie below or near $T = 1$, so that most of these molecules exhibit effects of MBL at low temperature. At 200 K the alkanes still exhibit MBL, but the longer perfluoroalkane and PEG_4 do not. By 300 K we find $T < 1$ only for hexane. Overall, the thresholds differ among these systems because of differences in the local density of states of the molecules, which becomes somewhat larger as the molecule becomes larger for these cases or becomes larger with addition of fluorine.

Using the values of T plotted in Fig. 2a, we have calculated the most probable relaxation rates with Eqs. (4), (5), and (6), where $T > 1$, and with Eqs. (7) and (8) where $T < 1$. The thermalization time may of course exceed the relaxation time for excess energy in a mode, but we expect the rates to be similar if T_1 and T_2, the temperatures of the substrates, are close, i.e., the linear response regime. We observe overall that the thermalization rates in the smaller molecules of a particular class appear to be slower than the larger molecules of that class. The variation of the thermalization rate with temperature for the smaller molecules is highly nonlinear, the result of effects of small T for those molecules, particularly at low temperature, yielding very small rates at those temperatures. At 300 K the thermalization rates are slowest for alkane chains, then for PEG oligomers, and fastest for the perfluoroalkanes. For example, at 300 K the thermalization rate for the larger alkane chain, $C_{12}H_{26}$, was calculated to be 0.7 ps^{-1}, slower than the
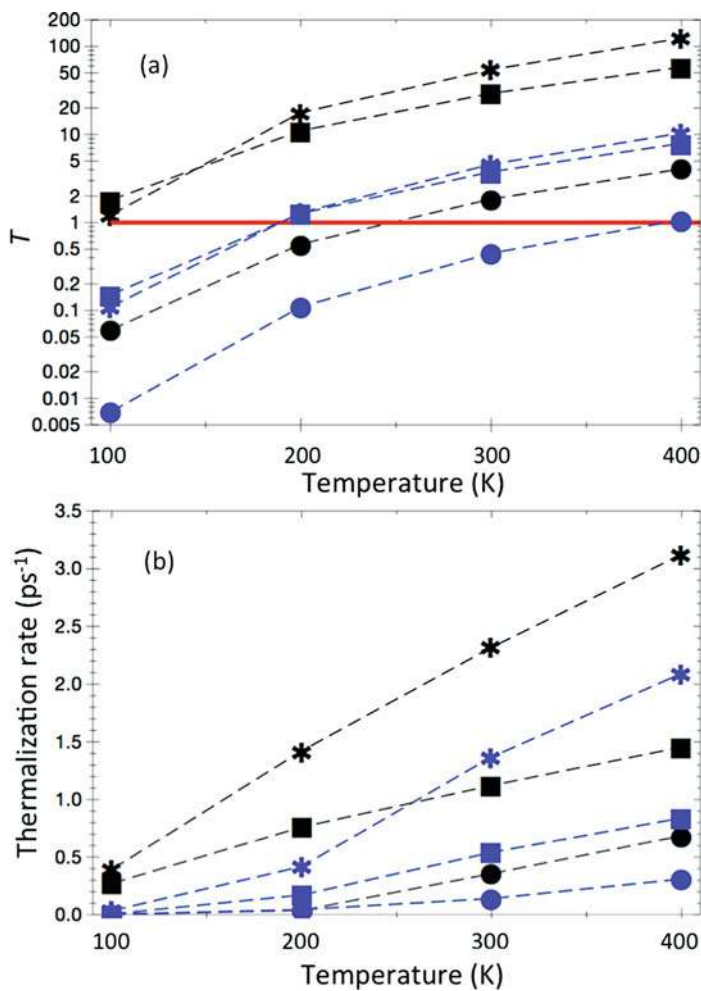
Fig. 2 (**a**) Transition parameter, T, plotted against temperature for alkanes (circles), perfluoroalkanes (squares), and PEG oligomers (*). Hexane, perfluorohexane, and PEG$_2$ are blue, and dodecane, perfluorododecane, and PEG$_4$ are black. The MBL threshold, $T = 1$, is shown as the red line. (**b**) Thermalization rate is plotted against temperature for the same systems

rate, 1.5 ps^{-1}, found for the larger PEG oligomer, and much slower than the rate of 3.1 ps^{-1} for perfluorododecane.

4.1   Thermalization Length

We refer to the thermalization length as the length in the molecule over which thermalization occurs. This length depends on the speed at which vibrational energy

propagates. We estimate the thermalization length as the ratio of the speed of sound along the molecule and the thermally averaged relaxation rate. For the molecules we consider here, alkanes, perfluoroalkanes, and PEG oligomers, we have noted the speed of sound obtained in earlier computational work (Pandey and Leitner 2017a) as 1768 ms^{-1}, 641 ms^{-1}, and 833 ms^{-1}, respectively. Using these speeds and the values of the thermalization rates plotted in Fig. 2, we have calculated the length over which thermalization occurs for PEG$_4$, perfluorododecane, and dodecane, which we plot in Fig. 3.

We also plot the result for pentadecane, where the effects of MBL on the most probable rate of thermalization are not included, i.e., we simply use the GR estimate for the energy relaxation rate. This is referred to in Fig. 3 as a long alkane, for which we expect effects of MBL on thermalization rates to be negligible. It is instructive to compare this result with the result for dodecane, since in the latter we find a substantial difference between the thermalization rate predicted by the GR and the actual thermalization rate, particularly at lower temperature. This gives rise to a sizable difference between the length over which thermalization occurs in dodecane at, say, 300 K, which is about 10 nm, and a long alkane, which is 5 nm. For PEG oligomers we have found that thermalization in PEG$_4$ is representative of that in a longer oligomer, and we find for this molecule and for longer perfluoroalkanes a thermalization length that is below 1 nm at 300 K.

Perfluoroalkanes and PEG oligomers also exhibit elastic scattering. We calculated the energy diffusion coefficient in harmonic approximation by propagating wave packets in terms of the normal modes for these molecules and found for perfluoroalkanes an energy diffusion coefficient of $D = 17.7$ Å^2 ps^{-1}, whereas
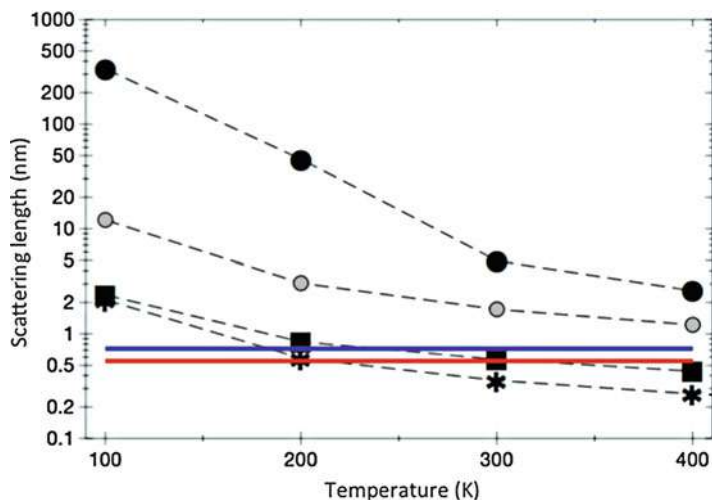
Fig. 3 Scattering length in PEG$_4$ (*), perfluorohexane (squares), a long alkane (gray circles), and dodecane (black circles) is plotted from 100 to 400 K. The mean free path for elastic scattering in PEG oligomers (blue line) and perfluoroalkanes (red line) is also plotted

for the PEG oligomers, we found $D = 30.0$ Å^2 ps^{-1} (Pandey and Leitner 2017b). The mean free path in these quasi-one-dimensional systems is given by $2D/v$. For perfluoroalkanes and PEG oligomers, the speed of sound is $v = 6.41$ Å ps^{-1} and $v = 8.33$ Å ps^{-1}, respectively, yielding, respectively, a mean free path for elastic scattering of 5.5 Å and 7.2 Å. We note that no elastic scattering was found for alkane chains up to at least C$_{15}$H$_{32}$.

4.2   Thermal Conduction

The results of the energy relaxation calculations help us assess which picture is appropriate for describing thermal conduction, one in which a vibrational quasiparticle travels across the molecule without thermalizing or one in which thermalization occurs over the length of a few atoms within the molecule. This picture, in turn, points us to the appropriate method to calculate thermal conductance across a molecular interface. For molecular junctions of about 2 nm, alkanes and fluorinated alkanes are not sufficiently thermalized, at least to about 400 K, to be able to apply a series model such as Eq. (15). In this case, a Landauer-like approach, which neglects thermalization, should provide a good estimate to the thermal conductance. We shall take this approach for those molecules and compare with experimental measurements of thermal conductance through similar molecules between gold and sapphire carried out by Hopkins and coworkers (Gaskins et al. 2015). However, we have seen that energy relaxation in PEG oligomers of comparable length ($n = 4$ or longer) occurs more rapidly. We therefore expect a Landauer approach, which does not capture this, to be of more questionable utility in predicting the thermal conductance. We thus compare the result of a Landauer model prediction and a resistance in series prediction for the thermal conductance between gold and sapphire.

To compare the predictions for alkanes and perfluoroalkanes with the measurements reported by Gaskins et al. (2015), we use Eq. (14) with a $L = 2$ nm perfluorinated alkane, the junction length in the experiments (Gaskins et al. 2015), and we use the mean free path $l = 0.55$ nm found in the elastic scattering calculation discussed above. A value of $B = 0.21$ was used, as this fits the measured data over a wide range of temperature. In Fig. 4 we plot the thermal conductance from Au to sapphire with a molecular layer of alkanes and a molecular layer of perfluoroalkanes between them, at 100 K, 200 K, 300 K, and 400 K. We observe in Fig. 4 good agreement with experiment. The calculated elastic mean free path introduced into the Landauer model appears to describe well the effect of fluorination on thermal transport in the molecular layer.

Turning to the Au-PEG$_4$-sapphire system, we again begin with the Landauer model in Eq. (14). We use for the speed of sound in sapphire 7229 m s^{-1} (Gaskins et al. 2015). The length of PEG$_4$ is about $L = 1.6$ nm and the mean free path plotted above is $l = 0.77$ nm. We plot the results in Fig. 4. We compare the results using the Landauer model with results calculated using the resistors-in-series model,
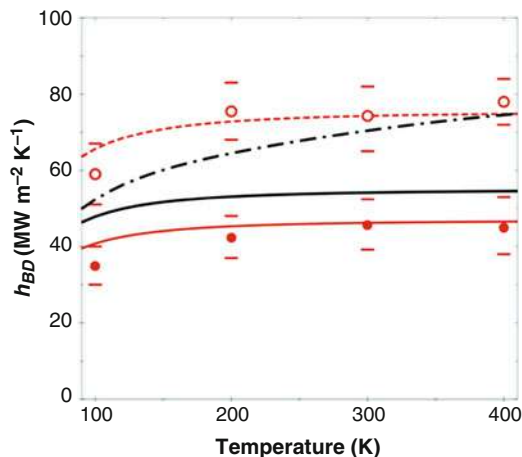
Fig. 4 Calculations of thermal conductance between Au and sapphire bridged by alkanes and perfluoroalkanes (red) and PEG oligomers (black) for temperatures from 100 to 400 K. A Landauer model is used with calculated elastic scattering length for the alkanes (dashed) and perfluoroalkanes (solid) and for PEG$_4$ (solid). A series model is also used for PEG$_4$ (dot-dashed) for comparison. Experimental results from Gaskins et al. (2015) are plotted (Au-sapphire red circles, closed and open, respectively, for carbon-chain interfaces that are fluorinated and not fluorinated) with reported error bars

where for the latter we introduce a cross-sectional area of 0.13 nm^2 so that the two calculations meet at low temperature.

We observe a striking difference in the temperature dependence of the thermal conductance produced by the Landauer and resistors-in-series pictures. Thermalization increases thermal conduction as additional transport channels are opened up by anharmonic interactions within the molecules at the interface, particularly at higher temperature. Because of those channels, phonons from gold that enter the molecule can up-convert and the resulting vibrational energy transferred to sapphire, a pathway that is unavailable in the absence of thermalization.

Enhancement of thermal conduction from one material to another via a molecular layer has been observed in other molecular simulations of heat transport. Thermal transport from gold to an organic solvent via an organic capping agent was found to be enhanced compared to thermal conduction across the gold-organic solvent interface in the results of MD simulations (Kuang and Gezelter 2011). The capping agent provides a bridge between gold vibrations and the surrounding liquid (Kuang and Gezelter 2011), which can result when vibrational energy is thermalized in the capping agent, facilitating thermal transport.

Some studies have compared results of calculations of thermal conductance in molecular junctions using a Landauer picture with the results of MD simulations as a check on the assumption that, over the length of the junction, thermalization effects on thermal conduction in the molecule are likely small. For example, a recent computational study using both a nonequilibrium Green's function approach (Buerkle et al. 2015) and nonequilibrium MD simulations of thermal transport in a

polyethylene junction (Buerkle and Asai 2017) found good agreement between the two calculations over a wide range of temperature, including room temperature, suggesting thermalization is likely negligible in this case. MD simulations have in other cases been found to exaggerate the role of anharmonicity, as discussed above. We thus prefer to calculate the rate of thermalization quantum mechanically as a check on the validity of a Landauer-like picture, where diffusive transport and thermal resistance may still arise by elastic scattering, but where the response to temperature change is quite different than if anharmonic effects were important.

5   Conclusions

In this perspective we have discussed the role energy relaxation plays in mediating thermal transport through molecular junctions. We have discussed energy relaxation in molecules, which arises from anharmonic interactions among the vibrational modes of the molecule, as well as from interactions between the molecule and its surroundings. Rapid energy relaxation is essential for local temperature to be established in a molecule, a condition that we have seen is often not met on the molecular scale. For that reason, approaches that neglect inelastic scattering within the molecule often predict thermal conductance in molecular junctions quite well. However, for many molecules, particularly of increasing length, complexity, and at room temperature or higher, inelastic scattering via anharmonic interactions is sufficiently rapid that it must be taken into account. We have summarized a general formalism for calculating the rate of energy relaxation via anharmonic interactions in a molecule, which includes the golden rule as a special case. In some molecules, even rather large ones, quantum mechanical effects can reduce the rate of relaxation below that predicted by the golden rule. We considered three classes of molecules, alkane and perfluoroalkane chains and PEG oligomers. Alkane chains, in particular, exhibit quantum mechanical effects that severely limit the rate of energy relaxation.

PEG oligomers exhibit the most rapid thermalization among these groups. While alkane chains thermalize over 10s or 100 s of nm, depending on temperature, perfluoroalkanes thermalize over shorter lengths on the order of 1 nm, and for PEG oligomers temperature is well defined over a few atoms at most temperatures of interest. Because thermalization is relatively rapid in PEG oligomers, molecular modeling methods that use temperature profiles to predict thermal boundary conductance for these systems should work well. That stands in contrast to alkane chains, where artifacts from the lack of thermalization in these systems have been found when using classical MD simulations to predict thermal conductance (Majumdar et al. 2015).

When thermalization in a junction is absent or occurs slowly, a Landauer model can be adopted to predict thermal transport through the junction. If there are vibrational modes of the molecule that are higher in frequency than those of the leads, they are not available as additional channels for vibrational energy transport. In this case, the increasing vibrational mismatch between two leads yields a smaller thermal conduction that cannot be remedied by a molecular interface between them, as has been observed for alkane chain junctions (Majumdar et al. 2015).

If thermalization occurs we expect the mismatch effect to be diminished and the temperature dependence of the conductance to be more pronounced. Indeed, we have illustrated how thermalization within a molecular junction can facilitate thermal transport by opening channels for vibrational energy to transfer to the leads.

References

Agbo JK, Xu Y et al (2014) Vibrational energy flow across heme-cytochrome *c* and cytochrome *c*-water interfaces. Theor Chem Accounts 133:1504

Allen PB, Feldman JL (1993) Thermal conductivity of disordered harmonic solids. Phys Rev B 48:12581–12588

Backus EHG, Nguyen PH et al (2008a) Structural flexibility of a helical peptide regulates vibrational energy transport properties. J Phys Chem B 112:15487–15492

Backus EHG, Nguyen PH et al (2008b) Energy transport in peptide helices: a comparison between high- and low-energy excitations. J Phys Chem B 112:9091–9099

Backus EH, Bloem R et al (2009) Dynamical transition in a small helical peptide and its implication for vibrational energy transport. J Phys Chem B 113:13405–13409

Berrios E, Pratt S et al (2013) More protected vibrational states at the dissociation limit of SCCl2. J Phys Chem A 117:12082–12090

Bigwood R, Gruebele M et al (1998) The vibrational energy flow transition in organic molecules: theory meets experiment. Proc Natl Acad Sci U S A 95:5960–5967

Botan V, Backus EHG et al (2007) Energy transport in peptide helices. Proc Natl Acad Sci U S A 104:12749–12754

Buchenberg S, Leitner DM et al (2016) Scaling rules for vibrational energy transport in proteins. J Phys Chem Lett 7:25–30

Buerkle M, Asai Y (2017) Thermal conductance of Teflon and polyethylene: insight from an atomistic, single-molecule level. Sci Rep 7:41898

Buerkle M, Hellmuth TJ et al (2015) First-principles calculation of the thermoelectric figure of merit for [2,2]paracyclophane-based single-molecule junctions. Phys Rev B 91:165419

Buldum A, Leitner DM et al (1999) Thermal conduction through a molecule. Europhys Lett 47:208–212

Burin AL (2015) Many-body localization in a strongly disordered system with long-range interactions: finite-size scaling. Phys Rev B 91:094202

Burin A (2017) Localization and chaos in a quantum spin glass model in random longitudinal fields: mapping to the localization problem in a Bethe lattice with a correlated disorder. Ann Phys 529:1600292

Cahill DG, Ford WK et al (2003) Nanoscale thermal transport. J Appl Phys 93:793–818

Cahill DG, Braun PV et al (2014) Nanoscale thermal transport. II. 2003–2012. Appl Phys Rev 1:011305

Chen R, Craven GT et al (2017) Electron-transfer-induced and phononic heat transport in molecular environments. J Chem Phys 147:124101

Craven GT, Nitzan A (2016) Electron transfer across a thermal gradient. Proc Natl Acad Sci U S A 113:9421–9429

Craven GT, Nitzan A (2017) Electron transfer at thermally heterogeneous molecule-metal interfaces. J Chem Phys 146:092305

Datta S (1995) Electronic transport in mesoscopic systems. Cambridge University Press, Cambridge, UK

Dhar A (2008) Heat transport in low-dimensional systems. Adv Phys 57:457–537

Duchemin I, Donadio D (2011) Atomistic calculation of the thermal conductance of large scale bulk-nanowire junctions. Phys Rev B 84:115423

Enright MB, Leitner DM (2005) Mass fractal dimension and the compactness of proteins. Phys Rev E 71:011912

Enright MB, Yu X et al (2006) Hydration dependence of the mass fractal dimension and anomalous diffusion of vibrational energy in proteins. Phys Rev E 73:051905

Fabian J, Allen PB (1996) Anharmonic decay of vibrational states in amorphous silicon. Phys Rev Lett 77:3839–3842

Feingold M, Leitner DM et al (1989) Semiclassical structure of Hamiltonians. Phys Rev A 39:6507–6514

Feingold M, Leitner DM et al (1991) Spectral statistics in semiclassical random matrix ensembles. Phys Rev Lett 66:986–989

Foley BM, Gorham CS et al (2014) Protein thermal conductivity measured in the solid state reveals anharmonic interactions of vibrations in a fractal structure. J Phys Chem Lett 5:1077–1082

Fujii N, Mizuno M et al (2014) Observing vibrational energy flow in a protein with the spatial resolution of a single amino acid residue. J Phys Chem Lett 5:3269––3273

Galperin M, Nitzan A et al (2007) Heat conduction in molecular transport junctions. Phys Rev B 75:155312

Gaskins JT, Bulusu A et al (2015) Thermal conductance across phosphonic acid molecules and interfaces: ballistic versus diffusive vibrational transport in molecular monolayers. J Phys Chem C 119:20931–20939

Giannoni M-J, Voros A et al (1991) Chaos and quantum physics. North-Holland, Amsterdam

Giri A, Niemela J-P et al (2016) Heat-transport mechanisms in molecular building blocks of inorganic/organic hybrid superlattices. Phys Rev B 93:115310

Gruebele M, Bigwood R (1998) Molecular vibrational energy flow: beyond the golden rule. Int Rev Phys Chem 17:91–145

Hassan S, Schade M et al (2014) Response of villin headpiece-capped gold nanoparticles to ultrafast laser heating. J Phys Chem B 118:7954–7962

Hopkins PE (2013) Thermal transport across solid interfaces with nanoscale imperfections: effects of roughness, disorder, dislocations and bonding on thermal boundary conductance (review article). ISRN Mech Eng 2013:682586

Ishikura T, Yamato T (2006) Energy transfer pathways relevant for long-range intramolecular signaling of photosensory protein revealed by microscopic energy conductivity analysis. Chem Phys Lett 432:533–537

Ishikura T, Iwata Y et al (2015) Energy exchange network of inter-residue interactions within a thermally fluctuating protein: a computational study. J Comput Chem 36:1709–1718

Keshavamurthy S (2013a) Eigenstates of thiophosgene near the dissociation threshold: deviations from ergodicity. J Phys Chem A 117:8729–8736

Keshavamurthy S (2013b) Scaling perspective on intramolecular vibrational energy flow: analogies, insights and challenges. Adv Chem Phys 153:43–110

Kondoh M, Mizuno M et al (2016) Importance of atomic contacts in vibrational energy flow in proteins. J Phys Chem Lett 7:1950–1954

Kuang S, Gezelter JD (2011) Simulating interfacial thermal conductance at metal-solvent interfaces: the role of chemical capping agents. J Phys Chem C 115:22475–22483

Kuharski RA, Chandler D et al (1988) Stochastic MD simulation of cyclohexane isomerization. J Phys Chem 92(11):3261–3267

Lee W, Kim K et al (2013) Heat dissipation in atomic-scale junctions. Nature 498:209–213

Lehmann KK, Scoles G et al (1994) Intramolecular dynamics from the eigenstate-resolved infrared spectra. Annu Rev Phys Chem 45:241–274

Leitner DM (1993) Real-symmetric random matrix ensembles of Hamiltonians with partial symmetry-breaking. Phys Rev E 48:2536–2546

Leitner DM (1999) Influence of quantum energy flow and localization on molecular isomerization in gas and condensed phases. Int J Quantum Chem 75(4–5):523–531

Leitner DM (2001a) Vibrational energy transfer and heat conduction in a one-dimensional glass. Phys Rev B 64:094201

Leitner DM (2001b) Vibrational energy transfer in helices. Phys Rev Lett 87:188102

Leitner DM (2002) Anharmonic decay of vibrational states in helical peptides, coils and one-dimensional glasses. J Phys Chem A 106:10870–10876

Leitner DM (2005) Heat transport in molecules and reaction kinetics: the role of quantum energy flow and localization. Adv Chem Phys 130B:205–256

Leitner DM (2008) Energy flow in proteins. Annu Rev Phys Chem 59:233–259

Leitner DM (2012) Mode damping rates in a protein chromophore. Chem Phys Lett 530:102–106

Leitner DM (2013) Thermal boundary conductance and rectification in molecules. J Phys Chem B 117:12820–12828

Leitner DM (2015) Quantum ergodicity and energy flow in molecules. Adv Phys 64:445–517

Leitner DM, Gruebele M (2008) A quantum model of restricted vibrational energy flow on the way to the transition state in unimolecular reactions. Mol Phys 106:433–442

Leitner DM, Pandey HD (2015a) Asymmetric energy flow in liquid alkylbenzenes: a computational study. J Chem Phys 143:144301

Leitner DM, Pandey HD (2015b) Quantum bottlenecks and unidirectional energy flow in molecules. Ann Phys 527:601–609

Leitner DM, Straub JE (2009) Proteins: energy, heat and signal flow. CRC Press/Taylor & Francis Group, Boca Raton

Leitner DM, Wolynes PG (1996a) Statistical properties of localized vibrational eigenstates. Chem Phys Lett 258:18–24

Leitner DM, Wolynes PG (1996b) Vibrational relaxation and energy localization in polyatomics: effects of high-order resonances on flow rates and the quantum ergodicity transition. J Chem Phys 105(24):11226–11236

Leitner DM, Wolynes PG (1997) Quantum energy flow during molecular isomerization. Chem Phys Lett 280:411–418

Leitner DM, Wolynes PG (1997a) Quantization of the stochastic pump model of Arnold diffusion. Phys Rev Lett 79:55–58

Leitner DM, Wolynes PG (1997c) Vibrational mixing and energy flow in polyatomics: quantitative prediction using local random matrix theory. J Phys Chem A 101(4):541–548

Leitner DM, Wolynes PG (2006) Quantum theory of enhanced unimolecular reaction rates below the ergodicity threshold. Chem Phys 329:163

Leitner DM, Köppel H et al (1994) Effects of symmetry breaking on spectra of chaotic Hamiltonian systems. Phys Rev Lett 73:2970–2973

Leitner DM, Levine B et al (2003) Quantum energy flow and trans-stilbene photoisomerization: an example of a non-RRKM reaction. J Phys Chem A 107:10706–10716

Leitner DM, Buchenberg S et al (2015) Vibrational energy flow in the villin headpiece subdomain: master equation simulations. J Chem Phys 142:075101

Lev YB, Reichman DR (2014) Dynamics of many-body localization. Phys Rev B 220201(R):89

Li Q, Duchemin I et al (2015) Mechanical tuning of thermal transport in a molecular junction. J Phys Chem C 119:24636–24642

Li Q, Strange M et al (2017) A strategy to suppress phonon transport in molecular junctions using π-stacked systems. J Phys Chem C 121:7175–7182

Logan DE, Wolynes PG (1990) Quantum localization and energy flow in many-dimensional Fermi resonant systems. J Chem Phys 93(7):4994–5012

Losego MD, Grady ME et al (2012) Effects of chemical bonding on heat transport across interfaces. Nat Mater 11:502–506

Luo T, Chen G (2013) Nanoscale heat transfer – from computation to experiment. Phys Chem Chem Phys 15:3389–3412

Majumdar S, Sierra-Suarez JA et al (2015) Vibrational mismatch of metal leads controls thermal conductance of self-assembled monolayer junctions. Nano Lett 15:2985–2991

Maradudin AA, Fein AE (1962) Scattering of neutrons by an anharmonic crystal. Phys Rev 128:2589–2608

Mehta ML (1991) Random matrices. Academic, San Diego

Meier T, Menges F et al (2014) Length-dependent thermal transport along molecular chains. Phys Rev Lett 113:060801

Nandkishore R, Huse DA (2015) Many-body localization and thermalization in quantum statistical mechanics. Annu Rev Cond Mat Phys 6:15–38

O'Brien PJ, Shenogin S et al (2012) Bonding-induced thermal conductance enhancement at inorganic heterointerfaces using nanomolecular monolayers. Nat Mater 12:118–122

Pandey HD, Leitner DM (2016) Thermalization and thermal transport in molecules. J Phys Chem Lett 7:5062–5067

Pandey HD, Leitner DM (2017a) Influence of thermalization on thermal conduction through molecular junctions: computational study of PEG oligomers. J Chem Phys 147:084701

Pandey HD, Leitner DM (2017b) Vibrational energy transport in molecules and the statistical properties of vibrational modes. Chem Phys 482:81–85

Qin Z, Bischof JC (2012) Thermophysical and biological responses of gold nanoparticle laser heating. Chem Soc Rev 41:1191–1217

Rubtsov IV (2009) Relaxation-assisted two-dimensional infrared spectroscopy (RA 2DIR) method: accessing distances over 10 Å and measuring bond connectivity patterns. Acc Chem Res 42:1385–1394

Rubtsova NI, Rubtsov IV (2015) Vibrational energy transport in molecules studied by relaxation-assisted two-dimensional infrared spectroscopy. Annu Rev Phys Chem 66:717–738

Segal D, Agarwalla BK (2016) Vibrational heat transport in molecular junctions. Annu Rev Phys Chem 67:185–209

Segal D, Nitzan A et al (2003) Thermal conductance through molecular wires. J Chem Phys 119:6840–6855

Sibert EL III, Gruebele M (2006) Molecular vibrational energy flow and dilution factors in an anharmonic state space. J Chem Phys 124:024317

Stocker KM, Gezelter JD (2013) Simulations of heat conduction at thiolate-capped gold surfaces: the role of chain length and solvent penetration. J Phys Chem C 117:7605–7612

Stocker KM, Neidhart SM et al (2016) Interfacial thermal conductance of thiolate-protected gold nanospheres. J Appl Phys 119:025106

Stuchebrukhov AA (1986) On the theory of intramolecular vibrational relaxation of polyatomic molecules. Sov Phys JETP 64(6):1195–1204

Stuchebrukhov AA, Marcus RA (1993) Theoretical study of intramolecular vibrational relaxation of acetylenic CH vibration for v=1 and 2 in large polyatomic molecules $(CX_3)_3YCCH$, where X=H or D and Y=C or Si. J Chem Phys 98(8):6044–6061

Swartz ET, Pohl RO (1989) Thermal boundary resistance. Rev Mod Phys 61:605–668

Uzer T (1991) Theories of intramolecular vibrational energy transfer. Phys Rep 199(2):73–146

Wang JS, Wang J et al (2008) Quantum thermal transport in nanostructures. Eur Phys J B 62: 381–404

Wilkinson M, Feingold M et al (1991) Localization and spectral statistics in a banded random matrix ensemble. J Phys A: Math Gen 24:175–181

Wu X, Varshney V et al (2017) How to characterize thermal transport capability of 2D materials fairly? – sheet thermal conductance and the choice of thickness. Chem Phys Lett 669:233–237

Xu Y, Leitner DM (2014) Vibrational energy flow through the green fluorescent protein-water interface: communication maps and thermal boundary conductance. J Phys Chem B 118: 7818–7826

Xu Y, Gnanasekaran R et al (2013) The dielectric response to photoexcitation of GFP: a molecular dynamics study. Chem Phys Lett 564:78–82

Yu X, Leitner DM (2003a) Anomalous diffusion of vibrational energy in proteins. J Chem Phys 119:12673–12679

Yu X, Leitner DM (2003b) Vibrational energy transfer and heat conduction in a protein. J Phys Chem B 107:1698–1707

Yu X, Leitner DM (2005) Heat flow in proteins: computation of thermal transport coefficients. J Chem Phys 122:054902

A Statistical Approach of Thermal Transport at Nanoscales: From Solid-State to Biological Applications

38

Yann Chalopin

Contents

Abstract

This collection of materials has been articulated to illustrate the idea that observing equilibrium short-scale processes and performing statistical measures through correlations contribute a great deal to revealing macroscopic thermal properties of condensed matter. The questions addressed here arise from a change of paradigm as soon as one comes to describe a system response at a scale shorter than the characteristic lengths of its underlying microscopic phenomena.

Y. Chalopin (✉)
EM2C Lab, Ecole CentraleSupelec/CNRS, Gif-sur-Yvette, France
e-mail: yann.chalopin@centralesupelec.fr

We restrict our analysis here to nanoscale heat transport by conduction or radiation and to a lesser extent in liquid and soft materials. A description of the mechanisms of phonon transmission at solid-state constrictions (or interfaces) is introduced. Infrared absorption in dielectric nano-objects is next discussed as well as a calculation of the absorption spectra in cellular membranes. Finally, the Brownian relaxation of magnetic nanoparticles is revisited in the light of accounting for long-range dipolar interactions.

1   A Brief Historical Review

As it is frequently the case in physical sciences, early developments come from experimental breakthroughs. Despite it was not a single man's contribution, the mercury thermometer can be considered as the first modern "device" capable to measure a quantity of energy: the temperature. Interestingly, the underlying concept of mean-free-path, which nowadays remains essential to formulate a microscopic description of transport phenomena, was first introduced by Clausius in 1857. In the middle of the nineteenth century, one realized that light would also allow a form of heat transfer, according to Kirchhoff and its formulation of emission and absorption of radiation, and shed the light, *Ex tempore* onto the concept of "black body." In a slightly different field, Einstein and Langevin discovered, respectively, the Brownian motion (1905) and the stochastic equation (1900), which are central in many areas of current research. Finally, Ritov (1959) did an important theoretical step forward by describing the structure of the radiation spectrum as being related to microscopic fluctuations of its current sources. Although this historical introduction remains shallow, it brings up most of the core concepts commonly discussed in this chapter: (1) Conduction occurs in solids and appears in the form of microscopic vibrations (the phonons). (2) Convection/diffusion in a gas or in a fluid corresponds to momentum exchanges in collisions between atoms and molecules. (3) Thermal radiation allows heat exchanges by photons between distant objects. These three modes of energy transfer are triggered by microscopic carriers that exhibit themselves characteristic time and length scales. Based on this very broad set, an important idea of this chapter is simply to demonstrate that observing the time fluctuations of such carriers (e.g., charge density, phonons, molecules) allows to predict the susceptibilities of these systems (and more specifically the dissipation properties). In this respect, we first introduce an atomistic formulation of the phonon conductance and transmission, to deal with heat transfer at molecular interfaces. To complement our approach, we present how infrared absorption in confined dielectrics can be calculated using molecular dynamics simulations. In both of these examples, we discuss the advantage of an atomistic formulation to unveil microscopic effects neglected when considering macroscopic properties. We next illustrate how such technics can be applied to nano-objects of biological relevance. We discuss the thermal properties of lipid bilayers which correspond to a model of a cellular membrane. Also, we address the problem of estimating thermal heat by coupled magnetic nanoparticle in a cellular (liquid) environment.

The practical application behind these topics is cancer hyperthermia, which of course encompasses many other areas of research.

2   Transport at Phonon Interfaces and Conductance in Nano-objects

Thermal properties of nanostructures are strongly influenced by phonon scattering at boundaries and interfaces (Carey et al. 2008; Cahill et al. 2003). Consequently, in recent years, interfacial thermal transport has become a major subject of interest with regard to two important technological applications, namely: thermoelectric materials development (Gao et al. 2010) and microelectronics thermal management. At the present stage of the physical understanding, two other key questions remain concerning heat conduction in materials with interfaces: what is the contribution of interfacial scattering compared to that of internal scattering in thick layer systems, and what are the mechanism of transmission and recombination of the vibrational modes at the contact?

2.1   A General Equilibrium Formulation of the Phonon Conductance at Molecular Interfaces

Thermal conductance at the nanoscale is extensively studied by various numerical approaches such as the Boltzmann transport equation, molecular dynamics (MD) simulations, or the atomistic Green's function (AGF) method (Mingo and Yang 2003; Zhang et al. 2007a, b) also referred as nonequilibrium Green's function (NEGF). As reported in the literature (Zhang et al. 2011; Schelling et al. 2002; Becker et al. 2006; Panzer et al. 2008; Ong et al. 2010), thermal conductance predictions based on MD simulations rely on a nonequilibrium procedure where a constant heat flux is fixed across the junction. From these considerations, we propose a formulation for calculating thermal conductance of any type of interfaces using an equilibrium approach analog to the well-known Green-Kubo (GK) method (Kubo et al. 1985) for thermal conductivity (Allen et al. 1987). We illustrate our technique on perfect Si/Ge superlattices and compared to experimental works.

2.1.1   Equilibrium Approach for Thermal Conductance Across an Interface

Let us assume two systems A and B exchanging energy and initially set at the same equilibrium temperature. The total Hamiltonian at equilibrium can be written as $H = H_A + H_B$, H_A and H_B being defined from a local energy at each atomic site as follows:

$$H_{A/B} = \sum_{i \in A/B} h_i = \sum_{i \in A/B} \frac{p_i^2}{2m_i} + V_i \qquad (1)$$

where p_i and m_i refer to the atom i momentum and mass. In the case of pair potential interactions, $V_i = \sum_{j \neq i} V(r_i - r_j)/2$, and half of the interaction belongs to H_A and to H_B. To obtain the interfacial thermal conductance, we use the linear response formalism (Pottier et al. 1999) where an infinitesimally small temperature difference is applied between the two systems A and B. $\Delta T = T_B - T_A$ is sufficiently small such that temperatures T_A and T_B in systems A and B are uniform. A and B are considered as thermal reservoirs. The static response function χ_o to the applied static temperature difference is defined by $<Q> = \chi_o \Delta T$, and taking the parameter $\lambda = 0$ provides the relationship between the static response and (*equal-time*) correlations in the system:

$$\chi_o = \frac{< Q H_B >}{k_B T^2} \tag{2}$$

Setting $Q = \dot{H}_A$ allows to find the rate of heat added to the subsystem *A* per unit time when a temperature difference ΔT is applied between A and B. The definition of the thermal conductance $\mathcal{G}$ is:

$$\frac{d < H_A >}{dt} = \mathcal{G} \Delta T \tag{3}$$

Using Eq. (2) obtains:

$$\mathcal{G} = \frac{1}{k_B T^2} < \frac{d H_A(t)}{dt} H_B(t) > . \tag{4}$$

2.1.2   Use of Equations of Motion to Compute $\mathcal{G}$

The Hamiltonian equations of motion can be used to obtain the time derivative of H_A:

$$\frac{d H_A(t)}{dt} = \sum_{i \in A \cup B} \left(\frac{\partial H_A}{\partial p_i} \cdot \dot{p}_i + \frac{\partial H_A}{\partial q_i} \cdot \dot{q}_i \right)$$

$$\dot{p}_i \equiv -\frac{\partial H}{\partial q_i} = F_i^A + F_i^B \tag{5}$$

$$\dot{q}_i \equiv \frac{\partial H}{\partial p_i} = \frac{p_i}{m_i} = v_i \tag{6}$$

$$\frac{\partial H_A}{\partial q_i} = -\left(F_i^A + \frac{1}{2} F_i^B \right) \text{ for } i \in A \tag{7}$$

$$\frac{\partial H_A}{\partial q_i} = -\frac{1}{2} F_i^A \text{ for } i \in B \tag{8}$$

F_i^S being the force from the side S on atom i. We include half of the $A - B$ interaction energy in H_A and the other half in H_B. Hence,

$$\frac{dH_A(t)}{dt} = \frac{1}{2}\left[\sum_{i \in A} v_i \cdot F_i^B - \sum_{j \in B} F_j^A \cdot v_j\right] \tag{9}$$

which corresponds to the net power (or heat flux) exchanged by system A. We can also derive a more familiar-looking formula by considering $H_B(t) = \int_{-\infty}^{t} \dot{H}_B(t')\,dt'$. The thermal conductance can be expressed as a time integral over the dissipated power correlator:

$$\mathcal{G} = \frac{1}{k_B T^2} \int_{-\infty}^{t} < \dot{H}_A(t)\dot{H}_B(t') > \, dt' \tag{10}$$

$$= \frac{1}{k_B T^2} \int_{0}^{\infty} < \dot{H}_A(\tau)\dot{H}_A(0) > \, d\tau \tag{11}$$

which looks like the usual GK formula.

Figures 1 and 2 illustrate our methodology by reporting autocorrelation function and the thermal conductance obtained for a $5 \times 5 \times 40$ Si/Ge superlattice at 440 K using the Stillinger-Weber potential (Stillinger and Weber 1985). Based on the assumption that the transport is ballistic within a period, interface scattering is predominant in Si/Ge, and the thermal conductivity of the superlattices is obtained by using the relation:

$$\lambda = GL/2, \tag{12}$$

Fig. 1 Net heat flux autocorrelation function at 440 K. Inset shows a zoom of the 0–2 ps time interval
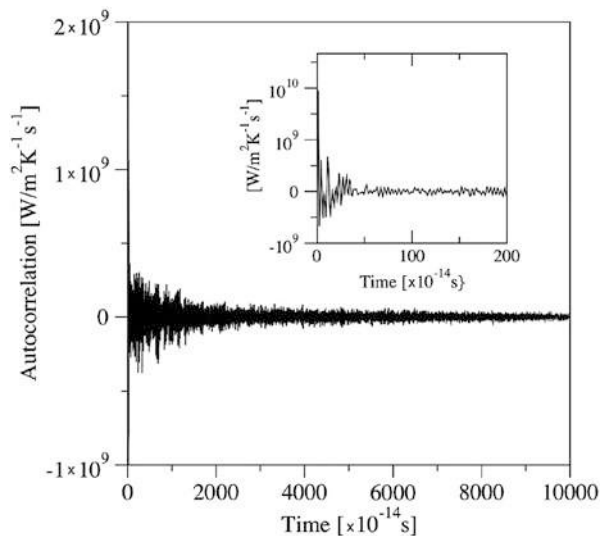
Fig. 2 First and second
neighbor contributions to the
thermal conductance of a
Si/Ge junction at 440 K
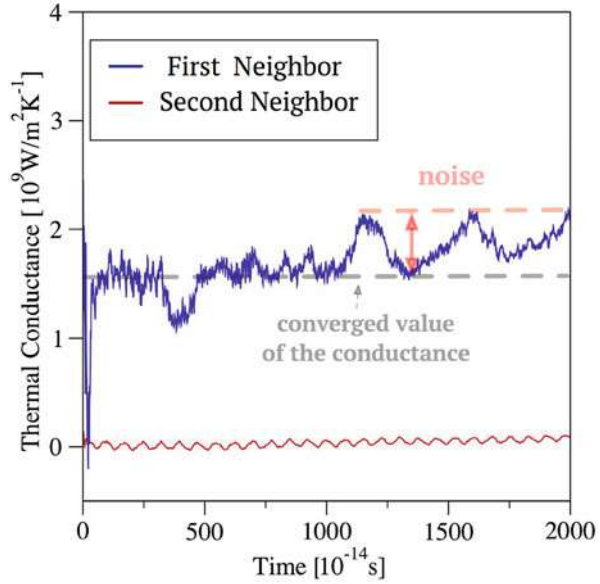
Fig. 3 Green's function
calculation at a single Si/Ge
interface (Zhang et al. 2007c)
(Green). Computed and
measured (Lee et al. 1997;
Borca et al. 2000) interface
thermal resistance as a
function of the period
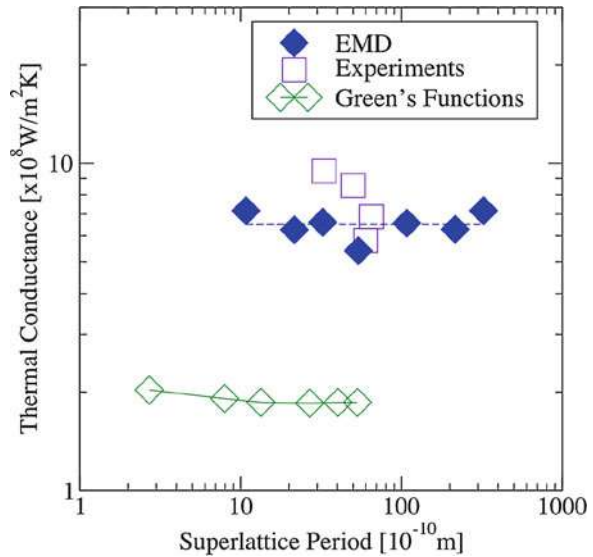
where G is the thermal conductance calculated by MD, L being the superlattice
period, and λ its corresponding thermal conductivity (Fig. 3).

It can be seen that above a critical thickness, phonon transport in the superlattice
becomes strongly impacted by the defects included during the growing/deposition
process in the fabrication step. This additional effect has not been included in MD
simulations, and the trend of the thickness dependence can no longer be compared

with experiments. EMD provides in this case only a good estimation of the order of magnitude of the thermal boundary resistance. The next part of this note is devoted to extend this formulation with a frequency dependence.

2.2   Spectral Conductance: An Equilibrium Approach

This section presents an approach for estimating the frequency vs. wave-vector-dependent phonon transmission across a solid-solid interface. We show that the spectral properties of the heat flux can be generally deduced from the equilibrium displacement fluctuations of the contact atoms. We have applied and demonstrated our formalism with molecular dynamics simulations to predict the angular and mode-dependent transmission. We first carry the study to the case of semiconductor interfaces (silicon and germanium thin films).

2.2.1   Derivation

To express the frequency and wave-vector-dependent phonon transmission, we start from the expression of the heat flux obtained from the Hamiltonian equation of motion (Chalopin et al. 2012b) of two interacting phonon systems whose Hamiltonians are H_A and H_B with $Q = - <\dot{H}_A> = <[H, \ H_A]>$,

$$Q(t) = 1/2 \sum_{\substack{i \in A, j \in B \\ \alpha, \delta \in \{x,y,z\}}} k_{i,j}^{\alpha,\delta} \left(\left\langle \dot{u}_i^\alpha(t) u_j^\delta(t) \right\rangle - \left\langle u_i^\alpha(t) \dot{u}_j^\delta(t) \right\rangle \right). \tag{13}$$

u_i refers to the atomic displacement of atom i, $\dot{u}_i$ the instantaneous velocity of atom i. The α and δ exponents refer to the x, y, or z component. $k_{i,j}^{\alpha,\delta}$ is the interatomic force constant between atoms i and j. Rewriting the correlation $\left\langle \dot{u}_i^\alpha(t) u_j^\delta(t) \right\rangle$ as

$$\left\langle \dot{u}_i^\alpha(t) u_j^\delta(t) \right\rangle = \lim_{t' \to t} \frac{d}{dt'} \left\langle u_i^\alpha(t') u_j^\delta(t) \right\rangle = \lim_{\tau \to 0} \frac{d}{d\tau} \left\langle u_i^\alpha(\tau) u_j^\delta(0) \right\rangle,$$

introduces a characteristic time $\tau = t - t'$ which is related to the vibrational dynamics. Applying the Fourier transform on the correlations yields

$$Q(\omega) = 1/2 \sum_{\substack{i \in A, j \in B \\ \alpha, \delta \in \{x,y,z\}}} k_{i,j}^{\alpha,\delta} \left[\dot{u}_i^\alpha(\omega) u_j^{*\delta}(\omega) - u_i^{*\alpha}(\omega) \dot{u}_j^\delta(\omega) \right]. \tag{14}$$

Here, u^* is the complex conjugate of u. Equation (14) can be compared to the following Landauer expression (Landauer 1970):

$$Q_{LD}(\omega) = \hbar\omega(n_a^0(\omega) - n_b^0(\omega))T(\omega), \tag{15}$$

where $\hbar\omega$ stands for the energy quantum and T for the phonon transmission. $n_{a,b}$ corresponds to the Bose-Einstein occupation numbers of systems A and B, respectively. Identifying the first term of the difference of Eq. (14) to the one in Eq. (15) and considering the classical limit ($\hbar\omega \ll k_B T$) allow us to extract the following frequency-dependent transmission:

$$T(\omega) = \beta/2 \sum_{\substack{i \in A, j \in B \\ \alpha,\delta \in \{x,y,z\}}} k_{i,j}^{\alpha,\delta} \left[\dot{u}_i^\alpha(\omega) u_j^{*\delta}(\omega) - u_i^{*\alpha}(\omega) \dot{u}_j^\delta(\omega) \right], \qquad (16)$$

which can be simplified into

$$T(\omega) = \beta \mathrm{Re} \left\{ \sum_{\substack{i \in A, j \in B \\ \alpha,\delta \in \{x,y,z\}}} k_{i,j}^{\alpha,\delta} \dot{u}_i^\alpha(\omega) u_j^\delta(\omega) \right\} \qquad (17)$$

where $\beta = 1/k_B T$ and k_B denotes the Boltzmann constant. Equation (17) highlights that the atoms involved are those in interaction across the interface. A further step can be taken by considering the atomic velocities in the wave-vector reciprocal space defined by:

$$\dot{u}_{\mathbf{k}}^\alpha(\omega) = \sum_i \dot{u}_i^\alpha(\omega) e^{i\mathbf{k}\cdot\mathbf{r}_i^0},$$

where $\mathbf{r}_i^0$ refers to the equilibrium position of atom i. Replacing the velocities by their expressions in Eq. (17), the transmission from wave-vector $\mathbf{k}$ to $\mathbf{k}'$ at frequency ω can be written as:

$$T(\omega, \mathbf{k}, \mathbf{k}') = \beta/2i\omega \sum_{\alpha,\delta \in \{x,y,z\}} \dot{u}_{\mathbf{k}}^\alpha(\omega) \dot{u}_{\mathbf{k}'}^{*\delta}(\omega) D^{\alpha,\delta}(\mathbf{k}, \mathbf{k}'), \qquad (18)$$

where $D^{\alpha,\delta}$ stands for the interfacial phonon dynamical matrix as:

$$D^{\alpha,\delta}(\mathbf{k}, \mathbf{k}') = \sum_{i \in A, j \in B} k_{i,j}^{\alpha,\delta} e^{i(\mathbf{k}\cdot\mathbf{r}_i^0 - \mathbf{k}'\cdot\mathbf{r}_j^0)}.$$

2.2.2 Example on a Si/Ge/Si Heterostructure

From Eq. (17), we have calculated the Si/Ge transmission for L ranging between 0.5 and 10 nm as well as that of the perfect Si/Si transmission.

Results are reported on Fig. 4 inset (a), (b), and (c) highlighting two important features. Inset (a) shows that the optical modes coming from the Si substrate above the frequency $\nu > 15$ THz are poorly transmitted while a significant contribution to transmission (noted S on inset (b)) is observed at $\nu \sim 14$ THz. This later peak does
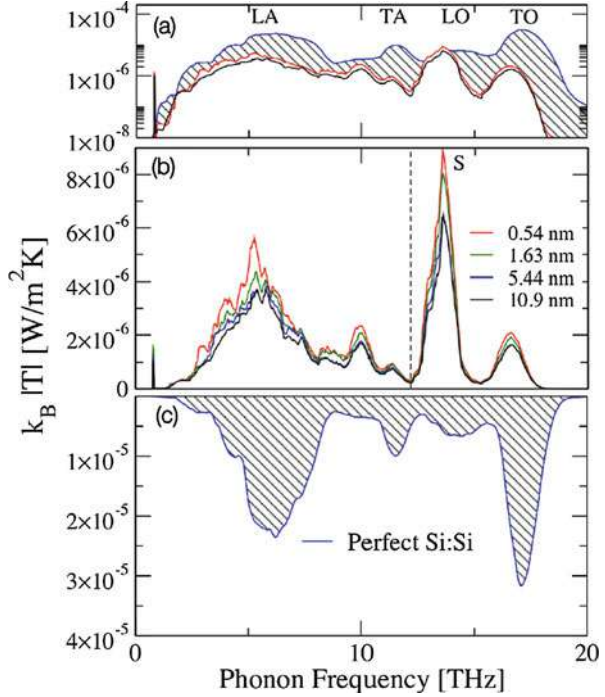
not appear in perfect bulk silicon (inset (c)) because it corresponds to an interface
mode. Additional modes are transmitted above the Ge bulk cutoff frequency ν_{Ge}
(Fig. 4 inset (a) represented by the dotted vertical line). This indicates that inelastic
processes occur at the contact layers with additional transport channels transferring
thermal energy above ν_{Ge}. The degree of accuracy of this microscopic description
is benchmarked by comparing the total thermal resistance of the junction to that
retrieved by more conventional approaches reported in the literature. We have
compared the thermal resistance R obtained from

$$R^{-1} = k_B \int_\omega T(\omega)d\omega, \tag{19}$$

to nonequilibrium molecular dynamics (NEMD) simulations (Landry and
McGaughey 2010) and AGFs (Zhang et al. 2007c) calculations. Both calculations
rely on the same interatomic potential (Stillinger and Weber 1985). As seen on
Fig. 5, thermal resistance increases with the layer thickness. The NEMD data trend
(red crosses) is recovered (blue triangles) within a satisfying range of accuracy.
AGFs overestimate the thermal resistance by less than 30 percent when compared
to other predictions. The harmonic feature of the theory can be considered at the
origin of this mismatch.

Figure 6 reports the wave-vector **k** and frequency-dependent transmission
$T(\omega, \mathbf{k})$ when Ge layers have two different thicknesses. **k** is constructed in the

Fig. 5 Thermal resistance obtained from the transmission of Fig. 4 and compared to the NEMD direct calculation (red circle) and AGFs (dark square) for several layer thicknesses
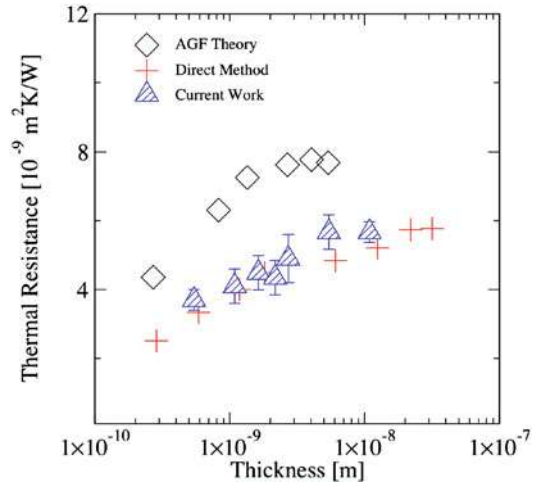
Fig. 6 Wave-vector-dependent phonon transmission calculated from MD at T = 400 K for Ge layer thicknesses of L = 0.54 nm (**a**) and L = 5.4 nm (**b**)
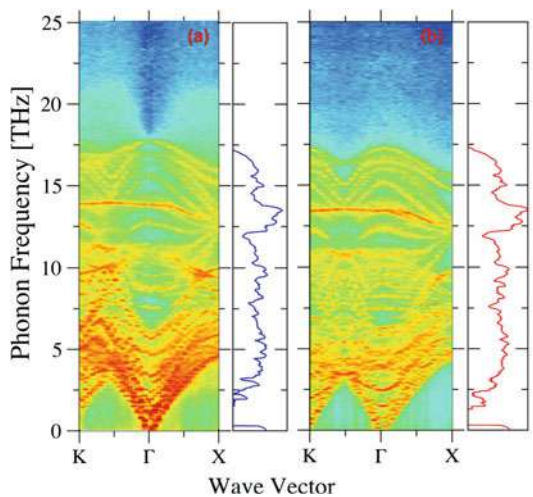
plane defined by the [100] and [110] directions. The right-hand side graphs represent the total transmissions previously calculated. A significant part of the transmitted energy arises in the low frequency range, i.e., below the Ge optical mode frequencies, where phonon modes exist in both materials. The localized interface mode (S) is found to be weakly dispersive and, as deduced from the spectral distribution, has important contributions to thermal transfer at the interface.

2.3 Concluding Remarks

An equilibrium formulation of thermal interface conductance has been derived according to the linear response theory for molecular dynamics simulations. It can

be applied to a wide range of systems, including the effect of anharmonicity which is central in thermal transport. While conductance is related to the fluctuations of the net heat flux obtained by the contact atoms, interface transmission can be viewed as the cross power spectral density of these later. A wave-vector decomposition and a detailed analysis of the modes that govern the phonon transmission for Si/Ge interfaces were presented. It turns out that the interfacial transport cannot be accurately estimated from bulk properties. The present formulation of the interfacial transport has the advantage of remaining general and rather simple to implement.

3   Anomalous Susceptibilities in Dielectric and Magnetic Nanoparticles

The description of the optical properties of nanomaterials involves two classes of physical systems: metals and dielectrics. At the nanoscale, both exhibit particularly interesting properties due to the resonances activated by the confining boundary conditions – e.g., surface phonon-polariton for dielectrics and surface plasmon for metals. Understanding the mechanism of coupling between light and nanomaterials forms the cornerstone of the fields of nanophotonics and nano-optics, and a large panel of applications have recently emerged such as the enhancements of Raman scattering (Nie and Emory 1997), the fluorescence imaging (Shafiei et al. 2013), or the control of coherent thermal radiation (Greffet et al. 2002). When studying how size confinement impacts optical properties at short scales, it has been reported that nonlocal effects take place in metals (McMahon et al. 2009; David and Garcia de Abajo 2011) such as the spill- out of electrons (Voisin et al. 2001; Gaudry et al. 2000): For systems smaller than 10nm, a strong increase of the intrinsic electron-phonon interaction can be observed. On the other hand, very few work reported peculiar size effects in dielectric material (Chalopin et al. 2012a) although there has been much effort devoted to exploring optical modes of vibration in these systems (Jones and Fuchs 1971; Tong and Maradudin 1969). This section is turned toward the study of the mechanisms of infrared absorption in confined dielectrics with the mean of molecular dynamics (MD) simulations. We found that light absorption in the THz occurs predominantly through anomalous surface modes which are reported to be confined in space, demonstrating that macroscopic description of electrodynamics fails at He nanoscale. This particular feature is illustrated on solid as well as in soft matter.

3.1   Microscopic Models of Nonlocal Dielectric Constant

Our approach is formulated on the basis of the linear response theory where the bulk susceptibility $\chi^{\alpha,\beta}(\omega, \mathbf{k}) = \varepsilon_r^{\alpha,\beta}(\omega, \mathbf{k}) - 1$ subjected to an external electric field $\mathbf{E}$ can be written as:

$$P^{\alpha}(\omega, \mathbf{k}) = \varepsilon_0 \chi^{\alpha, \beta}(\omega, \mathbf{k}) E^{\beta}(\omega, \mathbf{k}).$$

ε is the material dielectric tensor and P the polarization induced by E. α, β stands for the space coordinates. For any anisotropic materials, the fluctuation-dissipation theorem (Kubo 1966) allows to generally express χ as (Chalopin et al. 2011):

$$\chi^{\alpha, \beta}(\omega, \mathbf{k}) = \frac{1}{\varepsilon_0 k_B T V} \left[\left\langle P^{\alpha}(0, \mathbf{k}) P^{\beta}(0, \mathbf{k}) \right\rangle + i\omega \int\limits_0^{\infty} e^{i\omega t} \left\langle P^{\alpha}(0, \mathbf{k}) P^{\beta}(t, \mathbf{k}) \right\rangle dt \right],$$
(20)

where V is the volume of the dielectric system and $P(t)$ the polarization state at time t. In the reciprocal space P is expressed as:

$$P^{\alpha}(t, \mathbf{k}) = \sum_{i \in V} q_i \cdot \mathbf{r}_i^{\alpha}(t) e^{i\mathbf{k} \cdot \mathbf{r}_i^0},$$
(21)

$\mathbf{r}_i^0$ is the equilibrium position of atom i caring the charge q_i. Hence, the optical spectra of any polar material are extracted from the fluctuations of the system at thermal equilibrium, and the nonlocal description of the dielectric properties corresponds to the 2D Fourier transform of the charge fluctuations. Recalling that ε is complex $\varepsilon_r = \varepsilon_r' + i\varepsilon_r''$, one has $\varepsilon_r' = \text{Re}\{\chi\} + 1$ and $\varepsilon_r'' = \text{Im}\{\chi\}$. Using the convolution theorem, one can thus retrieve the absorption spectra through the imaginary part of the polarizability tensor $\alpha = \alpha' + i\alpha''$ as:

$$P^{\alpha}(\omega, \mathbf{k}) = \alpha^{\alpha, \beta}(\omega, \mathbf{k}) E^{\beta}(\omega, \mathbf{k})$$

with

$$\alpha''^{\alpha, \beta}(\omega, \mathbf{k}) = \frac{\omega}{k_B T V} \frac{P^{\alpha}(\omega, \mathbf{k}) P^{*\beta}(\omega, \mathbf{k})}{\tau}$$
(22)

3.2   Surface-Enhanced Infrared Absorption in Dielectric Thin Films with Ultra-strong Confinement Effects

3.2.1   The Bulk Dielectric Constant: An Example with MgO

We validate our approach by comparing the dielectric constant at $\mathbf{k} = 0$ with the available experimental data (see upper Fig. 7). The Matsui (1989) potential has been used. We have calculated the nonlocal (i.e., k-dependent) dielectric constant (Eq. 22) with $\mathbf{k}$ ranging along the main directions lying within first Brillouin zone (FBZ). Figure 7 reports the results obtained for $\mathbf{k}$ ranging along the points Γ [000], Σ [001], X [011], and L [111]. This physical process is already known and it has been demonstrated experimentally (Scarel et al. 2008; Fray et al. 1961) 8 cm.
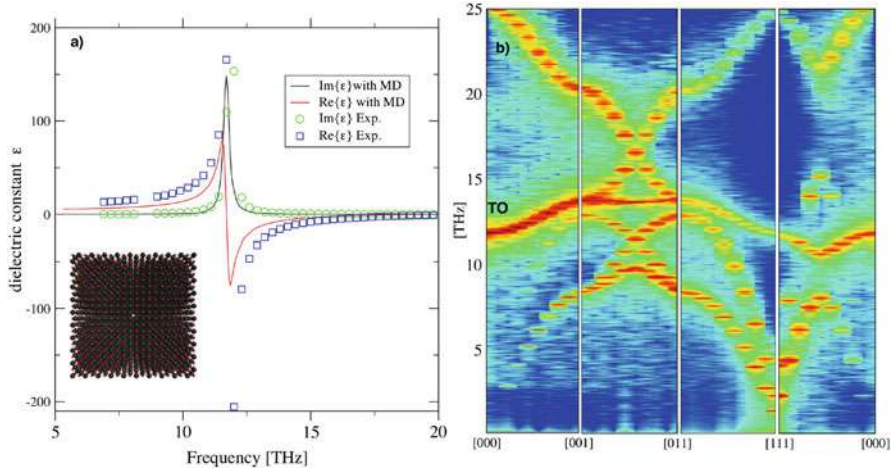
Fig. 7 Top figure: frequency-dependent bulk dielectric constant compared to Palik and Bennett (1991). Bottom figure represents the corresponding nonlocal description of the absorption along the main direction of the MgO first Brillouin zone. A log representation of Eq. 22 has been performed for more contrast

3.2.2   Confinement Effects and Surface Mode

We next consider the case of crystal slabs with finite thicknesses L in the [001] direction and periodic in two lateral ones ([100 and [010]). The translational symmetry preserved along two axes allows a wave-vector description of the in-plane mode $\mathbf{K} = (\mathbf{k_x}, \mathbf{k_y})$ as $\varepsilon^{\alpha,\beta}(\mathbf{K}, \mathbf{q}, \omega)$. $\mathbf{q}$ stands for the modes along the [001] confined direction. As for the bulk case, we first look at the absorption at long wavelength limit for thicknesses (L) ranging from 0.42 to 4.2 nm. An atomic representation of a 0.84 nm thick slab is illustrated in Fig. 8 as well as the calculated absorption spectra (Imα) B. It can be seen that IR absorption occurs through mainly two bands having particular resonances spreading around 12 and 24 THz. These two bands lie at the bulk TO and LO frequencies, respectively, and interestingly their structure strongly depends on the size:

3.2.3   Limit of the Classical Electrodynamics at the Nanoscale: The Dielectric Screening Length

From this study, we see that the 12 THz band corresponds to a competition between the TO polariton mode and a surface mode (also known as Lucas mode Chen et al. 1972; Lucas 1968) strongly localized at the topmost atomic layers: At the smaller thickness, absorption is driven by this surface mode, whereas the TO requires the system to have a sufficient number of layer to get activated: When increasing the layer thickness above 1 nm, the TO polariton starts to dominate over the surface mode. In a similar fashion, the optical absorption at higher frequencies (20–25 THz) corresponds to localized surface modes for the smaller thicknesses: The contribution of these resonances attributed to the surfaces get progressively replaced by that
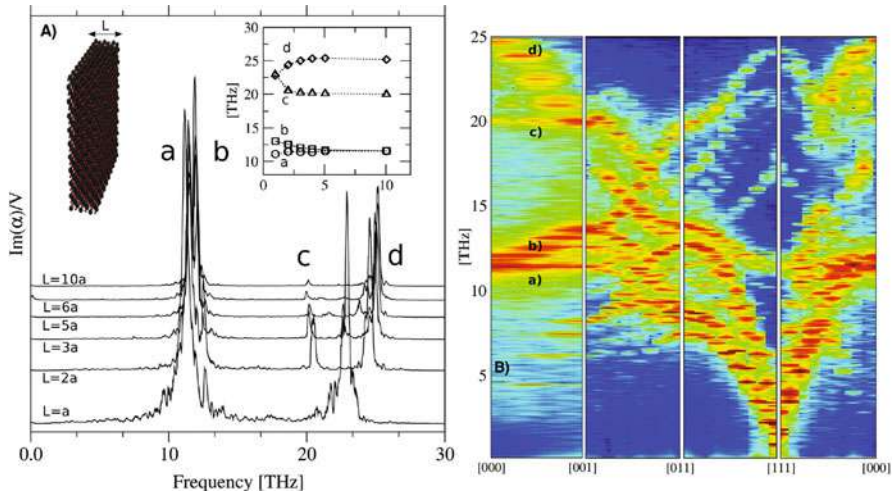
Fig. 8 Top figure (**a**): Size-dependent dielectric constant with thicknesses ranging from 1 to 10 MgO unit cells. Bottom figure represents the corresponding nonlocal description for the size of 3 unit cells with a log representation of Eq. 22 for more contrast

of the LO polariton for thickness >1 nm which is a volume mode (see bottom of Fig. 8): This effect arises from the fact that the topmost atomic layer undergoes interatomic force constants much weaker than that of the volume layers where the atoms are isotropically bonded and consequently their vibrational dynamics differs.

An important aspect of this analysis allows to conclude that none of the slabs' absorption spectra can be deduced from the calculated bulk dielectric constant; in particular, the well-known Clausius-Mossotti relation predicts a polariton resonance for $\epsilon = -1$ that is at the frequency of 17 THz if we consider the bulk ϵ calculated in Fig. 7. In fact the bulk LO polariton mode is found to occur at the frequency of 25 THz for all thickness which can be attributed to the effect of dielectric screening with a characteristic distance corresponding to the lattice parameter. Consequently, any macroscopic description of IR absorption in thin film clearly breaks down at this scale.

3.3   Infrared Absorption in Cellular Membrane Models Made of Lipid Bilayers

The dielectric properties of lipid bilayers have been studied for the past years for their central role in the protein and ion permeability of the membrane as well as electroporation. It has been reported that in lipid bilayers, the static dielectric permittivity is spatially nonuniform, but no work has addressed so far the infrared properties in the terahertz (THz) regime. The polarizability of two types of bilayers made of DPPC and POPC lipids is presented here. The two lipids have very similar
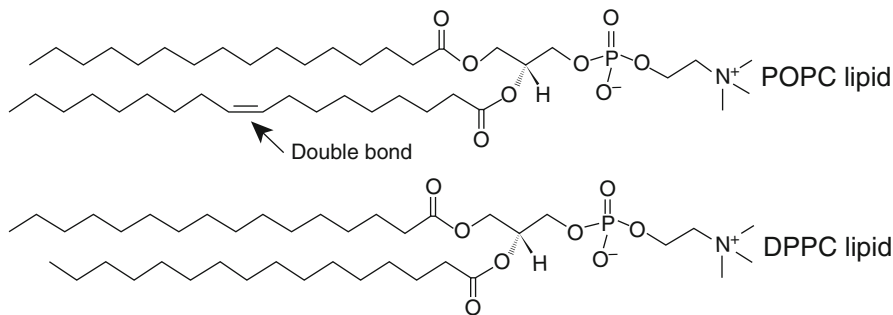
Fig. 9 Simplified Lewis formula of DPPC and POPC lipids

chemical formulas, and despite these similarities, their organization in bilayers displays dissimilar structure, which in turn dramatically affects the THz spectra. The calculations have been compared to Raman experiments and show a fair agreement.

3.3.1 Atomistic Modeling of the Lipids and Bilayers

The Lipid bilayers investigated here are made of two different lipids, namely, the 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and the 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC). Their chemical formulas are depicted in Fig. 9. The DREIDING (full atomistic) potential has been accounted for the lipids with an experimentally validated TIP3P interaction model for the water molecules (Jorgensen et al. 1983). Optical and morphological properties are extracted from the displacement and velocity fluctuations of the atoms in the micro-canonical ensemble (constant energy and volume) to ensure the validity of the fluctuation/dissipation theorem (FDT). Figure 10 depicts an atomic view of the relaxed bilayers at the temperature of 310 K. The resulting morphologies are in agreement with previous calculations reported in the literature.

3.3.2 Dielectric Properties

The static polarizability is first studied and compared to the literature. To confront our calculations with experiments, we have extracted the electrical capacitance C of the membrane according to $C = \frac{\alpha(\omega=0)\epsilon_0 S}{d}$, where d corresponds to the layer thickness. Various measurements have been reported in the literature (Benz et al. 1975; Toyama et al. 1991; Berestovsky et al. 1978; Montal and Mueller 1972; Lee and Bain 2005), where C ranges from 0.01 to 0.03 F.m^2 which is in agreement with our calculations (Fig. 11 (full lines)). Note that the capacitance does not vary in the same way as the polarizability due to change in volume by the temperature. More interestingly Fig. 11 reports that DPPC and the POPC bilayers have a static response which differs by a factor of 3. The imaginary part of the polarizability, $\alpha''(\omega)$ is shown in Fig. 12 (black line) and compared with Raman experimental data (Lee and Bain 2005) for a spectra ranging above 40 THz. Below that frequency, no data have been reported so far. The origin of the absorption can be attributed to specific
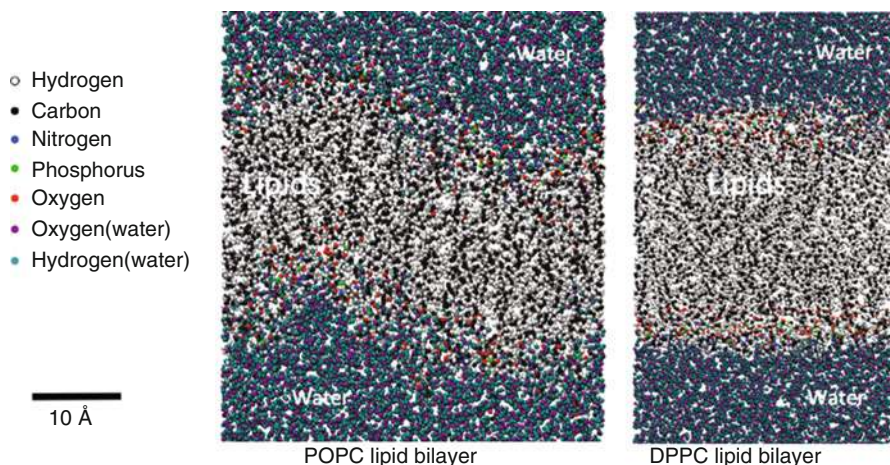
Fig. 10 Atomic view of POPC and DPPC lipid bilayers

Fig. 11 Cross-plane static polarizability and electrical capacity of DPPC and POPC bilayers as a function of temperature. Inset: Size effect (the number of lipids in the simulation box) on the static polarizability
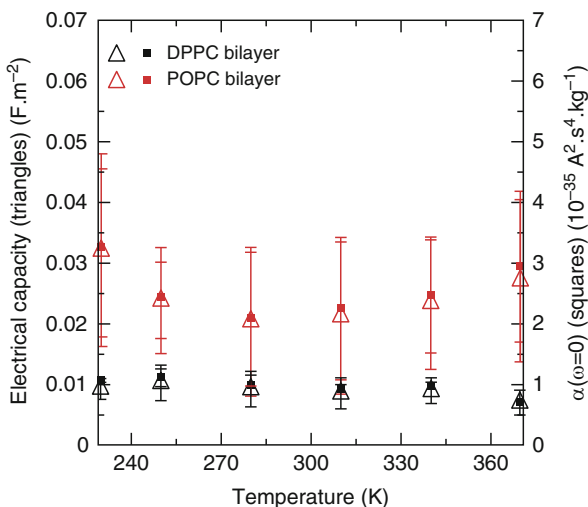
dipoles where, for instance, the bonds C-H and N-H are activate, respectively, at 85 and 100 THz. A very satisfactory agreement is found for those modes. The 50 THz band corresponds to the carbon bonds such as C-C, C-O, and C-N. The P-O bonds also lie in this frequency domain.

3.3.3 Interplay Between Absorption and Morphology

The IR properties of the bilayers are investigated here to identify where the absorption occurs in both configurations and what are the associated resonances.

Fig. 12 Imaginary part of the permittivity in DPPC and POPC bilayers and the comparison with Raman experimental data (Lee and Bain 2005) and the vibrational densities of states (DOS)
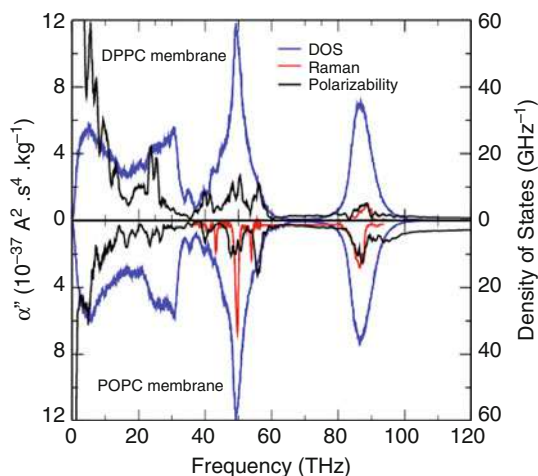
We have thus computed (Fig. 13) the densities of chemical elements forming each bilayers as well as the radial distribution function (Fig. 14). The density drop at the center of the DPPC bilayer (see Fig. 13 black arrow) indicates less interdigitation between the two lipid layers leading to a 0.5 nm thicker bilayer than the POPC as reported in the literature (Karjiban et al. 2013; Vermeer et al. 2007; Kupiainen 2005). In addition, the radial distribution function of the two membranes, Fig. 14, clearly indicates that the ordering of the POPC lipids is significantly reduced compared to that of the DPPC lipids: The density probability of the former becomes flatter behind the first neighbor peak. Furthermore, the first neighbor distance becomes larger for the case of POPC with a lower average density (i.e., tail of the correlation function). This is explained by the effect of the double bond at the tail of the POPC lipids that prevents perfect stacking. Thus the average distance between lipids in the same layer is larger which results in a lower surface density. This reduced surface density, which concomitantly involves a thinner bilayer, is not sufficient to fully explain the weaker static polarizability of the DPPC layers. The charge density is predominantly distributed at the hydrophilic heads (Fig. 13), and considering the highest mobility of the lipids, we can deduce that the local dipoles exhibit more angular fluctuations for the case of the POPC bilayer. To unveil this effect, we have computed the projection of the dipole moment along the surface ($p_{||}$) versus the perpendicular direction (p_z); hence the dipoles orientate predominantly in the planes as $\frac{p_z}{p_{||}} = 0.0309$ for POPC vs 0.0017 for DPPC: The difference in angular fluctuations leads to a higher projected dipole component (p_z) for the case of POPC. This explains the larger static polarizability.

It is possible to compute the spatially dependent absorption spectra in a direction perpendicular to the lipid planes with a 0.2 nm thick resolution (Fig. 15). The top

Fig. 13 Per-atom density
and charge density profile in
DPPC and POPC lipid
bilayers
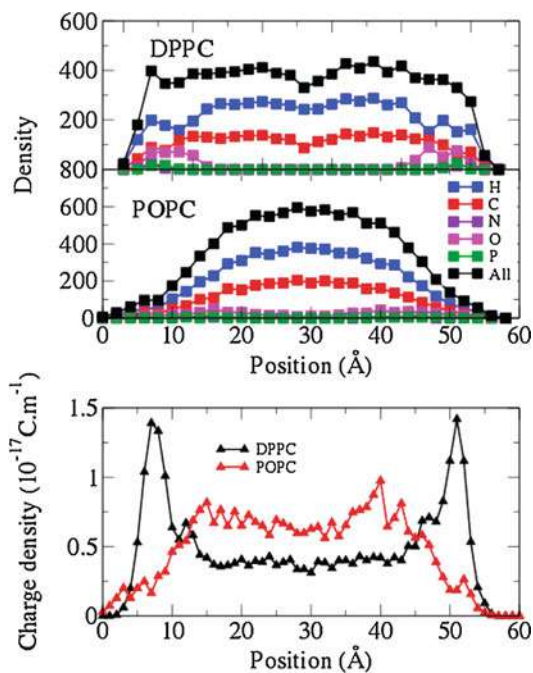
Fig. 14 In-plane radial
distribution function of the
DPPC and POPC lipid
bilayers
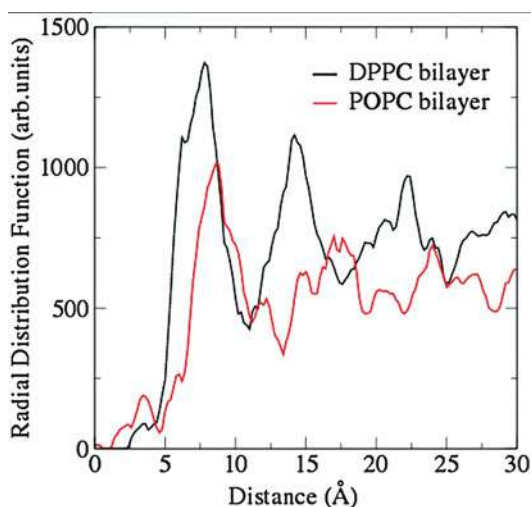
of Fig. 15 presents the total absorption integrated over the whole spectrum. It
demonstrates that the DPPC bilayer absorption is mainly localized in the heads
of the lipids (40 and 85 Å), whereas for the case of POPC, the maximum of the
absorption occurs more homogeneously within the bilayer.

Fig. 15 Total absorption and absorption at 53 THz as a function of the position in the bilayer
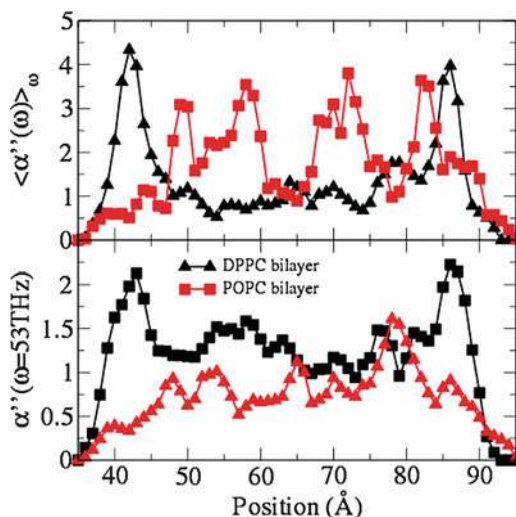
3.4 Concluding Remarks

To conclude, it has been demonstrated that optical absorption in polar nanomaterial is driven by anomalous absorption processes that cannot be predicted by any macroscopic description of electrodynamics. These optical properties are extremely confined in space and predominant in absorbing IR radiation. This approach provides a possible tool for predicting radiative properties of biological systems as well as in solid-state systems.

4 Brownian Dissipation by Magnetic Nanoparticles

Clusters of magnetic nanoparticles have received considerable interest in various research fields. Their capacity to generate heat under an alternating magnetic field has recently opened the way to applications such as cancer therapy by hyperthermia. The physics underlying the dynamics of interacting dipoles in systems of microscopic dimension is so rich that tremendous efforts have been devoted to address this subject from experimental (MacIsaac et al. 1996; Morup and Tronc 1994) and theoretical (Dejardin 2011) approaches. Predicting the dynamics of the dipolar sets, comprising magnetic and dielectric nanostructures, is an important challenge as the long-range and anisotropic-dipolar interactions (Corato et al. 2014) entail a wealth of nonindependent degrees of freedom, difficult to track analytically. The developments of both numerical and analytical methods assessing the particles' response to an external field (Branquinho et al. 2013; Ruta et al. 2015; Tan et al. 2014), and

powerful enough to handle the complexity of the couplings, are still to be found. Heat generation is driven by two processes: The Neel and the Brownian relaxation. Regarding the former, many works have reported a shortening of the magnetization relaxation time when increasing the dipolar interactions (Dejardin 2011; Berkov and Gorn 2001; Denisov and Trohidou 2001; Brinis et al. 2014). In addition, various studies have further addressed the case of two-dimensional square arrays (Denisov and Trohidou 2001; Brinis et al. 2014; Denisov et al. 2003; Sugano et al. 2007; Stariolo 2008; Bahia 2004; Farrell et al. 2004) which are suited for experimental investigations (Stariolo 2008). In a similar fashion, dielectric relaxation in systems subjected to rotational Brownian motion remains an important question (Jimenez et al. 2006; Zhou and Bagshi 1992) . It is relevant to the study of dynamic light scattering, Raman scattering, and liquid crystals. Some analytical approaches have accounted for approximations in which the many-body system is casted in terms of a two-body problem (Felderhof 2002) or a single dipole in a mean field.

Here, our interest focuses on the problem of evaluating the effect of the interactions between rigid dipoles carried by particles damped by a viscous torque (Coffey and Kalmykov 2012). Our analysis thus aims at identifying to what extent the Brownian relaxation process is affected by the particle interactions without any restriction on their density. For this purpose, we have formulated a statistical approach based on numerical simulations that reproduce the equilibrium orientational fluctuations of dipoles interacting in a heat bath. Hence, by means of the fluctuation-dissipation theorem, we have predicted the frequency-dependent imaginary part of the susceptibility of the system, subjected to an external ac field $\vec{B}_1$.

4.1 Statistical Modeling with the Langevin Equation

We considered a square array of magnetic rigid dipoles $\vec{\mu}_i$ carried by spherical particles with volume V and inertia momentum I, immersed in a liquid with viscosity η at the temperature of 320 K. With the help of the fluctuation-dissipation theorem, it is possible to formulate a general approach to derive the macroscopic nonequilibrium response of the system by tracking numerically the microscopic equilibrium fluctuations of the dipoles $\vec{\mu}_i$. The Hamiltonian of the system can be written as:

$$H = \frac{1}{2} \sum_{i,j=1}^{N} U_{ij} - \sum_{i=1}^{N} \vec{\mu}_i \cdot \vec{B}_1, \tag{23}$$

considering the dipole-dipole coupling energy

$$U_{ij} = \frac{\mu_0}{4\pi r_{ij}^3} \left[\vec{\mu}_i \cdot \vec{\mu}_j - 3 \frac{\left(\vec{\mu}_i \cdot \vec{r}_{ij}\right)\left(\vec{\mu}_j \cdot \vec{r}_{ij}\right)}{r_{ij}^2} \right], \tag{24}$$

where $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$. The magnetization of the system is defined here as:

$$\vec{M} = \sum_i \vec{\mu}_i. \tag{25}$$

We consider the rotation of the magnetic dipoles with a position fixed in a two-dimensional square lattice along the (x, y) plane. The nearest neighbor distance d corresponds to a dipole surface density $\rho = 1/d^2$. Each dipole $\vec{\mu}_i = \mu \vec{e}_i$ undergoes a viscous torque $\vec{\Gamma}_{\text{vis},i} = -\zeta \vec{\omega}_{p,i}$, with $\vec{\omega}_{p,i}$ the instantaneous rotation vector of dipole $\vec{\mu}_i$ and $\zeta = 6\eta V$ the angular viscous friction coefficient. In addition to this damping, each dipole i is subjected to a magnetic torque $\vec{\mu}_i \times \vec{B}_i$ with

$$\vec{B}_i = \vec{B}_1 - \frac{\mu_0}{4\pi} \sum_{j \neq i} (\vec{\mu}_j \cdot \vec{\nabla}_i) \left(\frac{\vec{r}_{ij}}{r_{ij}^3} \right).$$

To account for the collisions with the liquid molecules, we have considered an angular Langevin equation with a stationary random torque $\vec{\Gamma}_{\text{rnd},i}$ acting on $\vec{\mu}_i$. The correlation time of this torque is consequently the collision time, which remains much smaller than the relaxation time of the angular velocity $\vec{\omega}_{p,i}$. $\Gamma_{\text{rnd}}(t)$ is thus delta-correlated in time: $\langle \Gamma_{\text{rnd}}(t)\Gamma_{\text{rnd}}(t+\tau) \rangle = 2\zeta k_B T \delta(\tau)$. The complete set of differential equations describing the dynamics of the magnetic dipoles damped in a liquid is

$$\begin{cases} I d\vec{\omega}_{p,i}/dt = \vec{\mu}_i \times \vec{B}_i - \zeta \vec{\omega}_{p,i} + \vec{\Gamma}_{\text{rnd},i}(t) \\ d\vec{\mu}_i/dt = \vec{\omega}_{p,i} \times \vec{\mu}_i \end{cases} \tag{26}$$

In principle, it would be possible to study the response of the system illuminated by an external alternating field $\vec{B}_l(\omega)$ along direction l by including explicitly B_l in Eq. 26. The dynamics of a magnetic dipole in an oscillating field has been studied with a similar approach (Engel et al. 2003). Practically, this procedure requires too many calculations to determine the whole spectrum of the magnetic susceptibility tensor $\alpha(\omega)$ defined as:

$$\vec{M}_k(\omega) = \alpha_{kl}(\omega)\vec{B}_l(\omega). \tag{27}$$

Thanks to the fluctuation-dissipation theorem (Callen et al. 1951; Landau and Lifshitz 1980) (FDT), the components of the imaginary part of the magnetic susceptibility tensor $\alpha'' = Im(\alpha(\omega))$ can be obtained from the time correlation function of $\vec{M}$ as:

$$\alpha_{kl}(\omega) = \frac{\langle M_k(0)M_l(0) \rangle}{Nk_B T} + \frac{i\omega}{Nk_B T} \int_0^\infty \langle M_k(t)M_l(0) \rangle e^{-i\omega t} dt, \tag{28}$$

with

$$\langle M_k(t)M_l(0)\rangle = \int\limits_{-\infty}^{+\infty} M_k(t+\tau)M_l(\tau)d\tau, \qquad (29)$$

$k, l \in \{x, y, z\}$ and N is the number of particles considered. This approach can be compared with some correlation functions of the magnetization trackable experimentally (Grimm et al. 2011).

4.2   Collective Behaviors in Large Clusters of MN Nanoparticles

Let us consider the case of noninteracting rigid dipoles. It corresponds to the limit where the inter-particle distance $d \rightarrow \infty$, so that there is no effect of the dipolar interactions and $\rho = 0$. The dipoles are carried by spherical particles of various diameters. Figure 16 depicts the time fluctuations of the components of the magnetization $\vec{M}(t)$ and the imaginary part of the susceptibility spectra $\alpha''(\omega)$ extracted from these fluctuations following Eq. (28). As expected, it can be observed that $\alpha''(\omega)$ fits with a Debye law according to the formula:

$$\alpha''(\omega, \rho = 0) = \alpha(\omega = 0, \rho = 0)\frac{\omega\tau}{1 + (\omega\tau)^2}. \qquad (30)$$

The lowest frequencies are attributed to the slowest characteristic times; thus, the longer one increases the simulation duration, the better one estimates the lowest frequency part of the spectra (in our case, $N_s = 1^8$ steps with $dt = 1\,\text{ms}$). At the optimal excitation angular frequency $\omega = \tau^{-1}$, the dissipation $\alpha''(\rho = 0)$ reaches its maximum value $\alpha''_{\text{max}}(\rho = 0) \equiv \alpha(\omega = \tau^{-1}, \rho = 0) = \alpha(\omega = 0, \rho = 0)/2$, where τ is the relaxation time of the system (Fig. 16b). For MNPs with hydrodynamic diameters ranging from 5 to 50 nm, the simulations give relaxation times that quantitatively agree with the expected Nernst-Einstein relation $\tau = \tau_B = \zeta/2k_BT = 3\eta V/k_BT$ (Fig. 16c). This result validates the statistical calculation in the well-known case of noninteracting rigid dipoles experiencing angular Brownian relaxation. Note that, as expected, the in-plane component of the susceptibility $\alpha''_{n/p}$ ($\vec{e}_n$ and $\vec{e}_p$ being any couple of orthogonal in-plane directions) and the cross-plane components of the susceptibility α''_z are identical in this case. We have next included the dipole-dipole interactions between MNPs. The strength of the interactions is tuned with the particle surface density ρ. Figure 17b displays the anisotropic dissipation spectra for the two components α''_n and α''_p in the plane and the perpendicular component α''_z, at the density of $1.5 \cdot 10^{-2}\,\text{nm}^{-2}$. The corresponding time fluctuations of these magnetization components are illustrated on the inset. We have identified on the same figure a range of concentration (gray) where the Knudsen number $K_n = \Lambda/l$, obtained from the water mean-free path

Fig. 16 Top: Imaginary part of the susceptibility from equilibrium simulations. Inset shows the time evolution of the momentum. A Debye fit is shown with the bold lines. Bottom: Validation with the Nernst-Einstein formula
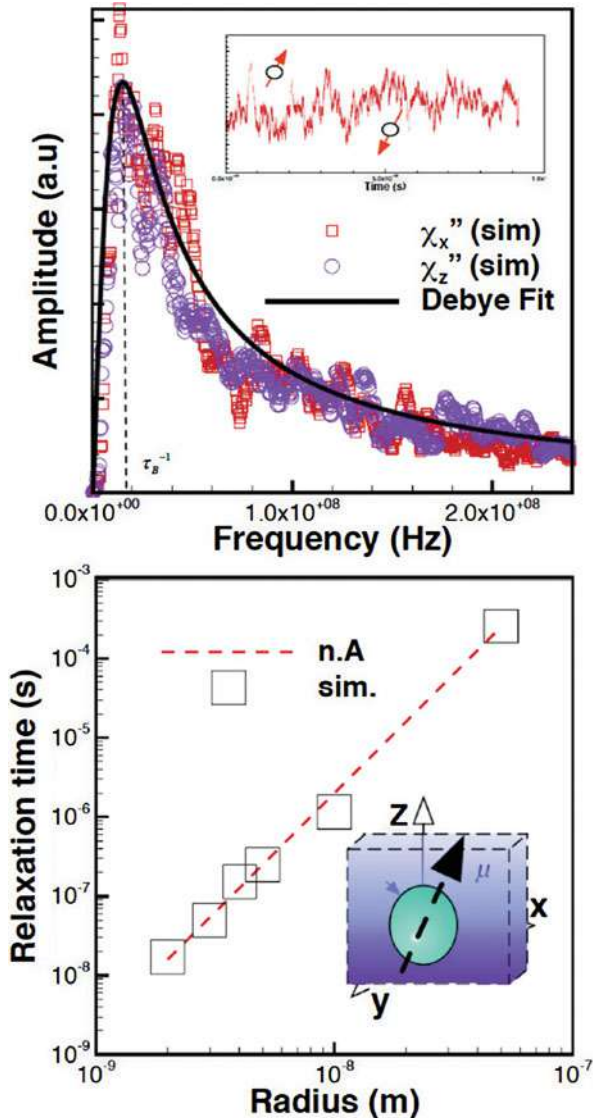
$(\Lambda = 2.5 \cdot 10^{-10} \, \text{m}$ in water) and the edge-to-edge inter-particle distance l, is not small enough to guaranty the absence of hydrodynamic effects. As shown on the same plot, this particular regime occurs in the restricted case where the particles are in contact $(\rho \sim 10^{-2} \, \text{nm}^{-2})$. As we will see later in this work, the damping by the liquid is so weak compared to the dipole-dipole interactions that the effect of the liquid can be completely neglected.

4.3   The Effect of the Dipole-Dipole Interaction

The dipole-dipole interactions (illustrated in Fig. 17) impact the amplitude of $\alpha''(\omega)$, which is dramatically decreased, and the maximum of dissipation is shifted toward higher frequencies. In the time domain, orientational fluctuations of interacting dipoles are both accelerated and limited in amplitude compared to noninteracting dipoles. This is consistent with what has already been reported independently (Landi 2014; Denisov and Trohidou 2001; Zhou and Bagshi 1992). To estimate the mass concentration of nanoparticle in g/l from which the effect of the interactions

Fig. 17 Top: Imaginary part of the susceptibility from equilibrium simulations. The in-plane (red) and cross-plane (blue) dissipation are plotted. Bottom: Evolution of the susceptibility spectra with the concentration. The simulation points have been fitted with continuous curves for the sake of clarity
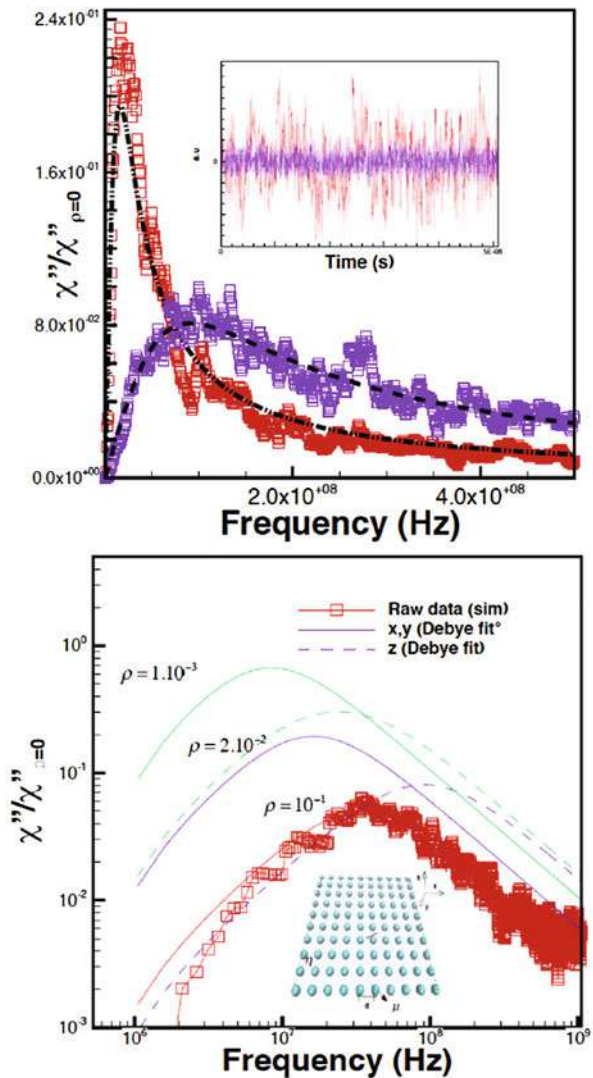
becomes observable, we consider the mass density of iron oxide ($5 \cdot 10^3$ kg/m^3). Dipolar effects occur as soon as the inter-particle distances become shorter than roughly three times the particle diameter (this correspondence between the dipolar energy cutoff and the particle diameter remains a good approximation even at other sizes, when accounting for a volume-dependent dipole moment). Considering the case of an isotropic particle distribution, the mass concentration is of the order of 250 g/L. Figure 17b presents the amplitude at the maximum of dissipation $\alpha''_{max}(\rho)$ as a function of the density, for each component. These amplitudes are normalized

Fig. 18 Power law with the inter-particle distance. (Bottom) The relaxation frequency as function of the concentration. (TOP): The amplitude of absorption as function of the concentration
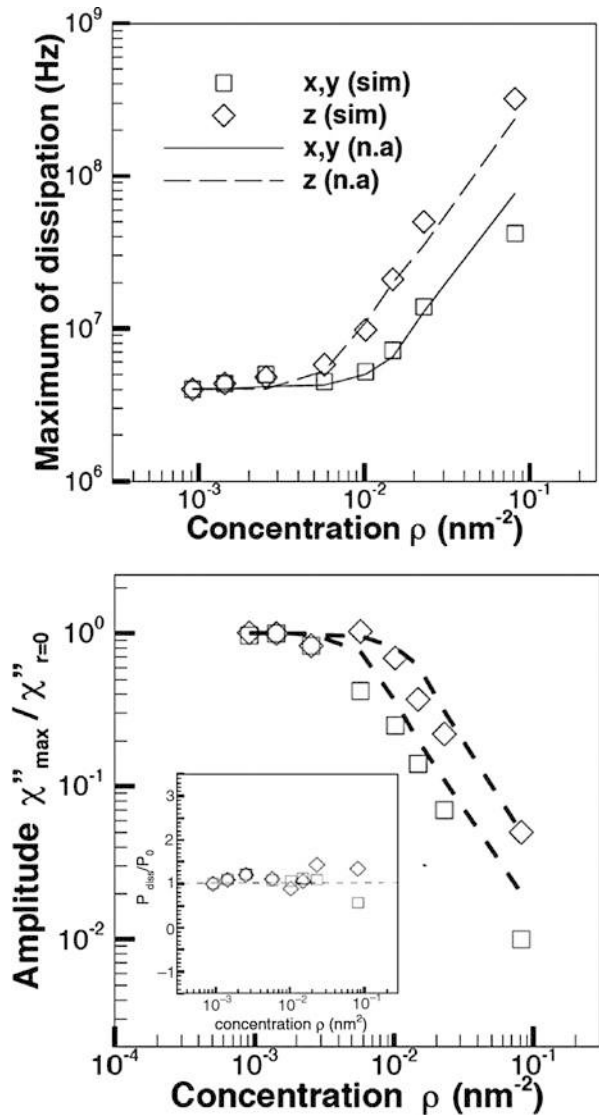
by the corresponding values in the noninteracting case. Similarly, Fig. 17c displays the frequency of maximum dissipation $1/\tau$ as a function of the MNP density. Importantly, above a density threshold of $\rho = 4 \cdot 10^{-4} \, \mathrm{nm}^{-2}$, the optimal frequency increases with density as $\rho^{3/2}$, similarly to the coupling energy U which scales as $d^{-3} = \rho^{3/2}$. According to the Debye structure of the dissipation spectra, we deduce that the relaxation rate of rigid interacting dipoles includes an additional term which accounts for the interactions (Fig. 18):

$$\tau^{-1} = \tau_B^{-1} + \tau_U^{-1}(\rho), \tag{31}$$

with $\tau_U(\rho) = \zeta / U(\rho)$.

4.3.1 Concluding Remarks

Considering the equilibrium orientational fluctuations of magnetic dipoles, we have demonstrated that one can quantify the Brownian dissipation without approximations. Our work suggests that the interactions between magnetic dipoles have a dramatic impact on the fundamental mechanisms of dissipation: Increasing the dipole pair energy enhances the Brownian relaxation rate and diminishes the magnitude of absorption. Practically, these mechanisms are essential in the context of hyperthermia research. A perspective of this work would be to carry this analysis in situ, in order to study the heat diffusion in more complex environments, where the intracellular confinement has affected the particle concentration.

References

Allen MP, Tildesley DJ (1987) Computer simulation of liquids. Clarendon, Oxford

Bahia M, Numes J, Altbir D, Vargas P, Knobel M, Bahia M, Numes J, Altbir D, Vargas P, Knobel M (2004) J Mag Mag Mat 281:372–377

Becker B, Schelling PK, Phillpot SR (2006) Interfacial phonon scattering in semiconductor nanowires by molecular-dynamics simulation. J Appl Phys 99:123715

Benz R, Fröhlich O, Läuger P, Montal M, (1975) Electrical capacity of black lipid films and of lipid bilayers made from monolayers. Biochimica et Biophysica Acta, 394, 323–334 .

Berestovsky G, Gyulkhandanyan M, Ivkov V, Razhin V (1978) Voltage-induced reflectivity relaxation of bilayer lipid membranes: On changes of bilayer thickness. J Membrane Biol 43:107–126

Berkov DV, Gorn NL (2001) Susceptibility of the disordered system of fine magnetic particles: a Langevin-dynamics study. J Phys Condens Matter 13:9369–9381

Borca-Tasciuc et al. (2000) Thermal conductivity of symmetrically strained Si/Ge superlattices. Superlattices and Microstructure 28:199

Branquinho LC, Carrio MS, Costa AS, Zufelato N, Sousa MH, Miotto R, Ivkov R, Bakuzis AF (2013) Effect of magnetic dipolar interactions on nanoparticle heating efficiency: implications for cancer hyperthermia. Sci Rep 3:2887

Brinis D, Laggoun A, Ledue D, Patte R (2014) Effects of dimensionality and spatial distribution on the magnetic relaxation of interacting ferromagnetic nanoclusters: A Monte Carlo study. J. App. Phys. 115:173906

Cahill DG, Ford WK, Goodson KE, Mahan GD, Majumdar A, Maris HJ, Merlin R, Phillpot SR (2003) J Appl Phys 93:793–818

Callen HB, Welton TA (1951) Irreversibility and generalized noise. Phys Rev 83:34

Carey VP, Chen G, Grigoropoulos C, Kaviany M, Majumdar A (2008) Nanoscale and microscale thermophysical engineering 12:1–60

Chalopin Y, Dammak H, Laroche M, Hayoun M, Greffet JJ (2011) Radiative heat transfer from a black body to dielectric nanoparticles. Phys Rev B 84:224301

Chalopin Y, Dammak H, Hayoun M, Besbes M, Greffet JJ (2012a) Size-dependent infrared properties of MgO nanoparticles with evidence of screening. App Phys Lett 100:241904

Chalopin Y, Esfarjani K, Henry H, Volz S, Chen G (2012b) Thermal interface conductance in Si/Ge superlattices by equilibrium molecular dynamics. Phys Rev B 85:195302

Chen TS, Alldredge GP, de Wette FW, Allen RE (1972) Surface Modes of Vibration in the Rigid-Ion Model of NaCl. Phys Rev B 6:627

Coffey WT, Kalmykov YP (2012) Thermal fluctuations of magnetic nanoparticles: fifty years after Brown. J App Phys 112:121301

David C, Garcia de Abajo FJ (2011) Spatial Nonlocality in the Optical Response of Metal Nanoparticles. J Phys Chem C 115:19470–19475

Dejardin PM (2011) Magnetic relaxation of a system of superparamagnetic particles weakly coupled by dipole-dipole interactions. J App Phys 110:113921

Denisov SI, Trohidou KN (2001) Fluctuation theory of magnetic relaxation for two-dimensional ensembles of dipolar interacting nanoparticles. Phys Rev B 64:184433

Denisov SI, Lyutyy TV, Trohidou KN (2003) Magnetic relaxation in finite two-dimensional nanoparticle ensembles. Phys Rev B 67:14411

Di Corato R, Espinosa A, Lartigue L, Tharaud M, Chat S, Pellegrino T, Menager C, Gazeau F, Wilhelm C (2014) Magnetic hyperthermia efficiency in the cellular environment for different nanoparticle designs. Biomaterials 35:6400–6411

Engel A, Muller HW, Reimann P, Jung A (2003) Ferrofluids as thermal ratchets. Phys Rev Lett 91:060602

Farrell D, Ding Y, Majetich SA, Sanchez-Hanke C, Kao CC (2004) Structural ordering effects in Fe nanoparticle two- and three-dimensional arrays. J. App. Phys. 95(11):6636–6638

Felderhof BU (2002) Rotational Brownian motion of a pair of linear molecules or dipoles with anisotropic interaction. J Chem Phys 117:3583

Fray SJ, Johnson FA, Quarrington JE, Williams N (1961) J Appl Phys 32:2102

Gao Y, Marconnet A, Panzer M, LeBlanc S, Dogbe S, Ezzahri Y, Shakouri A, Goodson K (2010) J Elec Mat 39:1456–1462

Gaudry JB, Capes L, Langot P, Marcen S, Kollmannsberger M, Lavastre O, Freysz E, Letard JF, Kahn O (2000) Chem Phys Lett 324:321

Greffet J-J, Carminati R, Joulain K, Mulet JP, Mainguy S, Chen Y (2002) Nature 416:61

Grimm M, Jeney S, Franosch T (2011) Brownian motion in a Maxwell fluid. Soft Matter 7:2076

Jimenez RP, Zayas FG, M'Peko JC, Eiras JA (2006) A numerical study of relaxation in a two dimensional dipolar lattice. J Appl Phys 99:64102

Jones WE, Fuchs R (1971) Surface modes of vibration and optical properties of an ionic - crystal slab. Phys Rev B 4:3581

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935

Karjiban RA, Shaari NS, Gunasakaran UV, Basri M (2013) A Coarse-Grained Molecular Dynamics Study of DLPC, DMPC, DPPC, and DSPC Mixtures in Aqueous Solution. J Chem 2013, 931051

Kubo R (1966) The fluctuation-dissipation theorem. Rep Prog Phys 29:255

Kubo R, Toda M, Hashisume N (1985) Statistical physics II. Springer, Berlin

Kupiainen M, Falck E, Ollila S, Niemel P, Gurtovenko A et al. (2005) Free volume properties of sphingomyelin, DMPC, DPPC, and PLPC bilayers. J Comput Theor Nanosci 2:401–413

Landau LD, Lifshitz EM (1980) Statistical Physics. 3rd Edition, Elsevier, Amsterdam

Landauer R (1970) Electrical resistance of disordered one-dimensional lattices Philos Mag 21:863

Landi GT (2014) Role of dipolar interaction in magnetic hyperthermia. Phys Rev B 89:14403

Landry ES, McGaughey AJH (2010) Effect of film thickness on the thermal resistance of confined semiconductor thin films. J App Phys 107:013521

Lee C, Bain CD (2005) Raman spectra of planar supported lipid bilayers. Biochim Biophys Acta 1711:59–71

Lee SM, Cahill DG, Venkatasubramanian R (1997) Thermal conductivity of Si-Ge superlattices. Appl Phys Lett 70(22):2957–9

Lucas AA (1968) Phonon modes of an ionic crystal slab. J Chem Phys 48:3156

MacIsaac AB, Whitehead JP, De'Bell K, Poole PH (1996) Phase diagram for a magnetic thin film with dipolar interactions and magnetic surface anisotropy. Phys. Rev. Lett. 77 :739–742

Matsui M (1989) Molecular dynamics study of the structural and thermodynamic properties of MgO crystal with quantum correction. J Chem Phys 91:489

McMahon JM, Stephen KG, Schatz GC (2009) Nonlocal optical response of metal nanostructures with arbitrary shape. Phys Rev Lett 103:097403.

Mingo N, Yang L (2003) Phonon transport in nanowires coated with an amorphous material: An atomistic Green's function approach Phys Rev B 68:245406

Montal M, Mueller P (1972) Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. Proc Nat Acad Sci 69:3561–3566

Morup S, Tronc E (1994) Superparamagnetic relaxation of weakly interacting particles. Phys Rev Lett 72:3278–3281

Nie S, Emory SR (1997) Probing single molecules and single nanoparticles by surface-enhanced raman scattering. Science 275:1102

Ong ZY, Pop E (2010) Molecular dynamics simulation of thermal boundary conductance between carbon nanotubes and SiO_2. Phys Rev B 81:155408

Palik ED, Bennett JM (1991) Handbook of optical constants of solids, vol II. Academic Press, New York

Panzer MP, Goodson KE (2008) Thermal resistance between low-dimensional nanostructures and semi-infinite media. J Appl Phys 103:094301-01

Pottier N (I999) Nonequilibrium statistical physics. Linear Irreversible Processes, Oxford Graduate Texts

Ritov S.M (1959) Theory of Electric Fluctuations and Thermal Radiation (Air Force Cambridge Research Center, Bedford, Mass.,), AFCRC-TR-59-162

Ruta S, Chantrell R, Hovorka O (2015) Unified model of hyperthermia via hysteresis heating in systems of interacting magnetic nanoparticles. Sci Rep 5:9090

Scarel G, Hyde GK, Hojo D, Parsons GN (2008) Berreman effect in infrared absorption spectroscopy of ionic oxide coatings formed by atomic layer deposition on three-dimensional structures. J Appl Phys 104:094314

Schelling PK, Phillpot SR, Keblinski P (2002) Phonon wave-packet dynamics at semiconductor interfaces by molecular-dynamics simulation. Appl Phys Lett 80:2484

Shafiei F, Wu C, Wu Y, Khanikaev AB, Putzke P, Singh A, Li X, Shvets G (2013) Nat Photon 7:367

Stariolo DA, Billoni OV (2008) Dipolar interactions and thermal stability of two-dimensional nanoparticle arrays. Phys D Appl Phys 41:20

Stillinger H, Weber TA (1985) Computer simulation of local order in condensed phases of silicon Phys Rev B 31:5262

Sugano R, Matsushita K, Kuroda A, Tomita Y, Takayama H. Magnetic properties of 2-Dimensional dipolar squares: boundary geometry dependence. J Phys Soc Jpn 76:44705 (2007)

Tan RP, Carrey J, Respaud M (2014) Magnetic hyperthermia properties of nanoparticles inside lysosomes using kinetic Monte Carlo simulations: influence of key parameters and dipolar interactions, and evidence for strong spatial variation of heating power. Phys Rev B 90:214421

Tong SY, Maradudin AA (1969) Normal modes of a semi-infinite ionic crystal. Phys Rev 181:1318

Toyama S, Nakamura A, Toda F (1991) Measurement of voltage dependence of capacitance of bilayer lipid membrane under the existence of ion carriers. Biophys J 59:939–944

Vermeer SL, de Groot BL, Rat V, Milon A, Czaplicki J (2007) Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with 2H NMR experiments. Eur Biophys J 36:919–931

Voisin C, Del Fatti N, Christofilos D, Vallée F (2001) Ultrafast electron dynamics and optical nonlinearities in metal nanoparticles. J Phys Chem B 105:2264

Zhang L, Keblinski P, Wang JS, Li B (2011) Interfacial thermal transport in atomic junctions. Phys Rev B 83:064303

Zhang W, Mingo N, Fisher TS (2007a) Simulation of phonon transport across a non-polar nanowire junction using an atomistic Green's function method. Phys Rev B 76:195429

Zhang W, Mingo N, Fisher TS (2007b) The atomistic green's function method: An efficient simulation approach for nanoscale phonon transport. Numer Heat Transfer 51:333

Zhang W, Fisher TS, Mingo N (2007c) Simulation of interfacial phonon transport in Si-Ge heterostructures using an atomistic green's function method J Heat Transfer 129:483

Zhou HX, Bagshi B (1992) Solvation dynamics in a Brownian dipole lattice: a comparison between theory and computer simulation. J Chem Phys 97:3610

Thermal Conductivity of Nanostructured Semiconductor Alloys

39

Meenakshi Upadhyaya and Zlatan Aksamija

Contents

Abstract

This chapter provides a comprehensive description of phonon transport in nanostructures of silicon-based Group-IV alloys. We employ the phonon Boltzmann transport equation (pBTE) formalism with full phonon dispersion and a partially diffuse momentum-dependent specularity model for boundary roughness scattering. The results show thermal conductivity in Si-Ge nanostructures including thin films, superlattices (SLs), nanowires (NWs), and nanocomposites (NCs) to be well below their bulk counterparts, reaching almost to the amorphous limit. Further minimization of lattice thermal conductivity is shown in binary (SiSn and GeSn) and ternary (SiGeSn) alloys and their thin films. Si-Sn alloys have the lowest conductivity (3 W/mK) of all the bulk alloys, which reduces further to around 1 W/mK in 20-nm-thick Si-Sn films. Thermal transport in nanostructures is tunable by extrinsic boundary effects such as sample size in thin films, period

M. Upadhyaya · Z. Aksamija (✉)
Department of Electrical and Computer Engineering, University of Massachusetts,
Amherst, MA, USA
e-mail: zlatana@engin.umass.edu

thickness in SLs, length/diameter in NWs, and grain size in NCs. Additionally, boundary/interface properties, such as roughness, orientation, and composition, also play a significant role in thermal transport and offer additional degrees of freedom to control the thermal conductivity in nanostructured semiconductor alloys. Thermal conductivity can be minimized in SLs by growing short-period $Si/Si_{1-x}Ge_x$ SLs with the Si-Ge alloy layer thicker than the Si one. We describe a Monte Carlo method of sampling the phonon mean free paths (MFPs), capable of capturing both partially diffuse boundaries and ballistic effects in the calculation of thermal conductivity of Si-Ge NWs. The computed phonon flights are comprised of a mix of long free flights over several μm interrupted by bursts of short flights, resulting in a heavy-tailed distribution of flight lengths, indicating superdiffusive phonon transport, which results in $L^{1/3}$ scaling across a wide range of NW lengths up to nearly $10\,\mu$m. Lastly, our pBTE model for nanocomposites, based on Voronoi tessellation to capture the grain structure in NCs, is described. The size scaling of thermal conductivity observed in NWs persists in NCs as well and is found to be insensitive to the variance in grain sizes.

1   Introduction

This chapter discusses thermal transport in silicon-based semiconductor alloys, which has drawn considerable interest in recent years. Much of the work employs a combination of alloying (Bera et al. 2010) and nanostructuring (Hicks and Dresselhaus 1993a; Dresselhaus et al. 2007) techniques to reduce the lattice thermal conductivity and improve the thermoelectric (TE) conversion efficiency of semiconductors (Joshi et al. 2008; Minnich et al. 2009; Zhu et al. 2009). TE energy conversion efficiency, given by the ratio between heat absorbed and the energy produced, is primarily dictated by a dimensionless figure of merit $ZT = S^2\sigma T/\kappa$, where S is the Seebeck coefficient, T is the absolute temperature, σ is the electrical conductivity, and κ is the thermal conductivity, composed of both lattice and electronic contributions (DiSalvo 1999). Most moderately doped semiconductors have high electrical conductivity σ and a moderate Seebeck coefficient S, with σ increasing and S decreasing with doping. Even when doping levels are optimized for a maximal power factor ($S^2\sigma$), ZT is still limited by the high thermal conductivity. The lattice contribution to thermal conductivity (κ_l) far outweighs its electronic counterpart (κ_e), and is the only property which can be tuned somewhat independently, without adversely affecting the power factor.

Silicon-based TEs are attractive due to the abundant availability of Si, relatively low cost, and previously established processing technology. In addition, they integrate easily with existing microelectronics microelectronics, opening up the possibility for on-chip cooling and improved thermal management in ultra-scaled CMOS circuits (Chowdhury et al. 2009; Pop 2010; Ryu et al. 2010). However, the high lattice thermal conductivity of bulk silicon at room temperature (146 W/m · K) (Glassbrenner and Slack 1964; Maycock 1967) limits the ZT of silicon to

approximately 0.05 (Liao et al. 1999), almost two orders of magnitude below the value of $ZT = 3$, which is considered desirable for solid-state TE devices to replace the conventional methods of cooling and power generation (Majumdar 2004). However, a more modest ZT of $1.5-2$ may suffice for kW-range applications such as vehicle heat recovery, car cooling/heating, and home cogeneration (Vining 2009).

Group-IV alloys of Si-Ge with varying composition have long been employed in TE energy conversion applications, such as NASA radioisotope TE generators, and have $ZT > 1$ around 900 K (Snyder and Toberer 2008). The improvement in ZT is primarily from the order-of-magnitude reduction in the lattice thermal conductivity κ_{ph} due to the mass-difference phonon scattering induced by alloying. However, the lattice thermal conductivity in $Si_{1-x}Ge_x$ alloys reaches a broad plateau between $0.2 < x < 0.8$ (Maycock 1967), which limits the lowest conductivity achievable through alloying. The mass-difference phonon scattering that dominates lattice transport in the alloy depends on the difference in atomic mass between the constituent materials (in this case Si and Ge) and the average mass in the alloy (given by $M_{Si}(1 - x) + M_{Ge}x$). Initially, increasing the Ge concentration moves the average further from the Si mass, resulting in an increase of the scattering rate (discussed in detail in Sect. 2). However, increasing the Ge concentration past a certain point (around x = 0.2) does move the average further from the Si mass but at the expense of moving it closer to the increasingly large fraction of Ge atoms in the alloy, resulting in a broad plateau in thermal conductivity vs. composition. Hence, additional methods of reducing the lattice thermal conductivity are required. Within this plateau, nanostructuring results in stronger decrease of thermal conductivity (three to five times lower than the bulk) than variation in Ge composition, while outside the plateau, compositional variation dominates.

The low-dimensional approach to reduce thermal conductivity was first proposed in seminal papers by Hicks and Dresselhaus (1993a, b), leading the way for numerous studies on $Si_{1-x}Ge_x$ alloy nanostructures, such as SiGe thin films (Cheaito et al. 2012), $Si/Si_{1-x}Ge_x$ superlattices (Lee et al. 1997; Lee and Venkatasubramanian 2008; Liu et al. 2001; Aksamija and Knezevic 2013; Upadhyaya et al. 2015), nanowires (Khitun et al. 1999; Wang and Mingo 2010; Lee et al. 2012; Hsiao et al. 2013; Upadhyaya and Aksamija 2016), and nanocomposites (Joshi et al. 2008; Wang et al. 2008; Aksamija 2014). Lattice thermal conductivity in nanostructures is dominated by scattering from the rough boundaries or interfaces between layers, even at room temperature. Therefore, in addition to composition, boundary/interface properties, such as roughness and orientation, are expected of freedom to control the thermal conductivity in semiconductor nanostructured alloys.

Our methodology is described in Sect. 2 and is based on solving the phonon Boltzmann transport equation (pBTE) in the relaxation time approximation, combining the full phonon dispersion calculated from Weber's adiabatic bond charge model (Weber 1974). All the relevant intrinsic scattering mechanisms spanning 3-phonon anharmonic umklapp and normal, isotope, impurity, and alloy mass difference are included. In order to accurately treat extrinsic phonon scattering from rough boundaries/interfaces, we employ a momentum-dependent specularity

parameter that depends on phonon wavelength and rms roughness amplitude. Several key results are discussed, beginning with Si-Ge alloy thin films in Sect. 3, and superlattices in Sect. 4, where diffuse interface scattering results in highly anisotropic thermal transport, with conductivity below that of a random alloy. Thermal conductivity can be further minimized by growing short-period $Si/Si_{1-x}Ge_x$ SLs with the SiGe layer thicker than the Si one. Calculation of the lattice thermal conductivity in Si-Ge nanowires (NWs) based on solving the pBTE by Monte Carlo method of sampling the phonon MFPs is discussed in Sect. 5. In alloy NWs, unlike their non-alloy counterparts, phonon transport is essentially superdiffusive, and thermal conductivity scales with the length of the NW as $\kappa \propto L^{1/3}$ over a broad range of wire lengths $10\,\text{nm} < L < 10\,\mu\text{m}$ regardless of diameter and roughness. This is a consequence of Levy-like phonon transport where series of short random flights are interrupted by several-μm-long leaps, leading to a heavy-tailed flight distribution.

Next, Sect. 6 on alloy NCs describes our Voronoi tessellation technique to compute the grain size distribution and interconnections. The size-dependent thermal conductivity in each grain is computed, and the resulting grain distribution is iterated until convergence to find the effective thermal conductivity in micron-sized samples, composed of tens of thousands of nanoscale grains. The size scaling observed in NWs persists even in nanocomposites, where the conductivity scales with the mean diameter of the grains as $\kappa \propto D^{1/3}$ in SiGe and $\kappa \propto D^{2/3}$ in Si NWs. However, the conductivity is found to be insensitive to the variance in grain sizes. Much work remains to be done in co-optimizing the complex morphology of NCs for simultaneous decrease in thermal and gain in electrical conductivity. Lastly, Sect. 7 discusses material advancement beyond SiGe, to Group-IV alloys containing tin, including binary Si-Sn and Ge-Sn and Si-Ge-Sn ternary alloys. Si-Sn was found to possess the lowest thermal conductivity among all the cases, and its nanostructures have conductivity nearing the amorphous limit. We conclude in Sect. 8 that Group-IV alloy nanostructures, in particular SLs, NWs, and NCs containing a mix of Si-Ge and Si-Sn, hold tremendous promise for future high-efficiency TE energy conversion applications.

2   Phonon Transport Model

Most theories that simultaneously capture both electron and phonon transport in crystalline solids employ the Boltzmann transport equation (BTE). The BTE theory can explain a broad range of relevant effects in bulk or homogeneous materials, including the dependence of the electrical conductivity, the thermal conductivity, and the Seebeck coefficient on temperature, impurity content, isotope scattering, and quantum confinement (Cahill et al. 2003). We focus herein on the lattice thermal conductivity via the transport of phonons, where the BTE approach can be complementary to other methods, such as molecular dynamics (MD). By solving the time-independent phonon Boltzmann transport equation (pBTE) in its relaxation-time-approximation (RTA) form, the steady-state distribution function

can be obtained, which is then used to determine lattice heat conductivity. In steady state, the time-independent pBTE is given as (Carruthers 1961)

$$\upsilon_{b,\mathbf{q}} \cdot \nabla N_{b,\mathbf{q}}(x, y, z) = -\frac{N_{b,\mathbf{q}}(x, y, z) - N_{b,\mathbf{q}}^0(T)}{\tau_{b}^{\text{int}}(\mathbf{q})}. \tag{1}$$

As indicated in the pBTE, the steady-state phonon distribution $N_{b,q}(x, y, z)$ is a function of the phonon branch b, wave vector $\mathbf{q}$, and position in 3D space (x, y, z). Here τ_q^{int} is the relaxation time due to all the intrinsic scattering mechanisms occurring in the interior of the material, including both resistive umklapp and non-resistive normal anharmonic phonon-phonon, isotope, impurity, and alloy mass-difference interactions. τ_q^{int} can be obtained using the standard single-mode relaxation time approximation from elasticity theory (Morelli et al. 2002) or from first principles, which typically leads to a quadratic (Ward and Broido 2010) or cubic (Esfarjani et al. 2011) frequency dependence of the anharmonic scattering rates in the long-wavelength limit. The equations hold for each branch, and interbranch scattering is included in $\tau_{b}^{\text{int}}(\mathbf{q})$. In the calculation of relaxation time $\tau_{b}^{\text{int}}(\mathbf{q})$ for a phonon in branch b and with wave vector $\mathbf{q}$, we consider normal (N) and umklapp (U) three-phonon scattering, impurity (I, if present through doping or defects), and mass-disorder scattering (Mass) with branch- and momentum-dependent relaxation times $\tau_{b,N}(\mathbf{q})$, $\tau_{b,U}(\mathbf{q})$, $\tau_{b,I}(\mathbf{q})$, and $\tau_{b,\text{Mass}}(\mathbf{q})$, respectively. The total intrinsic relaxation time is given by

$$\frac{1}{\tau_{b}^{\text{int}}(\mathbf{q})} = \frac{1}{\tau_{b,N}(q)} + \frac{1}{\tau_{b,U}(q)} + \frac{1}{\tau_{b,I}(q)} + \frac{1}{\tau_{b,\text{Mass}}(q)}. \tag{2}$$

When two materials are combined into an alloy, in this case $Si_{1-x}Ge_x$, variation in the local atomic mass leads to perturbations of lattice waves, which results in strong mass-difference scattering of phonons. In alloys, mass-disorder scattering will have three components: alloying, isotopic mass variation, and the small local strain field induced by variations in the atomic species (Si or Ge). The scattering strength will be proportional to the total fraction of mass-disordered constituents: (Carruthers 1961)

$$\Gamma_{\text{Mass}}(x) = \Gamma_{\text{Alloy}}(x) + \Gamma_{\text{Iso}} + \Gamma_{\text{Strain}}(x). \tag{3}$$

The alloy mass-difference constant is defined as (Abeles et al. 1962)

$$\Gamma_{\text{Alloy}} = \sum_i f_i (1 - M_i/\overline{M})^2, \tag{4}$$

where f_i is the proportion of material i with mass M_i while the average mass is $\overline{M} = \sum_i f_i M_i$ (Morelli et al. 2002). The strength of the alloy mass-difference scattering has a quadratic dependence on the ratio between the masses of constituent

atoms (M_i) and the average alloy mass ($\overline{M}$) in Eq. 4. In the present case of $Si_{1-x}Ge_x$ alloys, the mass disorder depends on the germanium fraction:

$$\Gamma_{Alloy}(x) = x(1-x)\frac{(M_{Ge} - M_{Si})^2}{(xM_{Ge} + (1-x)M_{Si})^2}. \tag{5}$$

The primary driver of thermal conductivity reduction from pure Si or Ge to the SiGe alloy is the quadratic dependence on the germanium concentration in alloy mass-difference $\Gamma_{Mass}(x)$. The energy dependence of the alloy scattering rate follows a Rayleigh-like ($\tau^{-1} \propto \omega^4$) trend and is calculated from (Tamura 1983; Maris 1990):

$$\frac{1}{\tau_{Mass}(\omega)} = \frac{\pi}{6}V_0\Gamma_{Mass}\omega^2 D(\omega), \tag{6}$$

where V_0 is the volume per atom and $D(\omega)$ is the vibrational density of states per unit volume (Garg et al. 2011). The total energy-dependent vibrational density of states is given by a sum over all phonon branches b:

$$D(\omega) = \sum_b \int \frac{d\mathbf{q}}{(2\pi)^3}\delta\left[\omega - \omega_b(\mathbf{q})\right]. \tag{7}$$

The volume integral of the energy-conserving delta function over the whole first Brillouin zone (BZ) is calculated from the full phonon dispersion using the extrapolation method of Gilat and Raubenheimer (1966). In this method, the BZ is discretized, and the area of the intersection between the constant-energy surface described by the delta function $\delta\left[\omega - \omega_b(\mathbf{q})\right]$ and each discretization cube is calculated. Then the areas are divided by the gradient of the dispersion and summed (Gilat and Kam 1969; Gilat 1972).

The contribution due to isotopic variation in each of the constituent materials can be obtained by combining the isotope constants for each pure material as

$$\Gamma_{Iso}(x) = \frac{(1-x)\Gamma_{Si}M_{Si}^2 + x\Gamma_{Ge}M_{Ge}^2}{(xM_{Ge} + (1-x)M_{Si})^2}, \tag{8}$$

where the pure silicon and germanium isotope scattering constants Γ_{Si} and Γ_{Ge} were taken from Morelli et al. (2002). An additional component to alloy scattering arising from the strain field due to the difference in lattice constants of pure Si and Ge and their alloys has been proposed. The contribution due to strain is then given by (Abeles et al. 1962)

$$\Gamma_{Strain}(x) = \varepsilon x(1-x)\frac{(a_{Ge} - a_{Si})^2}{a_{SiGe}^2(x)}, \tag{9}$$

where $a_{\mathrm{SiGe}}(x)$ is the composition-dependent alloy lattice constant, taken in the virtual crystal approximation, including bowing (Rieger and Vogl 1993). The empirical strain parameter is taken to be $\varepsilon = 39$ (Abeles 1963). It is estimated on the basis of the impurity model (Klemens 1960) and is in good agreement with experimental results for SiGe alloys. For most values of germanium concentration x, the strain contribution $\Gamma_{\mathrm{Strain}}(x)$ is found to be much smaller than the mass-difference component.

The resistive anharmonic phonon-phonon scattering rate was calculated in the standard approximation for dielectric crystals (Morelli et al. 2002):

$$\tau_{\mathrm{b,U}}^{-1}(\mathbf{q}) = \frac{\hbar \gamma_{\mathrm{b}}^2}{\overline{M} \Theta_{\mathrm{b}} \bar{\upsilon}_{\mathrm{b}}^2} \omega_{\mathrm{b}}^2(\mathbf{q}) T e^{-\Theta_{\mathrm{b}}/3T}, \tag{10}$$

where the speed of sound $\bar{\upsilon}_{\mathrm{b}}$ of each branch b is determined from the average slope of its dispersion curve near the Γ point and $\overline{M}$ is the average atomic mass. The Grüneisen parameter γ_{b} was obtained for each branch from the phonon dispersion and has the value of 1.1 for the longitudinal acoustic branch and -0.6 for the two transverse acoustic branches. The expression in (10) contains the exponential term $e^{-\Theta_{\mathrm{b}}/3T}$ in the temperature dependence, which controls the onset of resistive umklapp scattering for each phonon branch through the branch-specific Debye temperatures Θ_{λ}, which were obtained from (Slack 1979):

$$\Theta_{\mathrm{b}}^2 = \frac{5\hbar^2}{3k_B^2} \frac{\int \omega^2 D_{\mathrm{b}}(\omega) d\omega}{\int D_{\mathrm{b}}(\omega) d\omega}, \tag{11}$$

where the branch-wise vibrational density of states (vDOS) function $D_{\mathrm{b}}(\omega) = \sum_{\mathbf{q}} \delta \left[\omega - \omega_{\mathrm{b}}(\mathbf{q}) \right]$ was calculated for each phonon branch b from the full dispersion, as described previously. This way, the temperature dependence of the contribution from each phonon branch to the total thermal conductivity is accurately represented.

In the bulk case, the crystal is assumed infinite and uniform. The distribution is only a function of temperature, and in the absence of boundaries and interfaces, the solution of the pBTE equation is simply given by

$$n_{\mathbf{q}} = \tau_{\mathrm{int.}}(\mathbf{q}) \upsilon_{\mathbf{q}} \cdot \nabla_{\mathbf{r}} T \frac{\partial N_{\mathbf{q}}^0(T)}{\partial T}. \tag{12}$$

In contrast, boundaries and interfaces play an important role in the solution of pBTE in nanostructures. Hence, in nanostructures, an extrinsic relaxation rate $\tau_{\mathrm{b,B}}^{-1}(\mathbf{q})$ due to boundary-roughness scattering is added. Each time a phonon reaches the boundary, the probability of it not being scattered by the roughness is captured through the momentum-dependent specularity parameter $0 < p(\mathbf{q}) < 1$ given by

$$p(\mathbf{q}) = \exp\left(-4\Delta^2 q^2 \cos^2 \Theta\right), \tag{13}$$

with Δ being the rms boundary roughness (typically $0.1 < \Delta < 1$ nm, depending on sample quality and processing) and Θ being the angle between the direction of propagation of the phonon wave and the boundary normal. The specularity parameter allows us to capture both the magnitude and the angle dependence of the scattering and distinguish between the contribution to the heat flux from phonons traveling into the boundary (small Θ, hence smaller $p(\mathbf{q})$ and more scattering) and phonons traveling parallel to the boundary (large Θ and larger $p(\mathbf{q})$, leading to less scattering). The specularity parameter is used in solving the pBTE as a boundary condition, with $[1 - p(q)]$ giving the fraction of the incoming phonons which are scattered randomly.

As boundary scattering is a momentum-randomizing elastic process, the scattered phonons will carry zero heat flux, so they can be represented by the equilibrium Bose-Einstein distribution, leading to a boundary condition of the form:

$$N_{\mathrm{b}}(\mathbf{q})^{+} = p(\mathbf{q})N_{\mathrm{b}}(\mathbf{q})^{-} + [1 - p(\mathbf{q})] N_{\mathrm{b},T}^{0}(\mathbf{q}). \tag{14}$$

with $+$ and $-$ representing the solution before reaching and after leaving the boundary, respectively, and $N_0(\mathbf{q})$ is the equilibrium Bose-Einstein phonon distribution of phonon mode $\mathbf{q}$ in branch b. The boundary scattering rate for a film of thickness H is then obtained as (Aksamija and Knezevic 2013)

$$\tau_{\mathrm{b,B}}^{-1}(\mathbf{q}) = \frac{\upsilon_{\mathrm{b},\perp}(\mathbf{q})}{H} \frac{F_p(\mathbf{q}, H)}{1 - \frac{\tau_{\mathrm{b}}^{\mathrm{int}}(\mathbf{q})\upsilon_{\mathrm{b},\perp}(\mathbf{q})}{H} F_p(\mathbf{q}, H)}, \tag{15}$$

where a mode-dependent scaling factor $F_p(\mathbf{q}, H)$ is given by

$$F_p(\mathbf{q}, H) = \frac{[1 - p(\mathbf{q})]\left\{1 - \exp\left[-H/\tau_{\mathrm{b}}^{\mathrm{int}}(\mathbf{q})\upsilon_{\mathrm{b},\perp}(\mathbf{q})\right]\right\}}{1 - p(\mathbf{q})\exp\left[-H/\tau_{\mathrm{b}}^{\mathrm{int}}(\mathbf{q})\upsilon_{\mathrm{b},\perp}(\mathbf{q})\right]}. \tag{16}$$

This formulation of interface scattering allows for the rates of internal (intrinsic) and boundary roughness scattering to be added together, despite their interdependence (Turney et al. 2010). The factor given by Eq. (16) encapsulates the competition between boundary and internal scattering: the effective strength of boundary scattering will depend on the relative strength of the competing internal scattering mechanisms.

The full thermal conductivity tensor $\kappa^{\alpha\beta}$ is calculated as a sum over all phonon momenta and branches:

$$\kappa^{\alpha\beta}(T) = \sum_{\mathrm{b}} \sum_{\mathbf{q}} \tau_{\mathrm{b}}(\mathbf{q})C_{\mathrm{b}}(\mathbf{q}, T) \upsilon_{\mathrm{b}}^{\alpha}(\mathbf{q})\upsilon_{\mathrm{b}}^{\beta}(\mathbf{q}), \tag{17}$$

where $\tau_{\mathrm{b}}(\mathbf{q})$ is the total phonon relaxation time [for a bulk sample, $\tau_{\mathrm{b}}(\mathbf{q}) = \tau_{\mathrm{b,Internal}}(\mathbf{q})$ from Eq. (2)] and the phonon heat capacity per mode $C_{\mathrm{b}}(\mathbf{q}, T)$ is given by

$$C_{\mathrm{b}}\,(\mathbf{q},\,T) = \frac{[\hbar\omega_{\mathrm{b}}(\mathbf{q})]^2}{k_B T^2}\,\frac{e^{(\hbar\omega_{\mathrm{b}}(\mathbf{q})/k_B T)}}{\left[e^{(\hbar\omega_{\mathrm{b}}(\mathbf{q})/k_B T)} - 1\right]^2}. \tag{18}$$

All the phonon branches have to be considered in the calculation of heat transport (Klemens 1958), although optical branches represent standing waves with very low group velocities and hence contribute comparatively little to the overall thermal transport, relative to their acoustic counterparts. Despite this, in some nanostructures with strong boundary scattering, the contribution from the acoustic branches decreased to the point that the optical phonons, which do not scatter as often from boundaries due to their low group velocity, can make a non-negligible contribution (Tian 2011).

The $v_{\mathrm{b}}^{\alpha,\beta}(\mathbf{q})$ are the components of phonon velocity vector calculated from the full phonon dispersion based on Weber's adiabatic bond charge (ABC) model (Weber 1974). The ABC model includes interactions between ions, bond charges, bond bending, and long-range electrostatic interactions and has been shown to reproduce measured phonon vibrational frequencies in virtually all Group IV (Weber 1977; Rustagi and Weber 1976), III–V (Strauch and Dorner 1990; Rustagi and Weber 1976), and II–VI (Rajput and Browne 1996) semiconductors with excellent accuracy. It has also recently been re-parametrized from first principles (Colombo and Giannozzi 1995). The ABC phonon dispersions for Si can be found in Aksamija and Knezevic (2010a) and for Ge in Weber (1977) and Rustagi and Weber (1976). Vibrational properties of $Si_{1-x}Ge_x$ alloys, including phonon dispersion and velocity, are calculated in the virtual crystal approximation (VCA) (Abeles 1963), which has been shown to accurately capture the phonon modes of random alloys (Larkin and McGaughey 2013). The full phonon dispersion allows accurate treatment of the inherent anisotropy which arises out of the combination of phonon focusing and the momentum-dependent boundary-roughness scattering.

3   Lattice Thermal Conductivity in Si-Ge Alloy Thin Films

The calculated thermal conductivity of bulk and thin films as a function of varying Ge composition is shown in Fig. 1a. The symbols represent experimental measured data by Cheaito et al. (2012). The thermal conductivity rapidly decreases with increasing germanium composition in the range of $0 < x < 0.2$, and the strongest reduction is achieved as it reaches an almost flat plateau between $0.2 < x < 0.8$, after which the conductivity again increases between $0.8 < x < 1$. Virtual crystal approximation is applied to Si_xGe_{1-x} by replacing the mass-disordered lattice by an ordered virtual crystal with randomly distributed atoms of constituent materials. A similar reduction has been shown in $Al_xGa_{1-x}N$ alloys at high temperature using the VCA by Liu and Balandin (2005). They found that the thermal conductivity reduces from 200 W/mK for Al_1Ga_0N to 50 W/mK for $Al_{0.9}Ga_{0.1}N$ and 20 W/mK for $Al_{0.5}Ga_{0.5}N$, about a factor of 10 decrease in conductivity for the 50% GaN case.

We find that the reduction is much higher in Si-Ge alloys, about a factor of 18, as the conductivity reduces from 146 W/mK for Si_1Ge_0 to 8 W/mK for $Si_{0.8}Ge_{0.2}$.

The effect of nanostructuring in reducing thermal conductivity can also be observed in thin films. As shown in Fig. 1a, the strong quadratic dependence on germanium concentration dominates the variation in thermal conductivity. However, inside the plateau region, changing the thickness has a much greater effect on thermal conductivity than variation of Ge composition, while outside the plateau, compositional variation dominates. As the film thickness decreases, alloying induces smaller changes in the thermal conductivity as size effects begin to dominate. There is significant reduction in the thermal conductivities of thin films as compared to bulk, which is due to boundary scattering of the long wavelength phonons which are the primary thermal carriers. Consequently, achieving minimum thermal conductivity requires composition inside the plateau region in combination with nanostructuring well below the bulk limit.

In Fig. 1b, the calculated thermal conductivity of $Si_{1-x}Ge_x$ thin films is plotted against sample thickness for different boundary roughness, showing reduction in conductivity with decreasing film thickness. The behavior of thermal conductivity as the thickness is reduced to the subnanometer regime depends on the model and assumptions therein. Cahill's minimum thermal conductivity model (Cahill et al. 1992) implies that the scattering rate has an upper bound which is reached when the phonon lifetime equals one period of vibration. Hence the upper bound on the scattering rate is given by $\tau_{\max}^{-1} = \omega/\pi$, and setting this upper bound on the phonon scattering rate reproduces the experimental values for amorphous Si (1 W/mK) and Ge (0.6 W/mK), as shown by solid lines in Fig. 1b. In the completely diffuse limit (p = 0), the boundary scattering rate is proportional to thickness, and the conductivity scales linearly with thickness. If the boundary scattering rates are allowed to scale with decreasing thickness above the $\tau_{\max}^{-1}$, then the thermal conductivity gradually dips below the minimum value predicted by the minimum thermal conductivity model, often referred to as the amorphous limit, as shown by dashed lines in Fig. 1b. In that case, the trends are in agreement with recent measurements below the amorphous limit by Feser et al. (2013).

More generally, this model connects the boundary specularity $p(\mathbf{q})$ with roughness and finds that specularity depends exponentially on the rms roughness Δ; when thickness is reduced to the point that boundary roughness becomes comparable to the thickness, the exponential dependence of conductivity on roughness becomes more prominent, and the conductivity would tend to zero at a slower rate than the linear dependence predicted by the completely diffuse model. Hence, the dependence of the thermal conductivity on thickness would also include the consideration of roughness, and the amorphous limit is reached when both roughness and thickness become comparable to the unit cell size (0.543 nm in Si). Increasing the boundary roughness yields a reduction in thermal conductivity due to increased boundary scattering, more emphasized in larger samples as at smaller scales the conductivity is already limited by the size of the sample and increasing the boundary scattering will not reduce it further.
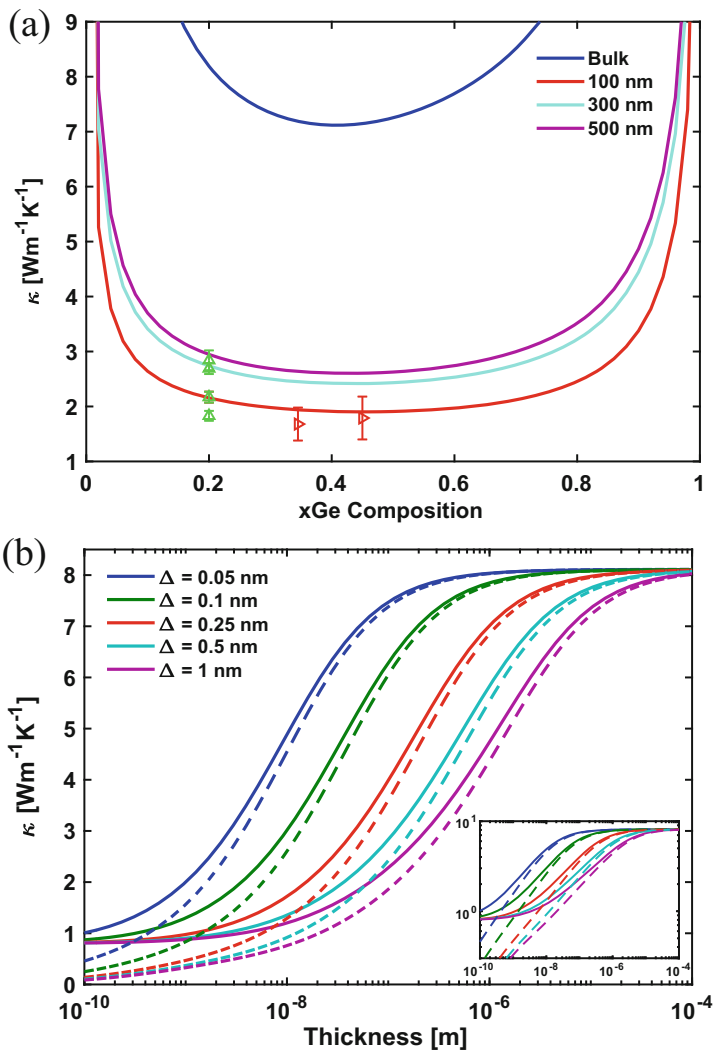
Fig. 1 (**a**) Thermal conductivity as a function of Ge composition is plotted for bulk and thin-film $Si_{1-x}Ge_x$ for three different thickness values of 100, 300, and 500 nm at room temperature. The symbols correspond to experimental data on the thickness series (upward triangles) and composition series (rightward triangles) by Cheaito et al. (2012). The plot demonstrates strong quadratic dependence on the germanium composition with a plateau between $0.2 < x < 0.8$. (**b**) Thermal conductivity of $Si_{1-x}Ge_x$ alloy is plotted versus film thickness for five different roughness values and germanium composition of 0.2. The solid lines represent the conductivity tending to the amorphous limit, and the dashed lines show the conductivity scaling with thickness below the amorphous limit. Increasing the roughness yields a gradual reduction in thermal conductivity, reaching the amorphous limit as the thickness approaches the nanometer scale. (Inset) The same plot on a log-log scale to show the conductivity scaling linearly with thickness below the amorphous limit (dashed lines). (Adapted from Upadhyaya et al. 2015)

4   Lattice Thermal Conductivity in Si-Ge Superlattices

SiGe SLs are bi-/multilayered structures composed of regular alternating layers of Si, Ge, or Si-Ge of varying composition, wherein the individual thickness of each layer is on the order of a few nanometers to several tens of nanometers. SLs have been found both experimentally and theoretically to have thermal conductivity below that of bulk $Si_{1-x}Ge_x$ alloys due to the combined effect of alloying and quantum confinement of particles due to nanostructuring. Bismuth-based SLs have been reported to possess very high figures of merit of around 2.6 (Venkatasubramanian et al. 2001). The impact of SL structure on lattice thermal conductivity has been attributed to a number of effects such as modification of the phonon spectrum, phonon localization, and scattering of phonons at the rough interfaces. Thermal conductivity of Si/Ge SL samples was measured to be lower than that of the corresponding alloy films (Huxtable et al. 2002), while in other materials, such as GaAs/InAs and bismuth-based SLs, the thermal conductivity was measured to be higher than that of the alloy film having equivalent overall composition.

Thermal transport in semiconductor SLs can be separated into two main directions based on symmetry: one being parallel called the in-plane transport (Khitun et al. 2001) and the other, perpendicular to the plane of the layers, called the cross-plane direction. The cross-plane direction is typically of much more interest for TE applications due to its lower lattice thermal conductivity (Cahill et al. 2003, 2014). In SLs with small period thickness, phonons are scattered by atomic-scale roughness present at the boundaries between successive layers. Roughness scattering is especially prominent in the cross-plane lattice thermal conductivity, which is governed by extrinsic rather than intrinsic effects (Chen 1998; Simkin and Mahan 2000; Yang and Chen 2001). This is particularly true for Si-Ge and $Si/Si_{1-x}Ge_x$ SLs, where the large mismatch between lattice constants and other material properties can lead to significant imperfections and appreciable roughness at the interfaces between layers, typically having several monolayers (ML, where $1\,ML \approx 0.13\,nm$) of roughness between the dissimilar layers in the SL (Huxtable et al. 2002). Therefore, interface properties, such as roughness (Martin et al. 2009), transport direction, and crystal orientation (Aksamija and Knezevic 2010a), in addition to alloy composition (Dismukes et al. 1964), are expected to play a significant role in thermal transport. They offer additional degrees of freedom to control the thermal conductivity in semiconductor nanostructures based on SLs, thereby opening up possibilities for further enhancement of the TE figure-of-merit through the reduction in lattice thermal conductivity.

Inside each layer of superlattice, every time the phonon reaches the boundary of that layer, one of two things must happen: the phonon will either continue traveling unscattered to the neighboring layer; this occurs in the case of a specular interaction, with probability given by the momentum-dependent specularity parameter $p(\mathbf{q})$, given earlier in Eq. (13). In the alternate case, called the diffuse case, the phonon will be scattered in a random direction with probability $1 - p(\mathbf{q})$. The addition of boundaries in one of the crystal directions breaks the symmetry and increases the anisotropy of the thermal transport between the direction of transport and

the direction normal to the boundaries. In general, since the boundary scattering rate for each phonon mode depends on the angle between its group velocity and the boundary normal, phonons interacting with the boundaries at low angles of incidence will undergo fewer and more specular collisions, thereby contributing more to transport in the direction parallel to the boundaries and increasing the anisotropy.

Once the thermal conductivity tensor κ for each of the two alternating layers is obtained, it can be decomposed into the two primary in-plane and cross-plane components. Owing to the cubic crystal symmetry of the underlying lattice, the tensor is diagonal. However, the diagonal components are not all equal due to the periodic SL structure breaking the rotational symmetry. From the full thermal conductivity tensor in Eq. (17), we can compute thermal conductivity by projecting it along any direction as $\kappa^{\alpha\beta} = \hat{n}^{\alpha}\kappa\hat{n}^{\beta}$ with $\hat{n}$ being the unit vector along the direction of transport (Aksamija and Knezevic 2010b). Subsequently, the conductivities of the two repeating layers in the SL will be combined in series for cross-plane transport and in parallel for in-plane transport; however, the cross-plane direction also has a small additional component of thermal resistance due to the acoustic mismatch between the layers, which adds a Kapitza resistance in series with the individual layers themselves:

$$\kappa_{\text{in-plane}} = \frac{L_1\kappa_1^{xx} + L_2\kappa_2^{xx}}{L_1 + L_2}$$

$$\kappa_{\text{cross-plane}} = \frac{L_1 + L_2}{\frac{L_1}{\kappa_1^{yy}} + \frac{L_2}{\kappa_2^{yy}} + \frac{1}{\sigma_1^{AIM}} + \frac{1}{\sigma_2^{AIM}}}. \quad (19)$$

The acoustic impedance mismatch (AIM) term has been found to add only a minor correction to the overall thermal conductivity in the extreme case of a Si/Ge SL and even smaller in alloy-based SLs due to smaller mismatch when both layers are composed of SiGe alloys (Aksamija and Knezevic 2013).

The additional scattering of phonons at the imperfect interfaces separating the SL layers leads to a two-order-of-magnitude reduction in thermal conductivity compared to bulk in Si/Ge SLs (Fig. 2a, b). The resulting thermal conductivity is lower than that of a $Si_{1-x}Ge_x$ alloy with the same proportions of silicon and germanium in agreement with Thumfart et al. (2017). This can be attributed to the additional boundary roughness scattering of phonons, especially the low-energy (long wavelength) acoustic phonons which are not strongly scattered by the Rayleigh-like mass-disorder scattering, given previously in Eq. (6). The lowest thermal conductivity is achieved when each of the layers is an alloy, which requires that the two layers are made up of different alloy compositions. In such a SL, thermal conductivity almost reaches the amorphous limit due to the combined effect of strong mass-difference scattering internal to the alloy layer and roughness scattering at the boundaries between the alloy layers, seen in Fig. 2c where both the layers are $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ alloys with different composition of x = 0.2 and y = 0.8. In SLs, both period thickness and total length play an important role; however,
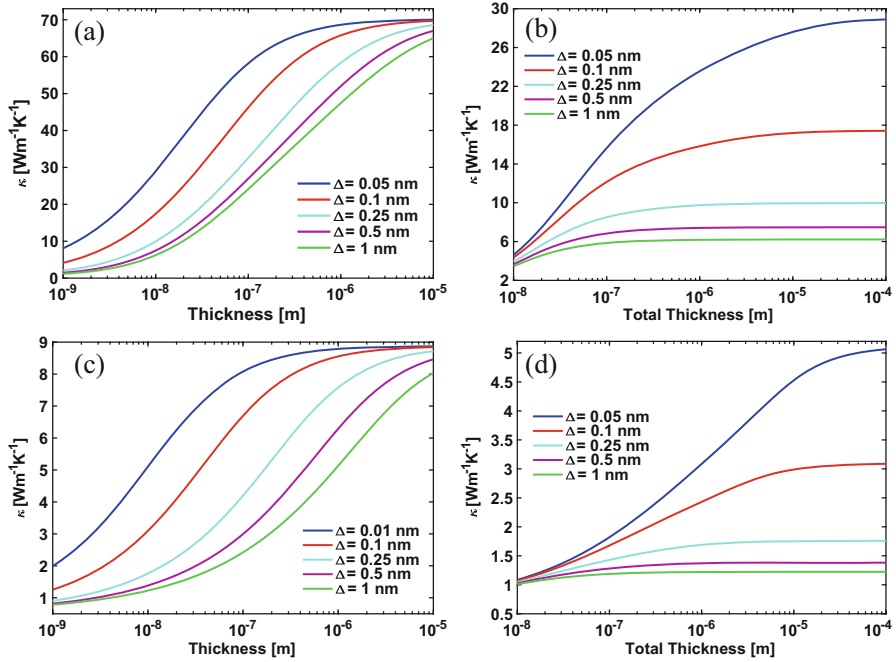
Fig. 2 Room temperature cross-plane thermal conductivity of a Si/Ge SL versus (**a**) SL bi-layer period thickness with fixed total sample thickness of $200\,\mu$m and (**b**) total sample thickness with fixed period thickness of 10 nm. Cross-plane thermal conductivity of a $Si_{0.8}Ge_{0.2}/Si_{0.2}Ge_{0.8}$ as a function of (**c**) bi-layer period thickness, with total sample thickness of $200\,\mu$m, and (**d**) total sample thickness with period thickness of 10 nm, for five different boundary roughness values of 0.05, 0.1, 0.25, 0.5, and 1 nm. Increasing the boundary roughness yields further reduction in thermal conductivity, more emphasized in larger samples because at smaller scales the conductivity is already limited by the overall size of the sample and increasing the boundary scattering between successive layers will not reduce it further. (Figure adapted from Upadhyaya et al. 2015)

total sample thickness is only significant when the roughness is small and the phonon mean free paths (MFPs) are not interrupted by the roughness scattering at the layer boundaries before they can reach the sample ends, as shown for both pure Si-Ge (Fig. 2b) and alloy SLs (Fig. 2d). The dependence on both sample and period thickness is more pronounced in pure Si/Ge SLs where the absence of alloy mass disorder implies that the thermal conductivity reaches the bulk limit in thick samples.

Figure 3a, b shows the in-plane conductivity for the Si-Ge and $Si_{0.2}Ge_{0.8}/Si_{0.8}Ge_{0.2}$ SLs. The in-plane thermal conductivity is higher due to the reduced effect of the boundary roughness scattering, which is more emphasized as the thickness of the individual layers is reduced. The inset figures show the anisotropy ratio (in-plane/cross-plane conductivity), which has the highest value at small layer thickness and saturates to a value of approximately as the layer thickness is increased above $1\,\mu$m.
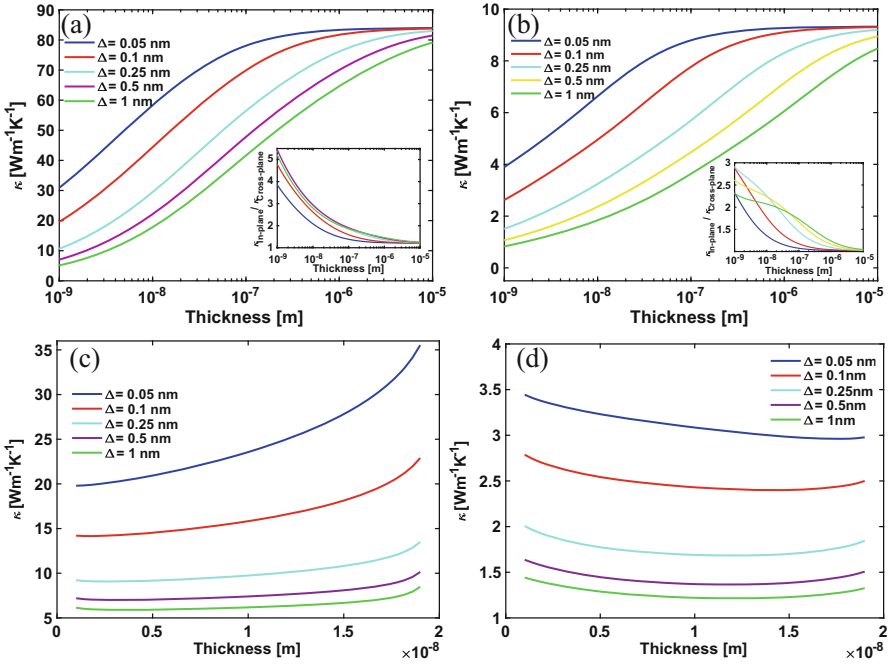
Fig. 3 Room temperature in-plane thermal conductivity of (**a**) Si/Ge SL and (**b**) $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ consisting of bi-layers with germanium composition of $x = 0.2$ and $y = 0.8$, respectively. Results are plotted as a function of bi-layer period thickness for five different boundary roughness values of 0.05, 0.1, 0.25, 0.5, and 1 nm (the total sample thickness is 200 μm). (Inset) The anisotropy ratio (in-plane/cross-plane conductivity) as a function of total sample thickness showing increasing anisotropy as thickness is reduced. Room temperature lattice thermal conductivity of (**c**) Si-Ge and (**d**) $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ consisting of bi-layers with germanium composition of $x = 0.2$ and $y = 0.8$. The period thickness is held fixed at 20 nm and total sample thickness is 200 μm, while the thickness of the first layer is varied from 0 to 20 nm. (Adapted from Upadhyaya et al. 2015)

Varying the relative thickness of the two layers, while keeping both period and total sample thicknesses fixed at 20 nm and 200 μm, respectively, shows that the lowest thermal conductivity occurs when the Si layer is thinner than the Ge layer in pure Si/Ge SLs (Fig. 3c). This is due to the lower intrinsic thermal conductivity of Ge ($\kappa \cong 56$ W/m/K at room T (Glassbrenner and Slack 1964)). In contrast, alloy SLs have lower thermal conductivity when the Si-heavy layer (with $x = 0.2$ of Ge) is dominant (thicker) and the second Ge-heavy layer (with $y = 0.8$) is thinner, as shown in Fig. 3d. This is again caused by the layer possessing lower thermal conductivity being preferred; in the alloy case, it is predominantly the Si layer which has lower conductivity (as seen from the asymmetry of the curves toward the Si side in Fig. 1a, especially in the bulk case) due to the smaller Si mass being further from the alloy average than the heavier Ge mass, causing more mass-disorder scattering, as defined in Eq. (5). Hence, composition must be precisely taken into account when

designing SiGe SLs for TE applications because the lattice thermal conductivity will in general be minimized by making the layer with lower thermal conductivity thinner; for example, in a SL where only one layer is composed of an alloy (while the other is either pure Si or pure Ge), the non-alloy layer should be made as thin as possible in order to emphasize the contribution from the roughness scattering, while the alloy layer can be made thicker since the mass disorder will keep its conductivity low.

5   Thermal Conductivity in Si-Ge Nanowires

Thermal transport along Si nanowires has been of particular interest owing to their potential applications in TE conversion (Boukai et al. 2008; Hochbaum et al. 2008). Ideally, the reduction of thermal conductivity due to alloy scattering could be combined with the effect of boundary scattering; then Si-Ge nanowires (SiGe NWs) could lead to even more dramatic improvements of ZT (Shi et al. 2009). MD simulations suggested that the reduction in alloy NWs (relative to small-diameter Si NWs) is only a factor of 5 (Chen et al. 2009). Several measurements of thermal conductivity in thin SiGe NWs with rough boundaries were performed (Kim et al. 2010; Yin et al. 2012; Lee et al. 2012) and confirmed the weak diameter dependence of thermal conductivity in Si-Ge NWs.

Semiconductor NWs are typically grown using the VLS procedure (Li et al. 2003), producing a circular geometry which leads to the usual Casimir limit $\tau_B^{-1} \propto v_s/D$ in the case where boundary scattering is independent of angle (Xie et al. 2014). The closed-form solution used in the planar (membrane or thin film) case (Aksamija and Knezevic 2013; Sellan et al. 2010) in the previous sections is not valid here because the distance from each point on the interior of the circular wire to the rough surface depends on both the angle (determined by the phonon group velocity vector) and the point of origin, in contrast to the planar case. In addition, the specularity of the boundary depends not only on the roughness but also on the angle of incidence between the phonon and the local boundary normal (Aksamija and Knezevic 2010a; Soffer 1967). To overcome these challenges, the phonon Monte Carlo (pMC) technique has been widely used to solve the phonon Boltzmann transport equation (pBTE) (Mazumder and Majumdar 2001; Chen et al. 2005; Lacroix et al. 2006; Jeng et al. 2008; Randrianalisoa and Baillis 2008; Péraud and Hadjiconstantinou 2011; Ramayya et al. 2012). The pMC allows us to sample the phonon lifetimes (McGaughey and Jain 2012) and find the combined phonon lifetime in the presence of both intrinsic scattering (from anharmonic phonon-phonon and mass-difference interactions inside the wire) and partially diffuse boundary scattering at the rough surface in circular wires (Martin et al. 2009).

In the phonon Monte Carlo algorithm, an ensemble of phonons is initialized according to the Bose-Einstein distribution (McGaughey and Jain 2012). Then the free-flight time until scattering of each phonon is determined by first selecting a random number $r_{int.}$, uniformly distributed between 0 and 1, and sampling the lifetime according to $t_{int.} = -ln(r_{int.})\tau_{int.}(\mathbf{q})$ (Maurer et al. 2015). The phonon

lifetime ($\tau_{\text{int.}}\mathbf{q}$) combines all the intrinsic scattering mechanisms. Once the phonon free-flight time is determined, each phonon travels along the propagation direction given by its group velocity $\mathbf{v}_g(\mathbf{q})$, until scattering at the end of the "free flight" at $t_{\text{int.}}$ or until it encounters a boundary or contact, whichever occurs first.

The boundary roughness scattering is characterized through a momentum-dependent specularity given in Eq. 13. When a phonon reaches the rough boundary, another random number $r_{\text{spec.}}$ is used to select between a specular and diffuse scattering. If $r_{\text{spec.}} < p(\mathbf{q})$, then the boundary interaction is specular and the phonon is simply reflected at the boundary by flipping its momentum about the boundary normal $\mathbf{q}_{\text{final}} = \mathbf{q}_{\text{init.}} - 2\mathbf{q}_{\text{init.}} \cdot \hat{s}$. Otherwise, the interaction is diffuse: the phonon path is terminated and the boundary scattering time t_B is recorded as the time at which the diffuse scattering occurred and the process is restarted. The quasi-ballistic contribution arising from phonons which reach the contacts (at time t_C) before scattering internally or at the boundaries is also captured. When all the phonon flights are terminated in either internal or boundary scattering or at a contact, thermal conductivity is computed from the average:

$$\kappa = \frac{1}{N_\mathbf{q} N_i} \sum_{\mathbf{q},i} v_g^2(\mathbf{q}) \min\{t_{int.}, t_B, t_C\} C(T, \mathbf{q}), \qquad (20)$$

where $C(T, \mathbf{q})$ is the modal volumetric heat capacity (McGaughey and Jain 2012) and $N_\mathbf{q}$, N_i are the number of phonons and iteration in the simulation, respectively, both being typically 100,000.

Figure 4a depicts the thermal conductivity vs. diameter for NWs of different lengths and surface roughness Δ. The results were computed for NWs with 20% Ge concentration at room temperature. The conductivity shows an almost linear diameter dependence at intermediate diameter values for pure Si (Xie et al. 2014), where boundary scattering is dominant. The deviation from this linear dependence increases with alloying. Alloy scattering rate follows a Rayleigh-like trend (Eq. 6) and suppresses most of the higher-frequency phonons, whereas the low-frequency (long-wavelength) phonons remain nearly unaffected.

The high-frequency phonons tend to undergo a more diffuse scattering at the boundaries causing a stronger diameter dependence seen in SiNWs, whereas the low-frequency phonons undergo a more specular boundary scattering and have smaller diameter dependencies, indicating that it is the midrange phonons that cause the weak diameter dependence in SiGe NWs. Surface roughness dependence of thermal conductivity is shown in Fig. 4b. There is a steady decrease in conductivity with an increase in surface roughness up to 1 nm due to more diffuse boundary scattering, but increasing the roughness further does not reduce the conductivity, which saturates to a value much lower than in pure SiNWs. However the reduction in conductivity due to increased diffuse boundary scattering in SiGe NWs is not as effective as in pure Si NWs, primarily in thin wires where boundary scattering is dominant, due to the suppression of the high-frequency phonons by alloy scattering. For a pure Si NW of 10 nm diameter and 100 nm length, there is a 75% reduction
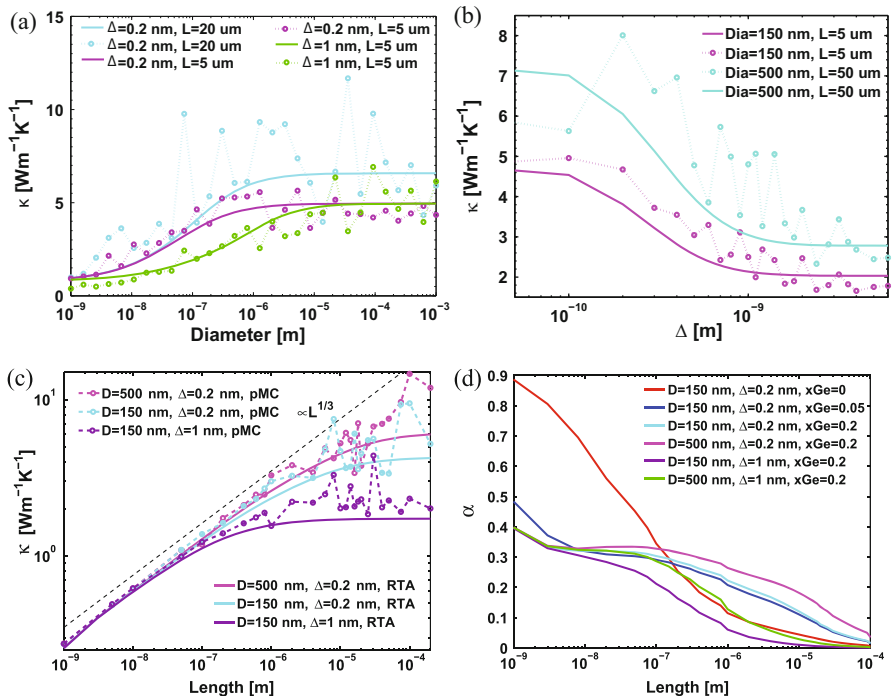
Fig. 4 (**a**) Thermal conductivity is plotted as a function of the wire diameter. The diameter dependence is weak due to strong intrinsic scattering. The solid lines represent the relaxation time approximation (RTA), and dashed lines represent the MC results. (**b**) Thermal conductivity is plotted as a function of surface roughness. For Δ values of 1 Å or less, the conductivity is unaffected, and it steadily decreases as the roughness is increased, without any further significant decrease beyond 1 nm. (**c**) Thermal conductivity as a function of length. The conductivity exhibits a $L^{1/3}$ behavior and gradually transitions into the diffusive regime at lengths exceeding 10 μm. (**d**) Exponent of length dependence α as a function of length. In SiGe NWs, $0.3 < \alpha < 0.4$ over a broad range of lengths, indicating non-diffusive transport. (Adapted from Upadhyaya and Aksamija 2016)

in conductivity when surface roughness is increased from 0 to 1 nm, whereas the reduction is about approximately 52% for a SiGe NW of the same dimension.

Figure 4c shows the dependence of thermal conductivity from pMC simulations on NW length. The dashed lines and symbols are the Monte Carlo results, while solid lines are the deterministic solution of the pBTE in the RTA, shown for comparison, and they are in close agreement. There is a gradual change in thermal conductivity with length, as seen in Fig. 4c, with the crossover to purely diffusive transport only occurring at lengths exceeding 10 μm, far in excess of the average phonon MFP and in agreement with measurements (Hsiao et al. 2013). Alloying suppresses most of the high-frequency phonons while the low-frequency phonons possess very long MFPs, allowing them to travel several microns without being scattered internally. However, even at very small NW lengths, the linear trend in the

length dependence is not present that would be characteristic of ballistic transport; instead, as seen in Fig. 4c, the conductivity scales as $L^{1/3}$.

A plot of the running exponent, defined as $\alpha(L) = d \ln \kappa(L)/d \ln L$ (Saito and Dhar 2010), in Fig. 4d shows that all SiGe NWs follow the same trend regardless of diameter, with 0.4, even when $L < 10$ nm. In contrast, short Si NWs reach the fully ballistic regime (characterized by $\alpha = 1$). The $\alpha \approx 0.33$ behavior has been observed in many momentum-conserving systems (Dhar 2008), including one-dimensional chains (Lepri et al. 2003; Mai et al. 2007), alloy thin films (Vermeersch et al. 2016), and, over a much narrower range of lengths, even thin Si NWs (Yang et al. 2010). However, the upper limit of length at which exponent $\alpha \approx 1/3$ depends on boundary scattering: in a rough wire ($\Delta = 1$ nm), the exponent reduces to the diffusive $\alpha = 0$ at a shorter length than in a smooth wire (where $\Delta = 0.2$ nm). Diffuse boundary scattering limits the longest MFP and thus results in a more uniform MFP distribution. Hence boundary scattering affects the range of length over which $\alpha = 1/3$ is observed, but not the length-scaling exponent α. Alloy scattering, on the other hand, results in an intrinsically different mode of transport having a broader range of MFPs with very few purely ballistic phonons.

Non-diffusive behavior of phonons can be readily seen in a plot of the distance along the NW traveled by one simulated phonon through a sequence of free flights interrupted by scattering events. Each flight number in the plot corresponds to a single free flight between successive scattering events for both Si (Fig. 5a) and $Si_{0.8}Ge_{0.2}$ (Fig. 5c). One phonon is chosen from the ensemble at random as a representative and plotted, as all of the simulated phonons exhibit qualitatively similar behavior. The scattering is predominantly elastic, arising in both cases from boundary roughness, while in alloy NWs there is a strong additional component due to mass disorder. It is interspersed by less frequent inelastic (anharmonic phonon-phonon) events which mix the phonon modes. Si NWs have a continuous distribution of distances due to the series of relatively uniform flight characteristic of diffusive transport, leading to a white noise-like appearance seen in Fig. 5a, which is readily associated with Brownian motion and predominantly diffusive transport.

In contrast, phonon flights in SiGe NWs are comprised of sequences of many short flights interrupted by rare long leaps, as evidenced by the micron-sized jumps in the distance traveled by the phonon shown in Fig. 5c. This behavior in the alloy is a consequence of the strong mass-disorder scattering, which affects the upper portion of the phonon spectrum far more than the low-frequency modes. Viewing the whole phonon ensemble collectively, the result is a heavy-tailed distribution of free-flight lengths, shown in Fig. 5e. The difference is especially prominent when alloy NWs are compared to pure Si NWs, in which the tail of the phonon flight distribution decays faster. Phonons making long jumps exceeding $1\,\mu$m are more than twice as frequent in SiGe NWs as they are in Si NWs, as seen in the distribution of flight lengths Δx as measured in the direction of heat flow along the NW. The difference between alloy and non-alloy steadily increases for longer leaps, in spite of alloy scattering causing phonons in SiGe to have MFPs more than an order of magnitude shorter on average than pure Si. The heavy-tailed behavior is characteristic of Lévy walk dynamics (Dhar et al. 2013; Zaburdaev et al. 2015), which has already been
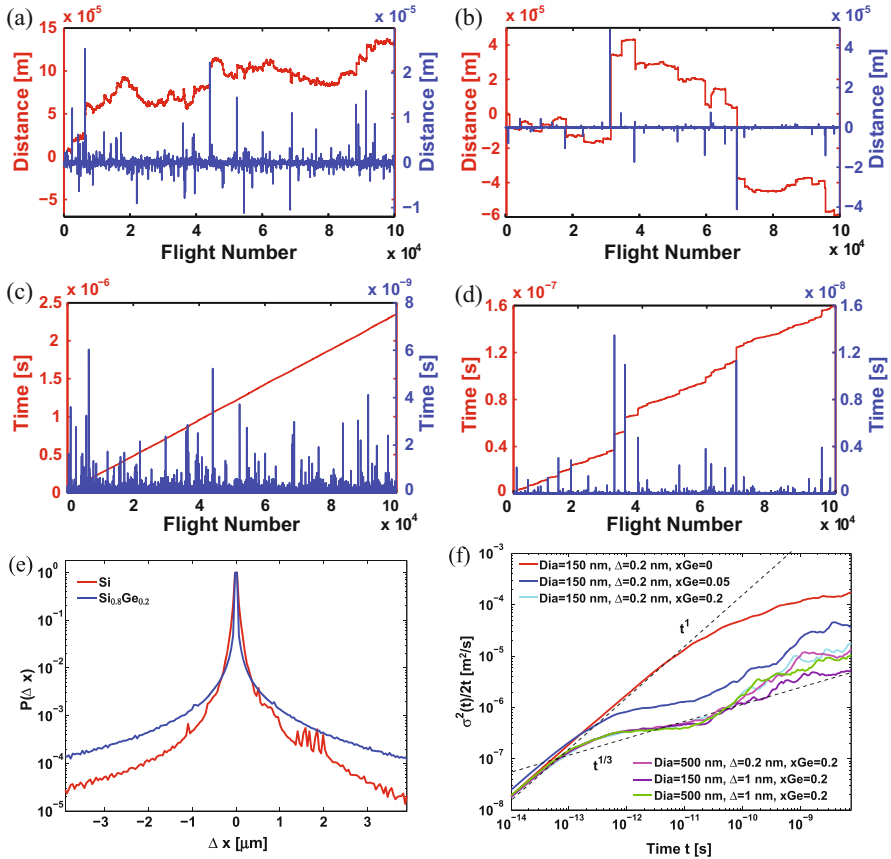
Fig. 5 The distance traveled by one simulated phonon between successive scattering events (blue dotted line) and the cumulative distance (red line) vs. flight number is plotted for (**a**) a pure Si NW and (**b**) a $Si_{0.8}Ge_{0.2}$ NW. Also, the time between scattering events (blue dotted line) and the cumulative time (red line) is plotted vs. the flight number in (**c**) pure Si NW and (**d**) $Si_{0.8}Ge_{0.2}$ NW. (**e**) Histogram of the individual free-flight lengths in Si (red) and $Si_{0.8}Ge_{0.2}$ NWs (blue). SiGe NWs show a larger proportion of long leaps, which leads to a heavy-tailed distribution. (**f**) Diffusion coefficient vs. time. In SiGeNWs there is a broad intermediate regime in which the exponent of the diffusion coefficient is ≈ 0.33 over several orders of magnitude in time, whereas in SiNWs $\alpha = 1$ in the ballistic regime followed by a smooth transition into the diffusive ($\alpha = 0$) regime. (Adapted from Upadhyaya and Aksamija 2016)

linked to superdiffusive phonon transport in low-dimensional (Cipriani et al. 2005) and alloy systems (Vermeersch et al. 2015).

Non-diffusive behavior of phonons in alloy NWs is further examined by the time-dependent phonon transport calculated from pMC simulations. The time dependence of the mean square energy displacement (MSD) (Denisov et al. 2003) is calculated from $\sigma^2(t) = \langle \Delta x^2(t) \rangle$ and related to an exponent $\sigma^2(t) \propto t^{\beta}$. Consequently, the diffusion coefficient $\sigma^2(t)/2t \propto t^{\beta-1}$. The exponent of length

dependence α, shown previously in Fig. 4d, has also been related to the MSD through $\alpha = \beta - 1$ (Liu et al. 2014). In normal diffusion, where Fourier's law remains valid, phonons undergo Brownian motion resulting in $\beta = 1$ (as $\sigma^2(t) \propto t$) (Shlesinger et al. 1993), which also means the exponent of length dependence $\alpha = 0$ and the conductivity is constant, independent of length. In contrast, when the system size is smaller than the MFP, phonon flights are uninterrupted by scattering, and their distance from origin grows in proportion with time; hence, ballistic MSD is quadratic in time ($\beta = 2$), implying that $\alpha = 1$ and the conductivity is linearly proportional to length (Li et al. 2015).

   $\alpha = 1/3$ scaling is attributed to an intermediate regime of superdiffusion: when $1 < \beta < 2$ transport is neither entirely ballistic nor diffusive. Instead, it is collectively characterized by a mix of long quasi-ballistic leaps, interrupted by bursts of short, diffusive steps. Consequently, the diffusion coefficient $\sigma^2(t)/2t \propto t^{\alpha}$ should also imply a length exponent $0 < \alpha < 1$ in the superdiffusive regime. The diffusion coefficient $\sigma^2(t)/2t$ is plotted as a function of simulation time in Fig. 5f, and $\alpha = 1/3$ (and $\beta = 4/3$) is observed over a wide range of timescales in SiGeNWs. In bulk SiGe alloys and SiGe thin films in the cross-plane direction, Vermeersch et al. (2015, 2016) have shown that $\beta = 1.34$, confirming that length scaling is a consequence of superdiffusion.

   The range of lengths over which superdiffusion is observed ($\alpha \approx 1/3$ over $10\,\text{nm} < L < 10\,\mu\text{m}$) far exceeds the average phonon MFP in the system; instead, it maps directly to the wide range of timescales over which superdiffusion is observed here in SiGeNWs (2 ps $< t <$ 2 ns in Fig. 5f) using the simple cutoff $t_s \approx L/v_s$ for superdiffusion in finite systems (Lepri et al. 2003; Liu et al. 2014), with v_s being the speed of sound ($v_s \approx 5000$ m/s in SiGe, depending on alloy composition). The diameter and roughness do not alter the exponent of the diffusion coefficient; instead, they affect only the onset of the transition from superdiffusion ($0 < \alpha < 1$) into purely diffusive regime ($\alpha = 0$), reducing the resulting conductivity in the steady state analogously to its length dependence in Fig. 4c. Thus, lattice conductivity is length tunable even in NWs several microns long, with potential applications to reducing thermal conductivity and thus increasing TE figure of merit in alloy NWs with sub-ten-micron lengths.

6   Thermal Conductivity in Si-Ge Nanocomposites

Despite their desirable properties, the cost of growing SLs is prohibitive, and nanocomposites (NCs) offer a tremendous advantage in their ease of manufacturing. Rather than having to grow SLs layer by layer from the ground up, NCs are made by relatively simple and cheap mechanical processing of materials, such as ball milling the bulk material into a fine powder of nanoscale grains, followed by spark plasma sintering into a nanostructured bulk composite (Liu et al. 2012), thereby making such nano-bulk materials much more amenable to scaling and commercialization (Lan et al. 2010). In addition to Si-Ge NCs, nanocrystalline Si has similar grain

structure and has also been studied as a platform to achieve high ZT in a Si-based system (Neophytou et al. 2013; Shiomi 2016).

However, the resulting NCs have a very heterogeneous structure with grains ranging in size from a few nanometers to hundreds of nanometers, typically following an exponential distribution (Lan et al. 2009). The grain properties, including atomic-scale grain boundary roughness (GBR), orientation, and composition, will also substantially affect thermal transport in NCs, offering numerous ways of adjusting their thermal conductivity by manipulation of grain size, shape, and crystalline angle distributions. Hence NCs can be designed to improve ZT by reducing lattice thermal conductivity by a combination of approaches including additional scattering of phonons from the interfaces between nanoscale grains of dissimilar materials, alloys of different composition, or grains with different crystal orientation (Kazan and Volz 2014).

Our model for NCs is based on solving the phonon Boltzmann transport equation (pBTE) with partially diffuse boundary conditions at the grain boundaries, similar to the model described previously in Sect. 3, but modified to capture fully the complex heterogeneous structure of the NC (Aksamija 2014). Voronoi tessellation (VT) is employed to discretize the material into grains which resemble closely the complex random grain structure of real NCs (Priolo et al. 1992; Bartkowiak and Mahan 1995). A cubic region is discretized, typically several micrometers in each direction, using the VT technique, which starts with a set of seed points and determines the division of the simulation domain into distinct and non-overlapping polyhedra. Each cell grows from its seed point until reaching its nearest neighbors at which point a grain boundary forms. The grain boundary always bisects the line connecting adjacent seed points. The VT follows the rules for constructing Wigner-Seitz cells except on the scale of grains rather than lattice cells (Aksamija and Knezevic 2014). Once the VT is complete, it provides critical information about the morphology of the NC, including the size of each individual grain, the grain size distribution, their interconnectivity (who the neighbors of each grain are), as well as the area of the contact region between neighboring grains. The grain size distribution is typically lognormal when starting from seed points uniformly distributed at random (Ferenc and Nda 2007) but can be modified to follow other prescribed size distributions (Fan et al. 2004).

When grain boundaries are introduced, the pBTE solution becomes position-dependent along the direction of propagation of the phonon mode (b, $\mathbf{q}$) because of scattering at the grain boundaries, which partially randomizes the direction of propagation of the incoming phonons, and scattering inside the grain. At the grain boundary, phonons will interact with the atomic-scale roughness, and a fraction $p_{\mathbf{q}}$ of the incoming phonons will pass through unchanged (coherent or specular part) while the rest $(1 - p_{\mathbf{q}})$ will be absorbed by the boundary and reemitted in a random direction (diffuse part). An effective scattering rate Γ_i^{GBR} due to the interactions of phonons with the GBR in the i-th grain can be defined by averaging the pBTE solution over the volume of each grain (Carruthers 1961; Wang and Mingo 2011) to obtain the final expression for the effective GBR scattering rate:

$$\frac{1}{\tau_{b,i}^{GBR}(\mathbf{q})} = \frac{v_b(\mathbf{q})}{D_i} F_{b,i}(\mathbf{q}) \left/ \left[1 - \frac{\Lambda_b^{int.}(\mathbf{q})}{D_i} F_{b,i}(\mathbf{q}) \right] \right. , \tag{21}$$

with D_i being the average distance a phonon travels inside the i-th grain. D_i can be related to the cubic root of the volume of the grain V_i, calculated for each grain from the VT, as $D_i = f_i S_i^{1/3}$, where f_i is the dimensionless form factor which captures the peculiarities of the specific geometry of each grain. For grains roughly approximated as disks, $f = \frac{8}{3\pi^{3/2}} \approx 0.48$. The parameter $F_{b,i}(\mathbf{q})$ fully captures the competition between GBR scattering at the boundaries of the grain and internal scattering inside the grain and is completely analogous to the one given previously for thin films, with D_i replacing the film thickness L. The GBR scattering rate $1/\tau_{b,i}^{GBR}(\mathbf{q})$ in Eq. (21) can be added to the rates due to intrinsic scattering processes internal to the grain to obtain the total scattering rate of mode $(b, \mathbf{q})$ in grain i as $1/\tau_{b,i}(\mathbf{q}) = 1/\tau_b^{int.}(\mathbf{q}) + 1/\tau_{b,i}^{GBR}(\mathbf{q})$.

In the steady state, the net heat flux through the boundaries of each grain must be zero. Therefore, fluxes in and out of its neighbors and any external heating source Q_i applied to i-th grain must add up to zero. This assertion allows us to calculate the temperature T_i inside the i-th grain from:

$$\sum_{j=\text{n.n.}} G_{ij} (T_i - T_j) + S_i Q_i = 0, \tag{22}$$

with the summation index j running over all the grains that neighbor grain i. The thermal conductance G_{ij} between neighboring grains i and j is given by the usual expression:

$$G_{ij} = \frac{\left[d_i \kappa_i(T_i) + d_j \kappa_j(T_j) \right] A_{ij}}{d_{ij}^2}. \tag{23}$$

with κ_i being the thermal conductivity of the i-th grain including the corresponding GBR scattering rate. Here, $d_{ij} = d_i + d_j$ is the distance between the centers of the two grains, i.e., the length of the straight line connecting the two centers, while d_i and d_j are the lengths of the line's segments inside the i-th and j-th grains, respectively. A_{ij} is the surface area of the grain contact region separating the two grains. This way, thermal resistor network is obtained, which can be solved iteratively (Braginsky et al. 2002), repeatedly updating the temperature of each grain based on the past values of the temperatures of its neighbors until convergence.

A heat sink is applied to the outer grains in the simulation domain so that the boundary condition on temperature outside of radius r_2 is kept constant. A heat source Q is applied to the grains inside a radius $r_1 = 1\,\mu$m around the center and is allowed to diffuse. Q is determined, upon iteration of the temperature, to ensure the peak temperature in the center to be a few Kelvin ($\approx 1\%$) above the temperature of

the heat sink. Convergence is reached when the total flux entering each grain equals the flux leaving it and the temperatures of all grains reach a constant value. After the temperature throughout the simulation domain reaches steady state (measured by the relative change in the temperature from one iteration to the next), the effective thermal conductivity is extracted using the below equation:

$$\kappa_{\text{eff}} = \frac{Q \log\left(\frac{r_2}{r_1}\right)}{2\pi (T_1 - T_2)}. \tag{24}$$

Effective thermal conductivity depends on the direction of heat propagation inside the NC, leading to strong directional anisotropy which arises from the local morphology of the grains, their sizes, and interconnections.

We find that the lattice thermal conductivity decreases with average grain size, as shown in Fig. 6b, analogous to the thickness dependence in thin films and SLs
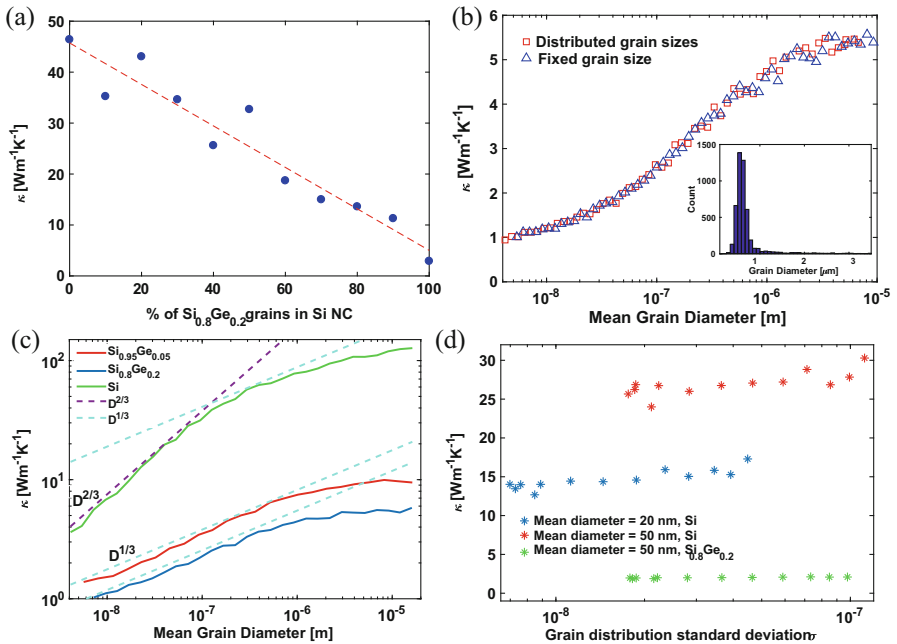
Fig. 6 (a) Thermal conductivity of a Si/Si$_{0.8}$Ge$_{0.2}$ NC is plotted versus the percentage of Si$_{0.8}$Ge$_{0.2}$ grains. (b) Thermal conductivity of a Si$_{1-x}$Ge$_x$ NC with germanium composition of x = 0.2 is plotted versus average grain size for both uniform and randomly distributed grains. (Inset) The distribution of grain sizes. (c) Thermal conductivity of a Si$_{1-x}$Ge$_x$ NC with germanium composition of x = 0, 0.05 and 0.2 is plotted versus average grain size. Conductivity scales as $D^{2/3}$ in Si NC and $D^{1/3}$ in SiGe NCs indicating superdiffusive transport. (d) Thermal conductivity vs. standard deviation of the grain size distribution showing negligible effect of size distribution on conductivity. (Partially adapted from Upadhyaya et al. 2015)

discussed in previous sections. Further reduction in thermal conductivity can be achieved by making a NC with grains comprising of different material/composition due to increased mass-difference scattering. This trend is observed when a NC is made of intermixed Si and SiGe grains as shown in Fig. 6a. The size dependence is gradual and begins at grain sizes of several micrometers and saturates toward the amorphous limit as grain size approaches the nanometer scale, as seen in Fig. 6c.

It is interesting to note that despite the heterogeneity and disorder in the NC structure, the crossover to purely diffusive transport occurs only when average grain diameter exceeds $10\,\mu m$, similar to what we observed in SiGe NWs. The conductivity scales as $D^{2/3}$ in Si NC and $D^{1/3}$ in SiGe NCs, indicating a breakdown of Fourier's law and superdiffusive transport in both these systems. Even in the absence of alloy scattering, phonon transport in Si NCs is never predominantly ballistic due to the strong GBR scattering at the interface of each grain. However, the wide distribution of grain sizes (Fig. 6b (inset)), coupled with momentum-dependent GBR scattering that is more specular for long wavelengths, suppresses mid- and high-frequency phonons while allowing low-frequency (long-wavelength) phonons with long MFPs to travel several microns before being scattered internally, somewhat analogous to alloy scattering. This leads to superdiffusive transport even in non-alloy NCs.

On the other hand, the shape of the grain size distribution appears to have little effect, at least in alloy NCs, as both uniform and randomly distributed grain sizes lead to the same thermal conductivity for a given average grain size. We study the dependence on the grain size distribution in more detail by starting with a uniform distribution of seed points, which leads to a VT with uniform grains and a tight distribution of grain sizes. Next, we increasingly perturb the initial seed points to reach a broader size distribution and compute the standard deviation of grain sizes from the mean and its corresponding thermal conductivity. Our finding that the shape of the distribution has modest effect is further confirmed by Fig. 6d, which shows that varying the standard deviation of the grain size distribution in Si-Ge NCs has no effect on the thermal conductivity, after averaging over all transport directions. This trend is explained by the gradual dependence of the thermal conductivity of each grain on its size, caused by the competing alloy scattering in their interior. On the other hand, in Si NCs where the size dependence is more pronounced, we observe a gradual increase with as much as 20–30% variation, in agreement with a recent study (Hori et al. 2015).

We conclude that our model, based on the Voronoi tessellation and the pBTE solution in each grain, captures the grain structure of the NC and its impact on the thermal conductivity much more directly than empirical approaches such as the effective medium approximation (EMA) (Aksamija 2014). Reducing grain size into the nanometer regime or adding nanoscale grains into the bulk matrix is the most effective way to reduce thermal conductivity in NCs, with the additional benefit of lower cost and simpler fabrication than SLs. At such scales, energy-filtering effects are expected to lead to a boost, rather than a reduction, in the electronic power factor (Zebarjadi et al. 2011).

7   Beyond SiGe: Lattice Thermal Conductivity in Binary and Ternary Group-IV Alloys

Lately, there has been some interest in Sn-based Group-IV alloys, in particular, because of their widely tunable band gap (Zhu et al. 2015), which has opened the possibility of Si-based direct-band-gap semiconductors (Moontragoon et al. 2012; Gaiduk et al. 2014) with potential application in optoelectronics (Kouvetakis et al. 2006; Soref et al. 2007). Ternary Group-IV alloys such as $Si_{1-x-y}Ge_xSn_y$ and their heterostructures are also potential candidates for IR devices and quantum-well photodiodes (Fischer et al. 2015). However, growth of Si-Sn and, to a lesser extent, Ge-Sn alloys with high Sn concentration remains challenging due to the large lattice mismatch (19.5% between Si and α-Sn) and low solubility of Sn in Si ($\approx 5 \times 10^{19}$ cm^{-3}) (Min and Atwater 1998). Spurred on in part by potential TE applications (Tonkikh et al. 2014; Khatami and Aksamija 2016), recent work has demonstrated growth of thin-film $Si_{1-x}Sn_x$ with large concentration x on Si, Ge, and InP substrates (Kurosawa et al. 2017).

The phonon transport model used for computing thermal conductivity in Si-Ge alloys described in Sect. 2 holds for ternary Group-IV alloys as well. However, the isotope scattering term is modified to include contribution due to isotopic variation in each of the three constituent materials by combining the isotope constants as

$$\Gamma_{\text{Iso}}(x, y) = \frac{(1 - x - y)\Gamma_{\text{Si}}M_{\text{Si}}^2 + x\Gamma_{\text{Ge}}M_{\text{Ge}}^2 + y\Gamma_{\text{Sn}}M_{\text{Sn}}^2}{(xM_{\text{Ge}} + (1 - x - y)M_{\text{Si}} + yM_{\text{Sn}})^2}, \tag{25}$$

where Γ_{Sn} is calculated from known isotope composition of naturally occurring tin to be $\Gamma_{\text{Sn}} = 3.64 \times 10^{-4}$ from Eq. 4, which holds for isotope scattering within each material in the alloy (Liu and Asheghi 2005). An additional component to alloy scattering arising from the strain field due to the difference in lattice constants of pure Si and Ge and their alloys has been proposed. The contribution due to strain is then given by (Abeles et al. 1962)

$$\Gamma_{\text{Strain}} = \varepsilon \sum_i f_i(1 - a_i/\bar{a})^2, \tag{26}$$

where f_i is the proportion of material i with lattice constant a_i in this case Si, Ge, and Sn while the average lattice is given by $\bar{a} = \sum_i f_i a_i$. Here $a_{\text{SiGeSn}}(x)$ is the composition-dependent alloy lattice constant, taken in the virtual crystal approximation (VCA), including bowing (Rieger and Vogl 1993). The empirical strain parameter is taken to be $\varepsilon = 39$ (Abeles 1963).

The lattice thermal conductivity of bulk ternary alloy SiGeSn is calculated from the full dispersion pBTE model. Figure 7a depicts thermal conductivity of bulk SiGeSn alloys against the contributed composition of each material. Thermal conductivity of binary alloy SiGe can be seen along the edge where Sn composition is equal to zero, similarly for SiSn and GeSn where Ge and Si composition,
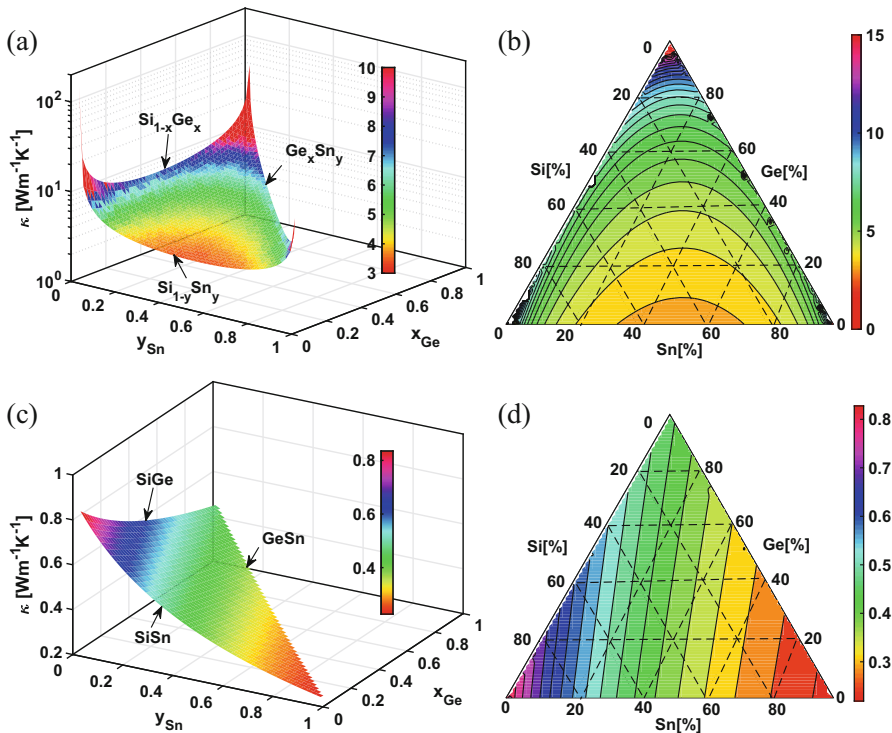
Fig. 7 (**a**) 3D plot of thermal conductivity vs. Sn and Ge composition in bulk SiSn, GeSn, and SiGe. SiSn has the lowest thermal conductivity in bulk form. (**b**) Ternary plot of thermal conductivity vs. alloy composition. (**c**) 3D surface plot of the lowest achievable thermal conductivity vs. alloy composition in SiGeSn. Binary alloys correspond to the data points on the edges of the triangle and show that Sn has the lowest thermal conductivity in the amorphous (disordered) limit. (**d**) Ternary plot of the same. (Figure adapted from Khatami and Aksamija 2016)

respectively, are equal to zero. Similar to SiGe (discussed in Sect. 3), thermal conductivity reaches a broad plateau in the alloy composition range $0.2 < y < 0.8$ of tin composition in $Si_{1-y}Sn_y$ and $Ge_{1-y}Sn_y$, due to dominance of mass-disorder scattering which depends quadratically on the difference in atomic mass between the constituent materials and the average mass. The lowest thermal conductivity for bulk binary alloy SiSn is 3 W/mK, at Sn composition of 0.5, is in good agreement with RNEMD calculations with mass ratio of 4.2 between Si and Sn (Frachioni and White 2012), and 5.86 W/mK in GeSn at Sn composition of 0.61, in agreement with the range of values (5–9 W/mK) measured in annealed poly-GeSn (Uchida et al. 2015). The large mass difference between Si and Sn results in a higher mass-disorder scattering rate, hence a lower lattice thermal conductivity, below the lowest value of 6.67 W/mK at Ge composition of 0.38 in SiGe. Fig. 7c shows the calculated amorphous (disordered) thermal conductivity, which is below 1 W/mK, and provides some indication of the minimum conductivity achievable through
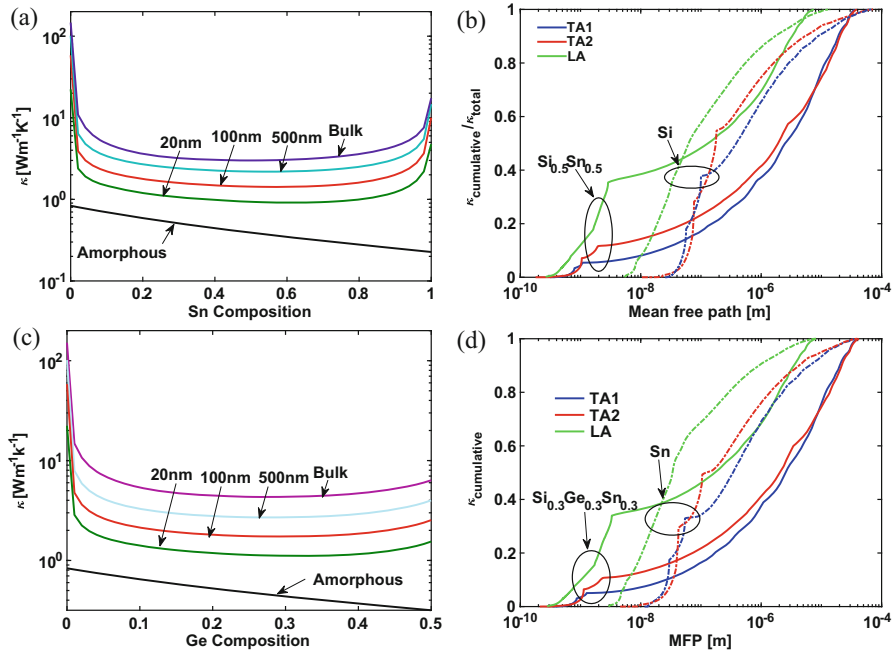
Fig. 8 Thermal conductivity of (**a**) SiSn vs. Sn composition and (**c**) SiGeSn vs. Ge/Sn composition, for bulk, 500, 100, and 20 nm thick samples, simulated at room temperature with roughness of 0.45 nm and sample thickness of 1 μm. Cumulative thermal conductivity vs. MFP of (**b**) pure Si (dashed lines) and SiSn alloy (solid lines), (**d**) pure Sn (dashed lines) and $Si_{0.3}Ge_{0.3}Sn_{0.3}$ alloy (solid lines). (Figure adapted from Khatami and Aksamija 2016)

disorder (a combination of mass/alloy and boundary roughness) in each of these binary and ternary alloys.

The lowest thermal conductivity of SiSn at 20 nm thickness is 0.91 W/mK (Fig. 8a), achieved at Sn composition of 0.59. This value is below that of a high-performance TE BiSbTe alloy NC (κ_{ph}=1.1 W/mK at room temperature) having comparable (20 nm average) grain size (Poudel et al. 2008). SiSn alloys with high Sn concentration have not been demonstrated previously, due to the low solid solubility of Sn in Si; nonetheless, thermal conductivity in 20 nm thick SiSn film at 18% Sn concentration, the highest Sn fraction demonstrated thus far (Kurosawa et al. 2015), only increases to 1.22 W/mK, which is nearly a 30% reduction from the lowest value achievable in SiGe thin films of equal thickness. A room temperature thermal conductivity of 2.5 W/mK, with a lattice contribution of 1.8 W/mK and corresponding ZT of about 1.3 at 900 °C, was reported in a $Si_{0.8}Ge_{0.2}$ NC with average nanoparticle size of about 20 nm (Wang et al. 2008; Lan et al. 2010). The calculated thermal conductivity of $Si_{0.82}Sn_{0.18}$ is 50% smaller that could potentially lead to a ZT of 1.7, assuming the same electron contribution to thermal conductivity $\kappa_e = 0.7$ W/mK. In p-type SiGe NCs, the peak ZT value is around 30% smaller than

n-type, likely due to lower hole mobility (Fischetti and Laux 1996), reaching 0.95 at 850 °C (Joshi et al. 2008; Zhu et al. 2009).

Figure 8c depicts the ternary alloy thermal conductivity vs. Ge and Sn composition, keeping Ge x and Sn y composition equal ($x = y$). x and y are varied from 0 to 0.5, while the silicon composition ($1 - x - y$) is reduced from 1 to 0. At $x = y = 0.32$ and the film thickness of 20 nm, the thermal conductivity reaches its lowest value of 1.11 W/mK, slightly higher than the lowest value achieved in SiSn at that same thickness. However, this is the closest value to the lowest achievable thermal conductivity given by the corresponding amorphous limit.

Figure 8b, d shows the normalized cumulative thermal conductivity vs. MFP in SiSn and SiGeSn, respectively. The optical modes are omitted due to their low contribution to thermal conductivity, caused by their very flat dispersion and consequently low phonon group velocities. The density of states peaks near van Hove singularities and the sharp bends in the slope of the cumulative thermal conductivity plot correspond to these peaks. In the diffuse case, boundary scattering limits the phonon MFP; hence, only phonons having MFP smaller than or equal to the film thickness will contribute to thermal conductivity, and its value can be estimated using the cumulative plot by taking the MFP equal to thickness and reading off the corresponding thermal conductivity. In SiSn, phonons having MFPs below 20 nm still contribute around 40% of thermal conductivity, significantly more than they contribute in SiGe (Khatami and Aksamija 2016). Alloying leads to a broader range of MFPs and a more gradual dependence of conductivity on MFP in Fig. 8c.

Group-IV alloys containing Sn have tunable band gaps which may expand the range of temperatures at which they are competitive with existing materials. For example, bulk Bi_2Te_3 is commonly used at room temperature because of its small band gap (150 meV at room temperature (Bejenari and Kantser 2008)) and low thermal conductivity (1.4 W/mK measured in a 1 μm film (Goncalves et al. 2010)); both of these are achievable with SiGeSn alloys, with the added advantage of compatibility with current Si-based nanoelectronics. While it is too early to fully evaluate the potential for SiGeSn to replace or improve existing TE materials, these recent findings will spur further research into the growth of bulk and nanostructured Group IV alloys containing Sn and their optimization for TE applications.

8   Conclusions

This chapter describes the lattice thermal conductivity in nanostructured SiGe alloys, including thin films, SLs, NWs, and NCs. The pBTE is solved with the full phonon dispersion and a momentum-dependent specularity model for boundary roughness scattering that effectively captures the interplay between intrinsic and interface scattering mechanisms. There's an order-of-magnitude reduction in thermal conductivity due to alloying, and further decrease is achieved in each of the nanostructures, with conductivity nearing the amorphous limit. Among Group-IV binary (SiSn and GeSn) and ternary (SiGeSn) alloys and their thin films, SiSn

has the lowest conductivity (3 W/mK) of all the bulk alloys, which reduces further to around 1 W/mK in 20-nm-thick SiSn films. Thermal conductivity is optimized in SLs by reducing the overall thickness of the superlattice period, increasing the roughness of the interfaces, and either introducing alloying into both layers of the superlattice or minimizing the thickness of the pure layer in silicon-alloy superlattices.

A Monte Carlo technique is used to compute the thermal conductivity in SiGe NWs. We find that thermal conductivity scales as $L^{1/3}$ over a broad range of lengths and the direct ballistic-to-diffusive crossover picture should be augmented by superdiffusion in the broad intermediate range of NW length from 10 nm to 10 μm. A Voronoi tessellation model was employed to treat the complex heterogeneous grain structure in NCs. The size scaling observed in NWs persists in nanocomposites and is insensitive to the variance in grain sizes. Alloy nanostructures exhibit novel heat dynamics and can be used as a unique fundamental platform to study the breakdown of Fourier's law and superdiffusive heat transport. The results demonstrate tunability of thermal transport through boundary and interface properties including structure size, roughness, and composition that are relevant to guiding the design of nanostructured SiGe alloys for thermoelectric applications.

References

Abeles B (1963) Lattice thermal conductivity of disordered semiconductor alloys at high temperatures. Phys Rev 131:1906–1911

Abeles B, Beers DS, Cody GD, Dismukes JP (1962) Thermal conductivity of Ge-Si alloys at high temperatures. Phys Rev 125(1):44–46

Aksamija Z (2014) Lattice thermal transport in si-based nanocomposites for thermoelectric applications. J Electron Mater 44:1644

Aksamija Z, Knezevic I (2010a) Anisotropy and boundary scattering in the lattice thermal conductivity of silicon nanomembranes. Phys Rev B 82:045319

Aksamija Z, Knezevic I (2010b) Thermoelectric properties of silicon nanostructures. J Comput Electron 9:173

Aksamija Z, Knezevic I (2013) Thermal conductivity of $Si_{1-x}Ge_x/Si_{1-y}Ge_y$ superlattices: competition between interfacial and internal scattering. Phys Rev B 88:155318

Aksamija Z, Knezevic I (2014) Lattice thermal transport in large-scale polycrystalline graphene. Phys Rev B 90:035419

Bartkowiak M, Mahan GD (1995) Nonlinear currents in Voronoi networks. Phys Rev B 51(16):10825–10832

Bejenari I, Kantser V (2008) Thermoelectric properties of bismuth telluride nanowires in the constant relaxation-time approximation. Phys Rev B 78(11):115322

Bera C, Soulier M, Navone C, Roux G, Simon J, Volz S, Mingo N (2010) Thermoelectric properties of nanostructured $Si_{1-x}Ge_x$ and potential for further improvement. J Appl Phys 108(12):124306

Boukai AI, Bunimovich Y, Tahir-Kheli J, Yu J, III WAG, Heath JR (2008) Silicon nanowires as efficient thermoelectric materials. Nature 451:168

Braginsky L, Lukzen N, Shklover V, Hofmann H (2002) High-temperature phonon thermal conductivity of nanostructures. Phys Rev B 66(13):134203

Cahill DG, Watson SK, Pohl RO (1992) Lower limit to the thermal conductivity of disordered crystals. Phys Rev B 46(10):6131–6140

Cahill DG, Ford WK, Goodson KE, Mahan GD, Majumdar A, Maris HJ, Merlin R, Phillipot SR (2003) Nanoscale thermal transport. J Appl Phys 93:793–818

Cahill DG, Braun PV, Chen G, Clarke DR, Fan S, Goodson KE, Keblinski P, King WP, Mahan GD, Majumdar A, Maris HJ, Phillpot SR, Pop E, Shi L (2014) Nanoscale thermal transport. II. 2003–2012. Appl Phys Rev 1(1):011305

Carruthers P (1961) Theory of thermal conductivity of solids at low temperatures. Rev Mod Phys 33(1):92

Cheaito R, Duda JC, Beechem TE, Hattar K, Ihlefeld JF, Medlin DL, Rodriguez MA, Campion MJ, Piekos ES, Hopkins PE (2012) Experimental investigation of size effects on the thermal conductivity of silicon-germanium alloy thin films. Phys Rev Lett 109:195901

Chen G (1998) Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices. Phys Rev B 57(23):14958–14973

Chen J, Zhang G, Li B (2009) Tunable thermal conductivity of $Si_{1-x}Ge_x$ nanowires. Appl Phys Lett 95(7):073117

Chen Y, Li D, Lukes JR, Majumdar A (2005) Monte Carlo simulation of silicon nanowire thermal conductivity. J Heat Transf 127(10):1129–1137

Chowdhury I, Prasher R, Lofgreen K, Chrysler G, Narasimhan S, Mahajan R, Koester D, Alley R, Venkatasubramanian R (2009) On-chip cooling by superlattice-based thin-film thermoelectrics. Nature Nano 4:235–238

Cipriani P, Denisov S, Politi A (2005) From anomalous energy diffusion to Levy walks and heat conductivity in one-dimensional systems. Phys Rev Lett 94:244301

Colombo L, Giannozzi P (1995) A first-principles derived parametrization of the adiabatic bond charge model. Solid State Commun 96:49

Denisov S, Klafter J, Urbakh M (2003) Dynamical heat channels. Phys Rev Lett 91:194301

Dhar A (2008) Heat transport in low-dimensional systems. Adv Phys 57(5):457–537

Dhar A, Saito K, Derrida B (2013) Exact solution of a Lévy walk model for anomalous heat transport. Phys Rev E 87:010103

DiSalvo FJ (1999) Thermoelectric cooling and power generation 285:703–706

Dismukes JP, Ekstrom L, Steigmeier EF, Kudman I, Beers DS (1964) Thermal and electrical properties of heavily doped ge-si alloys up to 1300[degree]K. J Appl Phys 35(10):2899–2907

Dresselhaus MS, Chen G, Tang MY, Yang RG, Lee H, Wang DZ, Ren ZF, Fleurial JP, Gogna P (2007) New directions for low-dimensional thermoelectric materials. Adv Mater 19:1043–1053

Esfarjani K, Chen G, Stokes HT (2011) Heat transport in silicon from first-principles calculations. Phys Rev B 84:085204

Fan Z, Wu Y, Zhao X, Lu Y (2004) Simulation of polycrystalline structure with Voronoi diagram in laguerre geometry based on random closed packing of spheres. Comput Mat Sci 29(3):301–308

Ferenc JS, Nda Z (2007) On the size distribution of Poisson Voronoi cells. Phys A Statis Mech Appl 385(2):518–526

Feser JP, Chan EM, Majumdar A, Segalman RA, Urban JJ (2013) Ultralow thermal conductivity in polycrystalline cdse thin films with controlled grain size. Nano Lett 13(5):2122–2127

Fischer IA, Wendav T, Augel L, Jitpakdeebodin S, Oliveira F, Benedetti A, Stefanov S, Chiussi S, Capellini G, Busch K et al (2015) Growth and characterization of SiGeSn quantum well photodiodes. Opt Express 23:25048

Fischetti MV, Laux SE (1996) Band structure, deformation potentials, and carrier mobility in strained si, ge, and sige alloys. J Appl Phys 80(4):2234–2252

Frachioni A, White BE (2012) Simulated thermal conductivity of silicon-based random multilayer thin films. J Appl Phys 112(1):014320

Gaiduk PI, Lundsgaard Hansen J, Nylandsted Larsen A, Bregolin FL, Skorupa W (2014) Suppression of tin precipitation in SiSn alloy layers by implanted carbon. Appl Phys Lett 104(23):231903

Garg J, Bonini N, Kozinsky B, Marzari N (2011) Role of disorder and anharmonicity in the thermal conductivity of silicon-germanium alloys: a first-principles study. Phys Rev Lett 106:045901

Gilat G (1972) Analysis of methods for calculating spectral properties in solids. J Comput Phys 10(3):432–465

Gilat G, Kam Z (1969) High-resolution method for calculating spectra of solids. Phys Rev Lett 22(14):715–717

Gilat G, Raubenheimer LJ (1966) Accurate numerical method for calculating frequency-distribution functions in solids. Phys Rev 144(2):390–395

Glassbrenner CJ, Slack GA (1964) Thermal conductivity of silicon and germanium from 3K to the melting point. Phys Rev 134:A1058–A1069

Goncalves LM, Couto C, Alpuim P, Rolo A, Völklein F, Correia JH (2010) Optimization of thermoelectric properties on Bi_2Te_3 thin films deposited by thermal co-evaporation. Thin Solid Films 518(10):2816–2821

Hicks L, Dresselhaus M (1993a) Effect of quantum-well structures on the thermoelectric figure of merit. Phys Rev B 47:12727

Hicks L, Dresselhaus M (1993b) Thermoelectric figure of merit of a one-dimensional conductor. Phys Rev B 47:16631

Hochbaum A, Chen R, Delgado R, Liang W, Garnett E, Najarian M, Majumdar A, Yang P (2008) Enhanced thermoelectric performance of rough silicon nanowires. Nature 451:163

Hori T, Shiomi J, Dames C (2015) Effective phonon mean free path in polycrystalline nanostructures. Appl Phys Lett 106(17):171901

Hsiao TK, Chang HK, Liou SC, Chu MW, Lee SC, Chang CW (2013) Observation of room-temperature ballistic thermal conduction persisting over $8.3\,\mu m$ in SiGe nanowires. Nat Nano 8(7):534–538

Huxtable ST, Abramson AR, Tien CL, Majumdar A, LaBounty C, Fan X, Zeng G, Bowers JE, Shakouri A, Croke ET (2002) Thermal conductivity of Si/SiGe and SiGe/SiGe superlattices. Appl Phys Lett 80(10):1737–1739

Jeng MS, Yang R, Song D, Chen G (2008) Modeling the thermal conductivity and phonon transport in nanoparticle composites using Monte Carlo simulation. J Heat Transf 130:042410

Joshi G, Lee H, Lan Y, Wang X, Zhu G, Wang D, Gould RW, Cuff DC, Tang MY, Dresselhaus MS, Chen G, Ren Z (2008) Enhanced thermoelectric figure-of-merit in nanostructured p-type silicon germanium bulk alloys. Nano Lett 8(12):4670–4674

Kazan M, Volz S (2014) Calculation of the lattice thermal conductivity in granular crystals. J Appl Phys 115(7):073509

Khatami SN, Aksamija Z (2016) Lattice thermal conductivity of the binary and ternary group-iv alloys Si-Sn, Ge-Sn, and Si-Ge-Sn. Phys Rev Appl 6:014015

Khitun A, Balandin A, Wang KL (1999) Modification of the lattice thermal conductivity in silicon quantum wires due to spatial confinement of acoustic phonons. Superlattice Microst 26(3):181

Khitun A, Balandin A, Liu JL, Wang KL (2001) The effect of the long-range order in a quantum dot array on the in-plane lattice thermal conductivity. Superlattice Microst 30(1):1–8

Kim H, Kim I, jin Choi H, Kim W (2010) Thermal conductivities of $si_{1-x}ge_x$ nanowires with different germanium concentrations and diameters. Appl Phys Lett 96:233106

Klemens PG (1958) Thermal conductivity and lattice vibrational modes. In: Seitz F, Turnbull D (eds) Solid state physics, vol 7. Academic Press, New York, pp 1–98

Klemens PG (1960) Thermal resistance due to point defects at high temperatures. Phys Rev 119(2):507–509

Kouvetakis J, Menendaz J, Chizmeshya AVG (2006) Tin-based group IV semiconductors: new platforms for opto- and microelectronic on silicon. Annu Rev Mater Res 36:497–554

Kurosawa M, Kato M, Yamaha T, Taoka N, Nakatsuka O, Zaima S (2015) Near-infrared light absorption by polycrystalline SiSn alloys grown on insulating layers. Appl Phys Lett 106(17):171908

Kurosawa M, Kato M, Takahashi K, Nakatsuka O, Zaima S (2017) Self-organized lattice-matched epitaxy of $Si_{1-x}Sn_x$ alloys on (001)-oriented si, ge, and inp substrates. Appl Phys Lett 111:192106

Lacroix D, Joulain K, Terris D, Lemonnier D (2006) Monte carlo simulation of phonon confinement in silicon nanostructures: application to the determination of the thermal conductivity of silicon nanowires. Appl Phys Lett 89(10):103104

Lan Y, Poudel B, Ma Y, Wang D, Dresselhaus MS, Chen G, Ren Z (2009) Structure study of bulk nanograined thermoelectric bismuth antimony telluride. Nano Lett 9:1419–1422

Lan Y, Minnich AJ, Chen G, Ren Z (2010) Enhancement of thermoelectric Figure-of-Merit by a bulk nanostructuring approach. Adv Func Mater 20(3):357–376

Larkin JM, McGaughey AJH (2013) Predicting alloy vibrational mode properties using lattice dynamics calculations, molecular dynamics simulations, and the virtual crystal approximation. J Appl Phys 114(2):023507

Lee EK, Yin L, Lee Y, Lee JW, Lee SJ, Lee J, Cha SN, Whang D, Hwang GS, Hippalgaonkar K, Majumdar A, Yu C, Choi BL, Kim JM, Kim K (2012) Large thermoelectric figure-of-merits from SiGe nanowires by simultaneously measuring electrical and thermal transport properties. Nano Lett 12:2918–2923

Lee ML, Venkatasubramanian R (2008) Effect of nanodot areal density and period on thermal conductivity in SiGe/Si nanodot superlattices. Appl Phys Lett 92(5):053112

Lee SM, Cahill DG, Venkatasubramanian R (1997) Thermal conductivity of Si-Ge superlattices. Appl Phys Lett 70(22):2957–2959

Lepri S, Livi R, Politi A (2003) Thermal conduction in classical low-dimensional lattices. Phys Reports 377(1):1–80

Li D, Wu Y, Kim P, Shi L, Yang P, Majumdar A (2003) Thermal conductivity of individual silicon nanowires. Appl Phys Lett 83:2934–2936

Li Y, Liu S, Li N, Hnggi P, Li B (2015) 1d momentum-conserving systems: the conundrum of anomalous versus normal heat transport. New J Phys 17(4):043064

Liao CN, Chen C, Tu KN (1999) Thermoelectric characterization of si thin films in silicon-on-insulator wafers. J Appl Phys 86(6):3204–3208

Liu S, Hänggi P, Li N, Ren J, Li B (2014) Anomalous heat diffusion. Phys Rev Lett 112:040601

Liu W, Asheghi M (2005) Thermal conduction in ultrathin pure and doped single-crystal silicon layers at high temperatures. J Appl Phys 98(12):123523

Liu W, Balandin AA (2005) Thermal conduction in alxga1xn alloys and thin films. J Appl Phys 97:073710

Liu W, Borca-Tasciuc T, Chen G, Liu J, Wang K (2001) Anisotropic thermal conductivity of ge quantum-dot and symmetrically strained si/ge superlattices. J Nanosci Nanotechnol 1: 39–42(4)

Liu W, Yan X, Chen G, Ren Z (2012) Recent advances in thermoelectric nanocomposites. Nano Energy 1(1):42–56

Mai T, Dhar A, Narayan O (2007) Equilibration and universal heat conduction in fermi-pasta-ulam chains. Phys Rev Lett 98:184301

Majumdar A (2004) Thermoelectricity in semiconductor nanostructures. Science 303(5659): 777–778

Maris HJ (1990) Phonon propagation with isotope scattering and spontaneous anharmonic decay. Phys Rev B 41(14):9736–9743

Martin P, Aksamija Z, Pop E, Ravaioli U (2009) Impact of phonon-surface roughness scattering on thermal conductivity of thin Si nanowires. Phys Rev Lett 102(12):125503

Maurer LN, Aksamija Z, Ramayya EB, Davoody AH, Knezevic I (2015) Universal features of phonon transport in nanowires with correlated surface roughness. Appl Phys Lett 106(13):133108

Maycock P (1967) Thermal conductivity of silicon, germanium, III-V compounds and III-V alloys 10:161–168

Mazumder S, Majumdar A (2001) Monte carlo study of phonon transport in solid thin films including dispersion and polarization. J Heat Transf 123(4):749

McGaughey AJH, Jain A (2012) Nanostructure thermal conductivity prediction by Monte Carlo sampling of phonon free paths. Appl Phys Lett 100(6):061911

Min KS, Atwater HA (1998) Ultrathin pseudomorphic Sn/Si and $Sn_x Si_{1-x}$/Si heterostructures. Appl Phys Lett 72(15):1884–1886

Minnich AJ, Lee H, Wang XW, Joshi G, Dresselhaus MS, Ren ZF, Chen G, Vashaee D (2009) Modeling study of thermoelectric SiGe nanocomposites. Phys Rev B 80(15):155327

Moontragoon P, Soref RA, Ikonic Z (2012) The direct and indirect bandgaps of unstrained $Si_xGe_{1-x-y}Sn_y$ and their photonic device applications. J Appl Phys 112(7):073106

Morelli DT, Heremans JP, Slack GA (2002) Estimation of the isotope effect on the lattice thermal conductivity of group IV and group III-V semiconductors. Phys Rev B 66(19):195304

Neophytou N, Zianni X, Kosina H, Frabboni S, Lorenzi B, Narducci D (2013) Simultaneous increase in electrical conductivity and seebeck coefficient in highly boron-doped nanocrystalline si. Nanotechnology 24(20):205402

Péraud JPM, Hadjiconstantinou NG (2011) Efficient simulation of multidimensional phonon transport using energy-based variance-reduced monte carlo formulations. Phys Rev B 84:205331

Pop E (2010) Energy dissipation and transport in nanoscale devices. Nano Res 3(3):147–169

Poudel B, Hao Q, Ma Y, Lan Y, Minnich A, Yu B, Yan X, Wang D, Muto A, Vashaee D, Chen X, Liu J, Dresselhaus MS, Chen G, Ren Z (2008) High-thermoelectric performance of nanostructured bismuth antimony telluride bulk alloys. Science 320(5876):634–638

Priolo A, Jaeger HM, Dammers AJ, Radelaar S (1992) Conductance of two-dimensional disordered Voronoi networks. Phys Rev B 46(22):14889–14892

Rajput BD, Browne DA (1996) Lattice dynamics of II-VI materials using the adiabatic bond-charge model. Phys Rev B 53(14):9052–9058

Ramayya EB, Maurer LN, Davoody AH, Knezevic I (2012) Thermoelectric properties of ultrathin silicon nanowires. Phys Rev B 86:115328

Randrianalisoa J, Baillis D (2008) Monte Carlo simulation of steady–state microscale phonon heat transport. J Heat Transf 130:072404

Rieger MM, Vogl P (1993) Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates. Phys Rev B 48(19):14276–14287

Rustagi KC, Weber W (1976) Adiabatic bond charge model for the phonons in A^3B^5 semiconductors. Solid State Commun 18(6):673–675

Ryu HJ, Aksamija Z, Paskiewicz DM, Scott SA, Lagally MG, Knezevic I, Eriksson MA (2010) Quantitative determination of contributions to the thermoelectric power factor in si nanostructures. Phys Rev Lett 105:256601

Saito K, Dhar A (2010) Heat conduction in a three dimensional anharmonic crystal. Phys Rev Lett 104:040601

Sellan DP, Turney JE, McGaughey AJH, Amon CH (2010) Cross-plane phonon transport in thin films. J Appl Phys 108(11):113524

Shi L, Yao D, Zhang G, Li B (2009) Size dependent thermoelectric properties of silicon nanowires. Appl Phys Lett 95(6):063102

Shiomi J (2016) Research update: phonon engineering of nanocrystalline silicon thermoelectrics. APL Mater 4(10):104504

Shlesinger MF, Zaslavsky GM, Klafter J (1993) Strange kinetics. Nature 363:31–37

Simkin MV, Mahan GD (2000) Minimum thermal conductivity of superlattices. Phys Rev Lett 84(5):927–930

Slack G (1979) Solid state physics, vol 34. Academic Press, New York

Snyder GJ, Toberer ES (2008) Complex thermoelectric materials. Nature Matter 7:105–114

Soffer SB (1967) Statistical model for the size effect in electrical conduction. J Appl Phys 38(4):1710–1715

Soref R, Kouvetakis J, Tolle J, Menendez J, D'Costa V (2007) Advances in SiGeSn technology. J Mat Res 22:3281–3291

Strauch D, Dorner B (1990) Phonon dispersion in GaAs. J Phys Condens Matter 2:1457–1474

Tamura SI (1983) Isotope scattering of dispersive phonons in Ge. Phys Rev B 27(2):858–866

Thumfart L, Carrete J, Vermeersch B, Ye N, Truglas T, Feser J, Groiss H, Mingo N, Rastelli A (2017) Thermal transport through Ge-rich Ge/Si superlattices grown on Ge(001). J Phys D Appl Phys 51(1):014001

Tian Z (2011) On the importance of optical phonons to thermal conductivity in nanostructures. Appl Phys Lett 99(5):053122

Tonkikh AA, Zakharov ND, Eisenschmidt C, Leipner HS, Werner P (2014) Aperiodic SiSn/Si multilayers for thermoelectric applications. J Crys Growth 392:49–51

Turney JE, McGaughey AJH, Amon CH (2010) In-plane phonon transport in thin films. J Appl Phys 107(2):024317

Uchida NI, Maeda T, Lieten RR, Okajima S, Ohishi Y, Takase R, Ishimaru M, Locquet JP (2015) Carrier and heat transport properties of polycrystalline GeSn films on SiO_2. Appl Phys Lett 107(23):232105

Upadhyaya M, Aksamija Z (2016) Non-diffusive lattice thermal transport in Si-Ge alloy nanowires. Phys Rev B 94(17):174303

Upadhyaya M, Khatami SN, Aksamija Z (2015) Engineering thermal transport in SiGe-based nanostructures for thermoelectric applications. J Mater Res 30:2649–2662

Venkatasubramanian R, Siivola E, Colpitts T, O'Quinn B (2001) Thin-film thermoelectric devices with high room-temperature figures of merit. Nature 413:597

Vermeersch B, Carrete J, Mingo N, Shakouri A (2015) Superdiffusive heat conduction in semiconductor alloys. I. Theoretical foundations. Phys Rev B 91:085202

Vermeersch B, Carrete J, Mingo N (2016) Cross-plane heat conduction in thin films with ab-initio phonon dispersions and scattering rates. Appl Phys Lett 108(19):193104

Vining CB (2009) An inconvenient truth about thermoelectrics. Nat Matter 8(2):83–85

Wang XW, Lee H, Lan YC, Zhu GH, Joshi G, Wang DZ, Yang J, Muto AJ, Tang MY, Klatsky J, Song S, Dresselhaus MS, Chen G, Ren ZF (2008) Enhanced thermoelectric figure of merit in nanostructured n-type silicon germanium bulk alloy. Appl Phys Lett 93(19):193121

Wang Z, Mingo N (2010) Diameter dependence of sige nanowire thermal conductivity. Appl Phys Lett 97(10):101903

Wang Z, Mingo N (2011) Absence of Casimir regime in two-dimensional nanoribbon phonon conduction. Appl Phys Lett 99(10):101903

Ward A, Broido DA (2010) Intrinsic phonon relaxation times from first-principles studies of the thermal conductivities of Si and Ge. Phys Rev B 81(8):085205

Weber W (1974) New bond-charge model for the lattice dynamics of diamond-type semiconductors. Phys Rev Lett 33(6):371–373

Weber W (1977) Adiabatic bond charge model for the phonons in diamond, si, ge, and a-sn. Phys Rev B 15(10):4789

Xie G, Guo Y, Wei X, Zhang K, Sun L, Zhong J, Zhang G, Zhang YW (2014) Phonon mean free path spectrum and thermal conductivity for si1xgex nanowires. Appl Phys Lett 104(23):233901

Yang B, Chen G (2001) Lattice dynamics study of anisotropic heat conduction in superlattices. Microscale Thermophys Eng 5(2):107–116

Yang N, Zhang G, Li B (2010) Violation of Fourier's law and anomalous heat diffusion in silicon nanowires. Nano Today 5:85–90

Yin L, Kyung Lee E, Woon Lee J, Whang D, Lyong Choi B, Yu C (2012) The influence of phonon scatterings on the thermal conductivity of SiGe nanowires. Appl Phys Lett 101(4):043114

Zaburdaev V, Denisov S, Klafter J (2015) Lévy walks. Rev Mod Phys 87:483–530

Zebarjadi M, Esfarjani K, Bian Z, Shakouri A (2011) Low-temperature thermoelectric power factor enhancement by controlling nanoparticle size distribution. Nano Lett 11:225–230

Zhu GH, Lee H, Lan YC, Wang XW, Joshi G, Wang DZ, Yang J, Vashaee D, Guilbert H, Pillitteri A, Dresselhaus MS, Chen G, Ren ZF (2009) Increased phonon scattering by nanograins and point defects in nanostructured silicon with a low concentration of germanium. Phys Rev Lett 102(19):196803

Zhu Z, Xiao J, Sun H, Hua Y, Cao R, Wanga Y, Zhao L, Zhuang J (2015) Composition-dependent band gaps and indirect-direct band gap transitions of group-IV semiconductor alloys. Phys Chem Chem Phys 17(33):21605–21610

Resonant Thermal Transport in Nanophononic Metamaterials

40

Mahmoud I. Hussein and Hossein Honarvar

Contents

Abstract

Thermoelectric materials allow heat to be converted directly into electricity or, conversely, to be absorbed or rejected when electrical power is provided. A key requirement for a thermoelectric material is to exhibit a low thermal conductivity while simultaneously possessing a high power factor, which is proportional to the electrical conductivity. The challenge, however, is that these two property targets are coupled and are at odds with each other in common materials. This natural

M. I. Hussein (✉) · H. Honarvar
Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado
Boulder, Boulder, CO, USA
e-mail: mih@colorado.edu; honarvar@colorado.edu

constraint has limited the performance of thermoelectric materials and impeded their industrial proliferation. In this chapter, we present a recently proposed material concept, termed *nanophononic metamaterial*, which overcomes this historical constraint. One promising configuration of a nanophononic metamaterial consists of a silicon membrane with inherently attached local resonators in the form of nanopillars distributed on the surface. The nanopillar local resonances, or *vibrons*, hybridize with the underlying phonon waves carrying the heat in the membrane, which leads to significant reductions in the phonon group velocities and to mode localizations within the nanopillars. These two wave-based phenomena, supplemented by reductions in phonon lifetimes, together cause a reduction in the lattice thermal conductivity along the base membrane. Since the nanopillars are located external to the main body of the membrane, changes to the electronic band structure and electron scattering are both minimized – thus a negligible effect on the generation and flow of electrons is expected. This novel nanostructure-induced mechanism therefore has all the ingredients to enable thermoelectric energy conversion at record high performance while using a low-cost and practical base material such as silicon. This chapter includes lattice dynamics- and molecular dynamics-based analyses of the underlying physical processes stemming from the presence of the local resonances and discusses how the nature of *resonant thermal transport* is shaped particularly by the wave-based mechanism of phonon-vibron coupling.

1 Introduction

Thermoelectric materials (TE) convert heat into electricity or vice versa through a solid-state process. For the conversion efficiency to be competitive with fluid-based technologies, a TE material must be a good insulator of heat while, simultaneously, exhibiting good electrical properties – a combination that is hard to find in common materials. Here we present the concept of a locally resonant nanophononic metamaterial (NPM) (Davis and Hussein 2014; Hussein and Davis 2016; Honarvar and Hussein 2016, 2018; Honarvar et al. 2016) to overcome this natural properties trade-off. One realization of an NPM is a freestanding silicon (Si) membrane (thin film) with a periodic array of nanoscale pillars erected on one or both free surfaces. Heat is transported along the membrane portion of this nanostructured material as a succession of propagating vibrational waves, *phonons*. The atoms making up the minuscule pillars on their part generate stationary vibrational waves, which we describe as *vibrons*. These two types of waves interact causing a mode coupling for each pair which appears as an avoided crossing in the pillared membrane's phonon band structure. This in turn (1) reduces the base membrane phonon group velocities, as well as (2) enables the generation of new modes localized in the nanopillar portion(s) in physical space; it also (3) drops the phonon lifetimes at and around the phonon-vibron coupling regions in the phonon band structure. These three effects in tandem cause a reduction in the in-plane lattice thermal conductivity of the underlying membrane.

Given that the number of vibrons scales with the number of degrees of freedom of a nanopillar, the three effects intensify as the size of the nanopillar(s) increases and in principle may be tuned to influence the entire phonon spectrum (which for silicon extends up to over 17 THz). This novel *full-spectrum* phenomenon thus provides an opportunity for achieving exceptionally strong reductions in the thermal conductivity. Furthermore, since the mechanisms concerned with the generation and carrying of electrical charge are practically independent of the phonon-vibron couplings, the Seebeck coefficient and the electrical conductivity are at most only mildly affected, if not at all. In this chapter, we introduce the concept of an NPM and resonant thermal transport, provide a lattice dynamics (LD) analysis highlighting the impact of local resonances on the phonon band structure, and present thermal conductivity predictions using molecular dynamics (MD) simulations. Finally, projections of record-breaking values of the TE energy conversion figure of merit ZT are provided.

2   Classical Macroscopic Metamaterials for Electromagnetic and Acoustic/Elastic Waves

Metamaterials are artificially structured materials that possess properties generally viewed to exceed what we expect to find in naturally occurring materials. The concept was first introduced in electromagnetics and optics (Pendry et al. 1999; Smith et al. 2000) and shortly afterward in acoustics (Liu et al. 2000). Examples of such unusual properties include subwavelength band gaps and negative refraction. In both electromagnetic and acoustic/elastic metamaterials, the feature of interest appears in the subwavelength regime and in the context of extreme effective properties. A form of local resonance is usually associated with this behavior. The local resonance is introduced by embedding or attaching resonating substructures into/to the host or base material medium (Hussein et al. 2014). At a resonance frequency, a coupling takes place between the resonating substructure and a traveling wave mode associated with the underlying medium. This coupling creates an avoided crossing in the electromagnetic or acoustic/elastic band structure of the overall system, which may result in the opening of a band gap and the creation of negative (or generally extreme) effective properties at and around the resonance frequency.

An example of a classical macroscopic metamaterial, proposed in a seminal paper by Liu et al. (2000), consisted of a relatively light and compliant matrix material, namely, epoxy, hosting heavy inclusions coated with a highly compliant material, namely, rubber-coated lead spheres. In this acoustic metamaterial, the lead spheres acted as the resonating masses, whereas the rubber coating served the role of supporting springs. An example of an elastic metamaterial was presented by Pennec et al. (2008) and Wu et al. (2008) in the form of a plate with pillars erected on one surface. The thickness of the plate and the size of the pillars were selected to be on the order of a few millimeters, which results in local resonances and hybridization band gaps at frequencies as low as a few hundred Hertz. In both sets of examples, the coated inclusions and the pillared plates, the resonating components

were placed in a spatially periodic fashion. While periodicity is advantageous in terms of compactness and ease of analysis and characterization, it is not necessary in a metamaterial (Achaoui et al. 2013).

3 Nanophononic Metamaterials for Thermal Transport

Similar to electromagnetic and acoustic/elastic metamaterials, an NPM also features resonating substructures, except as implied by the terminology; these substructures are of nanoscale size which makes their atomic structure directly relevant to the emerging resonance characteristics (Honarvar and Hussein 2018). Moreover, in classical metamaterials, typically the interest is in a single low-frequency resonance in the kHz or MHz regime. In an NPM, on the other hand, the entire set of resonances is of interest due to the nature of nanoscale thermal transport which in principle is formed from the collective contributions of all phonon modes. In a semiconducting material, the resonance spectrum spans an extremely broad frequency range, for example, reaching over 17 THz for silicon. The number of local resonances is equal to the number of atoms comprising a nanoresonator multiplied by three, which is the number of degrees of freedom for the motion of an atom. Thus for a resonating substructure on the order of a few tens of nanometers in size, the number of local resonances could be in the hundreds of thousands or millions.

3.1 Nanopillared Membranes for Thermoelectric Energy Conversion

In TE materials, the underlying process of power generation from temperature differences is referred to as the Seebeck effect, while the process of heat pumping or refrigeration driven by an electrical current is known as the Peltier effect (Vineis et al. 2010; Chen et al. 2013). Since no moving mechanical components, fluid systems, or chemical reactions are involved, TE materials provide good reliability and stability. On the other hand, their low efficiency is a significant obstacle that impedes their wide use for large-scale industrial and domestic energy conversion. The performance of a TE material is based on a well-defined figure of merit, $ZT = (S^2\sigma/k)T$, where S is the Seebeck coefficient (representing the amount of voltage generated per unit temperature gradient), σ is the electrical conductivity, k is the total thermal conductivity (consisting of a lattice component k_l and an electronic component k_e), and T is the average temperature between two material junctions (Rowe 2005).

A key requirement for a TE material is to exhibit a low thermal conductivity while simultaneously possessing a high Seebeck coefficient and a high electrical conductivity. The challenge, however, is that these two sets of requirements – the thermal and the electrical – are tightly coupled and cannot be both met, especially in industrial materials like silicon. At present, commercially available TE devices exhibit ZT values near 1; yet a value of 3 or higher is needed to enable industrial-
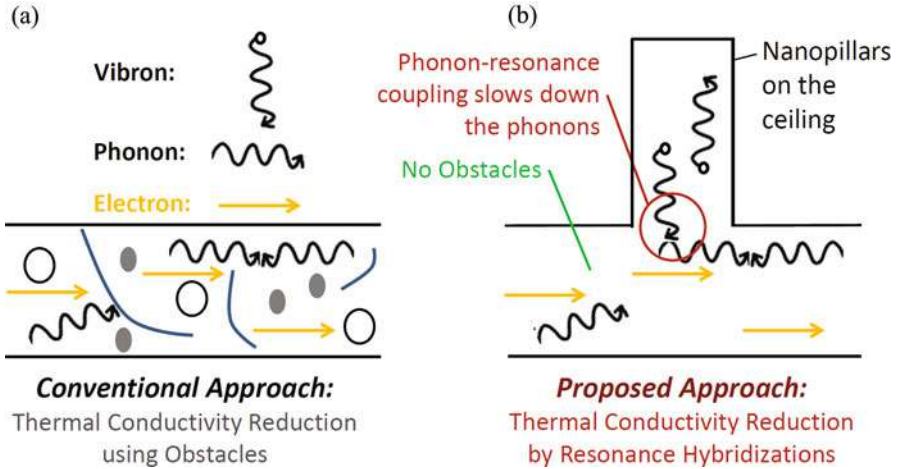
Fig. 1 Illustration of thermal conductivity reduction in a membrane by (**a**) scattering (using obstacles; conventional approach) versus by (**b**) the resonance hybridizations mechanism (without obstacles; proposed approach)

scale deployment of thermoelectrics for waste heat conversion to electricity and for cooling and refrigeration (Vineis et al. 2010). Over the last two decades, an overriding philosophy for increasing ZT has been to introduce small features (such as holes, particles, and/or interfaces) within the internal domain of the material in order to scatter the heat-carrying phonons and consequently reduce the thermal conductivity (Vineis et al. 2010; Snyder and Toberer 2008) (Fig. 1a). This strategy, however, still suffers from the properties trade-off mentioned above since the scatterers are also likely to affect charge generation and impede the transfer of electrons and thus negate any possibility of substantial increase in ZT.

The historical performance limitations of TE materials are potentially overcome via an NPM configuration that is based on a thin membrane of silicon with a periodic array of closely packed nanopillars standing on its surface (i.e., exterior to the membrane cross section) (Davis and Hussein 2014; Hussein and Davis 2016; Honarvar and Hussein 2016, 2018; Honarvar et al. 2016). A large portion of the heat transported in this nanostructured material is carried by phonons, a collection of propagating vibrational waves. The atoms making up the nanoscale pillars on their part generate stationary vibrational waves. These two types of waves, the traveling and the standing, interact and cause a substantial portion of the energy of the heat-carrying phonons to divert into the nanopillars and remain localized there (Fig. 1b). In addition, the coupling between the resonances and the underlying heat-carrying phonons traveling in-plane in the membrane causes the phonon group velocities to drop significantly. Furthermore, the nanopillars reduce the phonon lifetimes at and around the hybridization zones that feature in the phonon band structure. These three effects cause a reduction in the lattice thermal conductivity along the membrane. This novel mechanism of phonon-vibron coupling via nanostructuring is practically

independent of the mechanisms of generation and transport of electrical charge and therefore does not significantly affect the Seebeck coefficient and electrical conductivity, respectively.

3.2 Vibrons: Standing Phonons

In an NPM, the substructure resonances are denoted as vibrons, which unlike phonons are wavenumber-independent because they are localized standing waves. The frequency spectrum of vibrons made available by a resonating nanosubstructure may be designed to couple with heat-carrying phonon modes belonging to all or most of the dispersion branches across the full spectrum of the host medium. This atomic-scale coupling mechanism gives rise to a *resonance hybridization* between pairs of vibrons and wavenumber-dependent phonons of the host medium.

This phenomenon has been demonstrated theoretically using LD calculations and a Boltzmann transport model on a system based on a 3-nm-thick suspended Si membrane with Si nanopillars, yielding a factor of 2 reduction in the thermal conductivity due to the presence of the nanopillars (Davis and Hussein 2014). This discovery was the topic of a focus article in *Physics* (Monroe 2014) and, more recently, has been described in a German online book on renewable energy and energy efficiency as one of the recent ideas in thermoelectrics that "is radical and potentially overcomes the main problem that has limited the development of the field since its birth in 1821" (Khammas 2007–2018). In Fig. 2, we consider a freestanding Si membrane but use gallium nitride (GaN) for the nanopillars as another example of an NPM (modeling details are provided in the next subsection). As shown in Fig. 2a, avoided crossings between horizontal resonance modes (i.e., the vibrons) and the underlying membrane dispersion branches take place. In Fig. 2b, we clearly observe how an avoided crossing causes mode localization. We note that the nanopillar-membrane coupling in this multi-material structure is sufficiently strong to still generate these avoided crossings which lead to a reduction the in-plane thermal conductivity along the membrane.

3.3 Frequency Limits for "Active" Resonance Hybridization

Thermal transport in silicon is carried by very short wavelength waves which puts low-frequency limits on the nanopillar generated vibrons that couple with the phonons. To investigate these limits, we switch back to a simpler nanopillared membrane configuration where both the membrane and nanopillar portions are made out of silicon. We use this model throughout the rest of the chapter. Two NPM configurations are considered, namely, a membrane with nanopillars (1) on a single surface and (2) on each of the two surfaces. We only examine base membrane and nanopillars formed from defect-free single-crystal silicon. Figure 3 displays the unit cells of these two NPM configurations as well as the structure of a conventional cell (CC) and a unit cell of a uniform (unpillared) membrane with the same thickness as
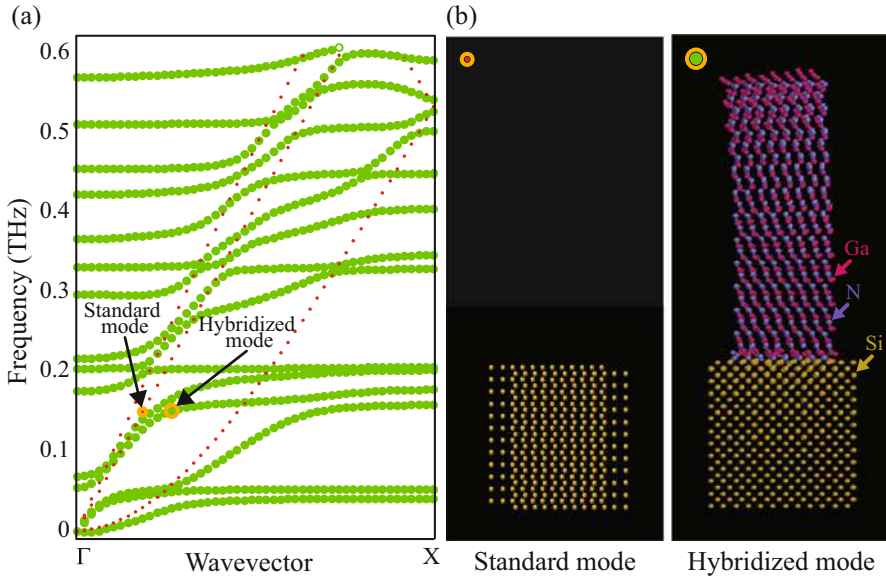
Fig. 2 Illustration of the resonance hybridization phenomenon from a lattice dynamics perspective. (**a**) Phonon band structure of a Si membrane with GaN pillars standing on one surface (green) versus that of a uniform Si membrane with the same thickness (red). (**b**) Atomic displacements for a standard heat-carrying phonon indicating activity within the Si membrane portion. (**c**) Atomic displacements of a hybridized mode revealing localized pillar motion and almost "thermal silence" in the Si membrane portion. A small model (3264 atoms in the unit cell) is displayed for clarity

Fig. 3 Atomic-scale unit-cell models. (**a**) Conventional 8-atom unit cell for silicon and unit cells for a (**b**) uniform membrane, (**c**) single-pillared NPM, and (**d**) double-pillared NPM. In (**b**), (**c**), and (**d**), the top and bottom surfaces are free
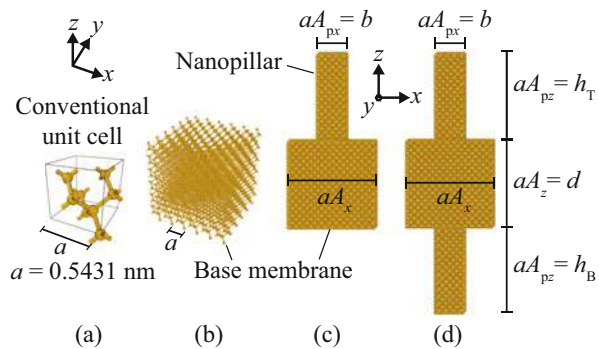
the thickness of the base membrane in the NPM. The geometry of a membrane with nanopillars on each surface is represented as $aA_x \times aA_y \times d + b \times b \times h_T + b \times b \times h_B$ which may be converted to CC by dividing each dimension by a. This same representation is used for an unpillared surface by simply dropping each of the last two terms. Figure 3 provides a pictorial definition of all geometric parameters. In all our analyses, room temperature at $T = 300\,\text{K}$ is assumed, and the Stillinger-Weber

Fig. 4 Phonon band
structure of a
$6 \times 6 \times 6 + 4 \times 4 \times 12$ CC
NPM unit cell (right) and a
corresponding unpillared
membrane modeled as a
$1 \times 1 \times 6$ CC unit cell (left).
Frequency-wavenumber
intercepts are marked to
illustrate the lower frequency
limits for "active" vibrons
considering phonons with
wavelengths ranging from
0.54 to 10.86 nm
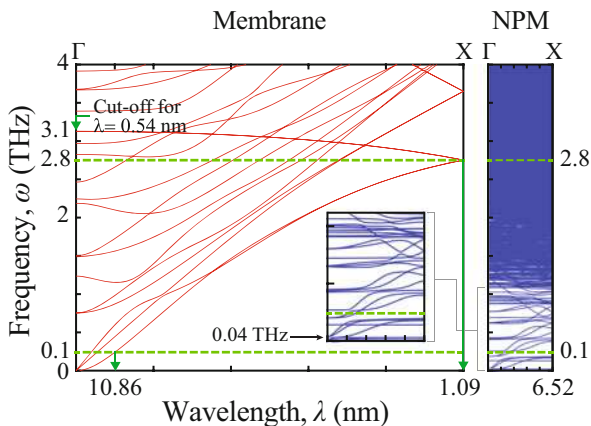
empirical potential is used to represent the interactions between neighboring atoms
(Stillinger and Weber 1985).

In bulk silicon, thermal transport is carried by phonons with wavelengths ranging,
roughly, from 0.4 to 10 nm at room temperature with the majority of the distribution
being between 0.4 and 2.2 nm (Esfarjani et al. 2011). In Fig. 4, we show the phonon
band structure for an unpillared membrane unit cell of size $1 \times 1 \times 6$ CC (i.e., 0.543-
nm wide) and a nanopillared membrane unit cell of size $6 \times 6 \times 6 + 4 \times 4 \times 12$
CC (i.e., 3.259-nm wide). In principle, the Brillouin zone of this NPM unit cell
may be unfolded six times and mapped onto the unit cell of the uniform membrane
to enable direct comparison. In the figure, we consider the wavelengths of $\lambda =$
0.54, 1.09, and 10.86 nm and mark the frequency where each of these wavelengths
intersects the lowest dispersion curve of the uniform membrane. The intercepts
correspond to frequencies $\omega = 3.1$, 2.8, and 0.1 THz, respectively. These values
represent the lower frequency limits for vibrons that are capable of coupling with
phonons at each of these wavelengths and lower. For example, any local resonance
that is 2.8 THz or higher is, in principle, available for coupling with phonons with a
wavelength of 1.09 nm or lower. Furthermore, we observe from an extended version
of Fig. 4 that the longitudinal acoustic branch reaches nearly 7 THz upon unfolding
once. Thus the numerous vibrons spanning the 3–7 THz range are available for
coupling and hybridization with a wide selection of phonons with wavelengths
between 0.5 and 1 nm, and many more vibrons are active when also considering
phonons with wavelengths larger than 1 nm. From the inset of Fig. 4, we observe
that the vibrons for this example NPM structure start at 0.04 THz and populate as we
move upward in the frequency domain; therefore there is an abundance of vibrons
available for resonance hybridizations to take place within the full range of relevant
phonon wavelengths.

The other critical size-dependent factor for the NPM concept is the distribution
of the phonon mean free path (MFP). The MFP distribution for bulk silicon
at $T = 300$ K ranges from a few nanometers to a few microns (Esfarjani

et al. 2011). For Si membranes with a thickness on the order of a few tens or hundreds of nanometers, the room-temperature MFP distribution was shown in recent experimental investigations to comfortably cover a range that is at least on the order of the membrane thickness; see, for example, Neogi et al. (2015), Johnson et al. (2013), and Minnich (2012). The active wavelengths (which as discussed above can be as small as a few angstroms) are able to travel relatively long distances within the nanopillared membrane structure as long as these distances fall within the range of the MFP distribution. In order to directly account for the effects of phonon-phonon and boundary scattering, MD simulations are needed. In MD simulations, any limitations on the dynamical processes involving the nanostructure phonon and vibron modes are inherently incorporated in the results. The reader is referred to Honarvar and Hussein (2016) where direct evidence of the existence of the resonance hybridization phenomenon in room-temperature MD simulations was demonstrated using spectral energy density (SED) calculations (Thomas et al. 2010; Larkin et al. 2014).

4   Thermal Conductivity Prediction by Molecular Dynamics Simulations

4.1   Simulation Parameters and Analysis Tools

To incorporate the effects of anharmonic phonon-phonon interactions and to predict the lattice thermal conductivity, we resort to equilibrium MD (EMD) simulations. In these simulations, the computational domain consists of one or more of the unit cells shown in Fig. 3, and standard periodic boundary conditions are applied at the in-plane boundaries. The MD results are post-processed using the Green-Kubo (GK) formulation to predict the thermal conductivities (Zwanzig 1965; Ladd et al. 1986; McGaughey and Kaviany 2004; Landry et al. 2008). Classical MD simulations like the ones we conduct here are valid at room temperature where quantum effects are negligible. The GK approach follows the linear response theorem and is based on the dissipation of equilibrium fluctuations for the heat current vector, $\boldsymbol{J}$. Defining the time average of the heat current autocorrelation function (HCACF) for a generally anisotropic material as $\langle \boldsymbol{J}(0) \otimes \boldsymbol{J}(t) \rangle$, where $\otimes$ denotes the tensor product, the thermal conductivity tensor is expressed as

$$\boldsymbol{k} = \frac{1}{k_B V T^2} \int \langle \boldsymbol{J}(0) \otimes \boldsymbol{J}(t) \rangle \mathrm{d}t, \qquad (1)$$

where k_B is the Boltzmann constant, V is the system volume, and t is time.

In order to examine whether the resonance hybridization mechanism does indeed exist and unfolds within the MD simulations (and as suggested by the quasi-harmonic phonon band diagrams), we compute the spectral energy density (Thomas et al. 2010; Larkin et al. 2014), which is a quantity obtained directly from the simulations. There are two SED formulations reported in the literature for phonon

transport problems. In one SED expression, referred to as Φ, the MD atom velocities are projected onto the phonon normal modes of the constituent unit cell, which are obtained separately from phonon band structure calculations. This approach allows for an accurate prediction of both phonon frequencies and lifetimes (Larkin et al. 2014). In an alternative formulation, the SED expression requires knowledge of only the crystal unit-cell structure and does not require any *a priori* knowledge of the phonon mode eigenvectors. This alternative SED expression, referred to as Φ', accurately predicts only the phonon frequencies and not the lifetimes (Larkin et al. 2014). In seeking to obtain evidence for the existence of local resonances and phonon-vibron couplings within an MD simulation, we intentionally seek an SED technique that enables us to extract the frequency-wave vector spectrum directly from the simulations without any prior knowledge of the phonon band structure. Thus the Φ' formulation is perfectly suited for this aim.

As provided by Thomas et al. (2010), the SED expression, Φ', is a function of wave vector, $\boldsymbol{\kappa}$, and frequency, ω, and for an all-silicon system is given by

$$\Phi'(\boldsymbol{\kappa}, \omega) = \mu_0 \sum_{\alpha}^{3} \sum_{b}^{n} \left| \sum_{l}^{N} \int_{0}^{\tau_0} \dot{u}_\alpha \binom{l}{b}; t \right) e^{i\left[\boldsymbol{\kappa}.\boldsymbol{r}_0 \binom{l}{0} - wt \right]} dt \right|^2, \qquad (2)$$

where $\mu_0 = m/(4\pi\tau_0 N)$, m is the mass of a Si atom, τ_0 is the total simulation time, $\boldsymbol{r}_0$ is the equilibrium position vector of the lth unit cell, and $\dot{u}_\alpha$ is the α-component of the velocity of the bth atom in the lth unit cell at time t. There are a total of $N = N_x \times N_y \times N_z$ unit cells in the simulated computational domain with n atoms per unit cell. The phonon frequencies in Eq. (2) can only be obtained at the set of allowed wave vectors as dictated by the crystal structure. For our model, the ΓX-path wave vectors are $\kappa_x = 2\pi j/(N_x A_x)$, $j = 0$ to $N_x/2$.

4.2   Simulation-Based Evidence of Phonon-Vibron Coupling

In implementing the Φ' form of the SED method, we consider a unit cell similar to the one depicted in Fig. 3 with $A_x = A_y = A_z = 6$, $A_{px} = A_{py} = 2$, and $A_{pz} = 6$ considering only a single nanopillar. For the computational domain, we set $N_x = 50$ and $N_y = N_z = 1$, which gives a ΓX wave vector resolution of $\Delta\kappa_x = 0.04$. MD simulations under NVE (constant mass, volume, and energy) conditions are executed for this system for 2^{22} time steps where the duration of each time step is $\Delta t = 0.5$ fs. Equation (2) is evaluated by computing the Fourier transform of the velocity trajectories extracted every 2^5 time steps.

The SED spectrum is shown in Fig. 5a for the uniform membrane and in Fig. 5b–e for the NPM with different nanopillar heights. Only the frequency range $0 \leq \omega \leq 1.5$ THz is shown because higher frequencies are difficult to distinguish in the SED field. This frequency range is responsible for roughly 40% of the total lattice thermal conductivity (Davis and Hussein 2014). The phonon
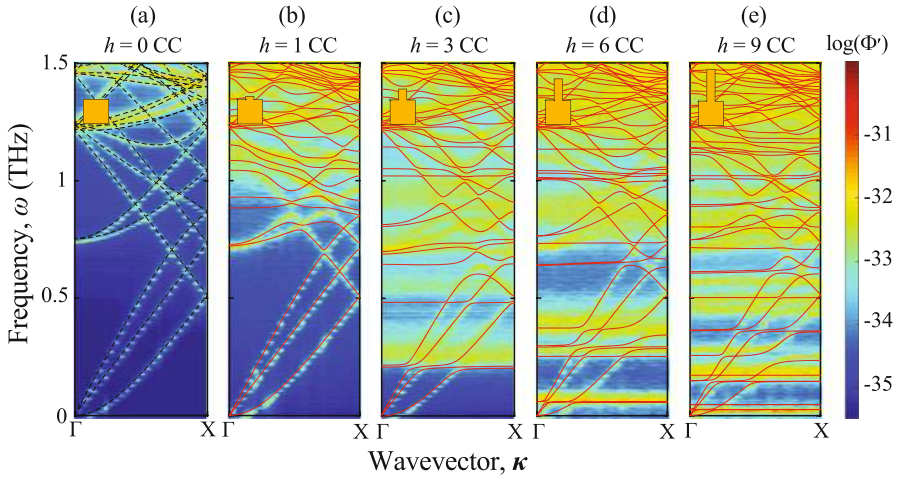
Fig. 5 Phonon dispersion of several NPMs with various nanopillar heights. The subfigures show the phonon dispersion as revealed by the SED field as a function of wave vector and frequency, and superimposed are the dispersion curves as directly obtained from harmonic LD calculations. The NPM unit-cell dimensions are $6 \times 6 \times 6 + 2 \times 2 \times h$ CC; the uniform membrane unit-cell dimensions are $6 \times 6 \times 6$ CC, i.e., $h = 0$ for the uniform membrane. Each inset presents a schematic of the unit cell analyzed

dispersion emerging from the MD simulations matches very well with that obtained by the independent LD calculations, thus providing confidence in both sets of simulations/calculations (upon close inspection, there is a slight upward deviation in the SED phonon dispersion curves compared to the LD dispersion curves, the extent of which is indicative of the degree of anharmonicities in the MD simulations). Considering the NPM case with $h = 3$ CC (Fig. 5c) as an example, we observe in the NPM SED spectrum the first two nanopillar local resonances appearing as horizontal lines at nearly 0.2 THz. More importantly, the interaction of these resonances with the acoustic branches of the underlying membrane is distinctly observed and follows closely the hybridization profiles featured in the LD dispersion curves. Resonance hybridizations are also seen at higher frequencies where vibrons interact with optical phonon dispersion branches. Upon comparing Fig. 5b–e with each other, the effects of changing the nanopillar height on the locations of the vibrons within the frequency spectrum are shown. The taller the nanopillar, the lower the first resonant frequency.

4.3   Maximization of Phonon-Vibron DOS Conformity

As we have described earlier, the key mechanism in an NPM is the coupling between phonons whose motion is confined to the in-plane directions within the base membrane and vibrons whose motion is limited primarily to the domain of each

nanopillar. The more effective this coupling across the full spectrum, the stronger we expect to achieve a reduction in the thermal conductivity. In Fig. 6, we examine how the overall size of the NPM affects the phonon and vibron density of states (DOS) distributions, separately and with respect to each other. A key factor in this comparison is the degree of DOS conformity, which may be quantified. For this purpose, we introduce a metric to represent the inverse of conformity, namely, the *nonconformity factor* $\hat{R}_{pv}$, which is defined as

$$\hat{R}_{pv} = \|R_{pv} - R_{pp}\|, \tag{3}$$

where R_{pv} is the cross-correlation between the phonon and the vibron DOS, R_{pp} is the autocorrelation of the phonon DOS, and $\|.\|$ denotes the double norm. The reader may refer to Rabiner and Gold (1975) for the definitions of correlation functions. The nonconformity factor varies between 0 (perfect conformity) and ∞ (no conformity). We see in Fig. 6 that as the sizes of the membrane and nanopillar portions are increased at an identical rate (the top path), the value of the nonconformity factor rises. On the other hand, we observe that when the size of the nanopillar portion is increased at a higher rate compared to the membrane portion (the bottom path), this value decreases, i.e., the degree of conformity intensifies. The former scenario corresponds to the set of cases considered by Honarvar et al. (2016). On the other hand, the latter scenario represents a design pathway that creates what we describe as *vibrons compensation* (Honarvar and Hussein 2018). Under this design scheme, nanopillar vibrons are added at a higher rate than membrane phonons as the overall size of the unit cell examined is increased. The outcome is that not only (1) the phonon band structure gets enriched with a higher vibrons-to-phonons ratio, but also (2) the level of conformity between the phonon and vibron DOS distributions gets significantly increased.

4.4 NPM Performance: Ultralow Thermal Conductivity

The size and geometry of the nanopillars are key to increasing the intensity of the resonance hybridization effect, as demonstrated in Fig. 6. In this subsection, we provide predictions of the thermal conductivity and seek to further explore the conditions for its extreme reduction. Equilibrium MD simulations and the GK method are used for the thermal conductivity calculations. The simulations are conducted using the LAMMPS software (Plimpton 1995). The systems are equilibrated under NVE ensembles with a time step $\Delta t = 0.5$ fs for a time span of 6 ns, which is sufficiently large compared to the longest phonon lifetime. The HCACFs converge within the first 500 ps. To minimize statistical error, the average of values from six independent simulations with different initial velocities is considered. Furthermore, the thermal conductivity predictions are averaged in the x- and y-directions, effectively resulting in an averaging over 12 predicted values.
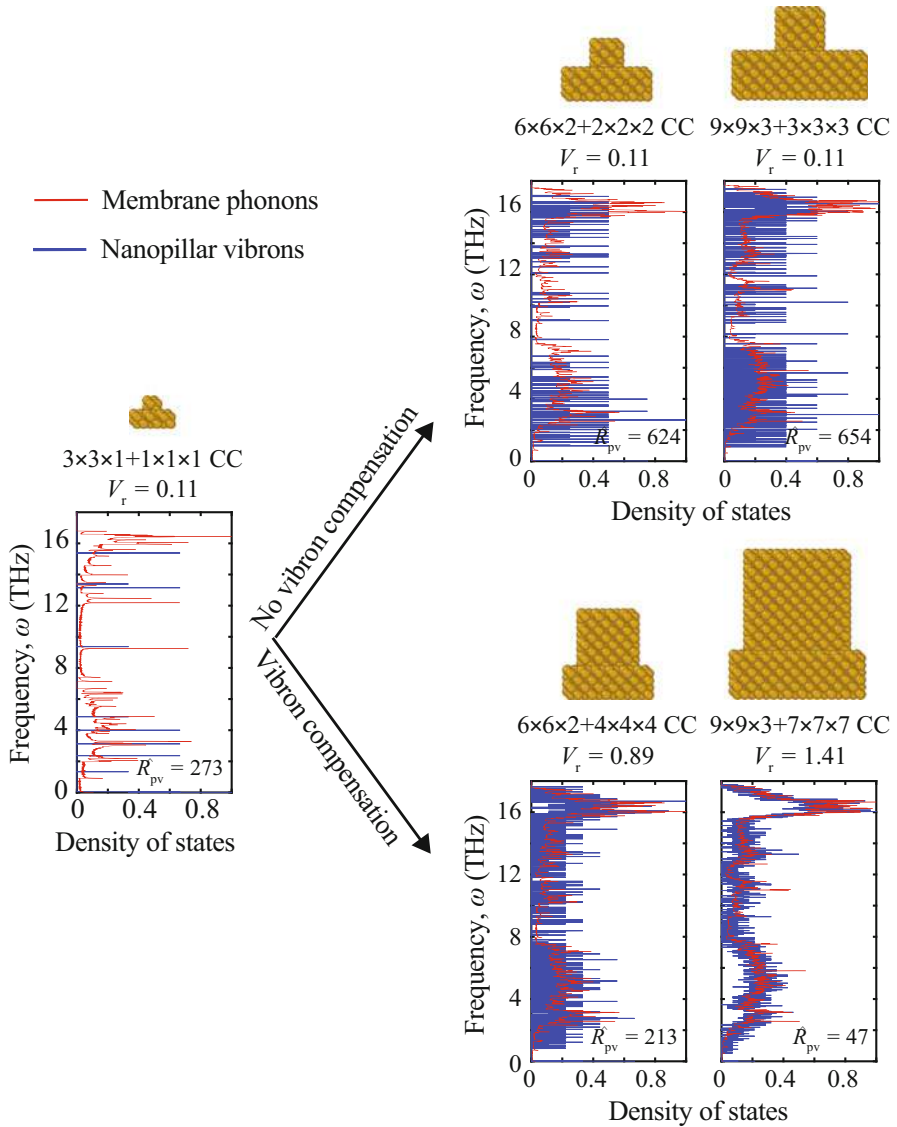
Fig. 6 Demonstration of phonon-vibron conformity as a function of unit-cell size and V_r. Relative distributions of the phonons and vibrons DOS are shown for three NPMs whose V_r values are kept constant (top route) and increased as the overall unit-cell size grows (bottom route). The leftmost NPM configuration has a $3 \times 3 \times 1$ CC membrane and a $1 \times 1 \times 1$ CC nanopillar. The vibrons DOS are obtained by considering the nanopillar as an independent nanostructure with free boundary conditions. All quantities are normalized with respect to their maximum values. Increasing the size of the nanopillar at a higher rate than the size of the membrane leads to higher phonon-vibron conformity, which corresponds to lower values of $\hat{R}_{pv}$ and a more intense resonance hybridization effect

Fig. 7 With nanopillar design optimization, the reduction in the thermal conductivity is shown to reach a factor of 130 (compared to a corresponding unpillared membrane). These predictions are obtained by equilibrium molecular dynamics simulations and the Green-Kubo method for Si-on-Si NPM models
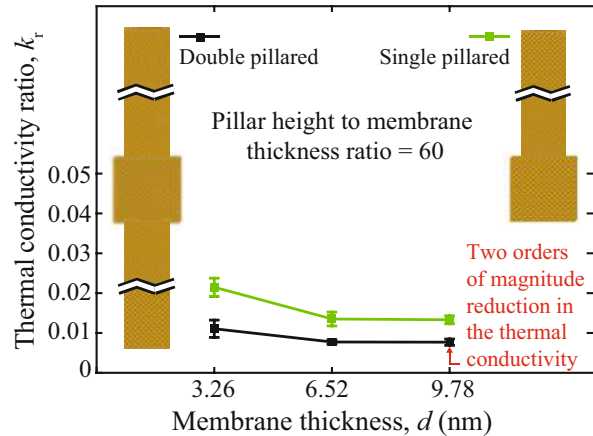
Figure 7 presents the reduction in the in-plane thermal conductivity for various NPM sizes and geometries. The reduction is represented by k_r, the lattice thermal conductivity of a membrane with nanopillar(s) in the unit cell divided by that of a uniform membrane with the same thickness. Key factors in this analysis are (1) the relative volumetric size of the nanopillar(s) with respect to the base membrane within the unit cell (this quantity is denoted V_r and is equal to DOF_{Pillar}/DOF_{Base} where DOF_{Pillar} and DOF_{Base} denote nanopillar(s) and base membrane number of degrees of freedom, respectively) and (2) the rate of change of V_r and the manner by which the geometry of the different NPM components (membrane and nanopillar) change relative to each other, as the size is increased. We consider a Si membrane with a Si nanopillar at each of the top and bottom surfaces. Using α as a size parameter, we specifically consider configuration $6\alpha \times 6\alpha \times 6\alpha + (6\alpha - 2) \times (6\alpha - 2) \times 6\alpha\beta + (6\alpha - 2) \times (6\alpha - 2) \times 6\alpha\beta$ where β is the nanopillar height divided by the membrane thickness. We also consider a single nanopillar unit cell which is described by the same configurational formula minus the last term.

The three data points shown in Fig. 7 represent $\alpha = 1$, 2, and 3, respectively, and for all the results shown, we have selected $\beta = 60$. These parameters give us $V_r = 53.3$, 83.3, and 94.8 for these three α values, respectively, which correspond to a case of profound vibron compensation. The thermal conductivity results, in turn, do yield improvements with upscaling as shown for both the single-pillared and double-pillared cases. We observe a reduction in the thermal conductivity by a factor of 75 ± 14 for the former case and a factor of 130 ± 28 for the latter case, for a membrane 9.78-nm thick with each nanopillar extending up to 586.5 nm in height. These are exceptionally strong reductions given that no scatterers have been introduced in the interior domain of the membrane and the membrane surfaces have not been roughened or oxidized. Yet more reductions are possible with further increases in nanopillar size and augmentation with conventional treatments such as optimized alloying (Xiong et al. 2016). Future research will explore different nanoresonator configurations to reduce the need for occupying a large volume. For

example, the nanopillars may be made from a higher atomic density material in order to generate the same number of vibrons using smaller heights or may be made from a material with heavier atoms in order to shift the vibron DOS to lower frequencies and thus be effective in reducing the thermal conductivity with shorter nanopillars.

4.5   Nanopillars Versus Nanowalls

The resonating substructures in an NPM may take a variety of forms. For example, instead of nanopillars we could instead erect nanowalls with a finite thickness along one direction in the plane of the membrane and extended to infinite in the orthogonal direction. A schematic of this modified NPM configuration is shown in Fig. 8a where a nanowall extends along the full length of the periodic unit cell in the x-direction and has a specified thickness along the y-direction. Compared to a squared nanopillar with the same thickness as the nanowall, the latter has more DOF; thus the NPM exhibits a higher value of V_r and is therefore expected to deliver a stronger level of thermal conductivity reduction. Furthermore, the asymmetry of the nanowall substructure is expected to produce anisotropic phonon dispersion characteristics and anisotropic planar thermal conductivity.

To confirm and quantify these expected results, we investigate a nanowalled membrane with dimensions $6 \times 6 \times 6 + 6 \times 4 \times 6$ CC and, for direct comparison, a corresponding nanopillared membrane with dimensions $6 \times 6 \times 6 + 4 \times 4 \times 6$ CC. The nanowall on top of the membrane has the same length as the unit cell along the x-direction and is therefore a contiguously extended substructure in this direction due to the application of the periodic boundary conditions. Along the y-direction, both the nanowall and nanopillar have the same thickness. In Fig. 8b, the dispersion curves for a uniform, a nanopillared, and a nanowalled membrane are shown; for the nanowalled case, phononic band structures are shown for wave propagation along both the ΓX and ΓY directions. We observe the ΓX dispersion diagram is relatively similar to that of the uniform membrane; indeed there is no periodicity in that direction and the nanowalled membrane is in fact a uniform waveguide. The ΓY dispersion diagram, on the other hand, is comparable to the nanopillared case because the nanowalls form a periodic system with freely vibrating substructures that exhibit a wide range of vibrons.

In Fig. 8, the ΓX thermal conductivity k_x and the ΓY thermal conductivity k_y of the nanowalled membrane are computed using EMD simulations and the GK method and compared to the nanopillared case. A total of six simulations were performed for each case. It is clear that the nanopillared membrane is practically isotropic with respect to the x- and y-directions. The nanowalled membrane, on the other hand, displays a strong reduction in the k_y/k_x ratio. In the y-direction, the thermal conductivity of the nanowalled system is noticeably smaller than the nanopillared system and reduces with increasing nanowall height (following a similar saturating trend as for nanopillared membranes). For a nanoresonator height of $h_T = 15$ CC, we predict an average k_r value of 0.1966 ± 0.0572 for the two

(a)

Nanopillar   Nanowall

z

x y

Y

Γ X

2D Brillouin zone

Base membrane

(b)   Uniform   Pillared   Walled   Walled
membrane membrane membrane membrane

1.5

1.2

0.9

0.6

0.3

0

Frequency, ω (THz)

Γ X Γ X Γ X Γ Y

Wavevector

(c)

Pillar or wall height, h_T (nm)

3.3   4.9   6.5   8.1

Anisotropic ratio, k_y/k_x

Isotropic line

Base   Pillar
$6{\times}6{\times}6{+}4{\times}4{\times}h_T$ CC
Base   Wall
$6{\times}6{\times}6{+}6{\times}4{\times}h_T$ CC

1

$k_{P_x}=$
12.7 ± 1.2
$k_{P_y}=$
11.4 ± 0.5

$k_{P_x}=$
8.7 ± 0.4
$k_{P_y}=$
8.8 ± 0.6

$k_{P_x}=$
9.1 ± 0.6
$k_{P_y}=$
8.2 ± 0.7

$k_{P_x}=$
9.6 ± 0.8
$k_{P_y}=$
8.9 ± 0.5

0.5

$k_{w_x}=$
17.9 ± 2.1
$k_{w_y}=$
9.0 ± 0.5

$k_{w_x}=$
15.2 ± 1.2
$k_{w_y}=$
9.4 ± 1.2

$k_{w_x}=$
15.7 ± 1.0
$k_{w_y}=$
6.4 ± 0.4

$k_{w_x}=$
14.8 ± 0.9
$k_{w_y}=$
5.9 ± 0.4

0

6   9   12   15

Nanopillar or nanowall height, h_T (CC)

Fig. 8 (**a**) Nanowalled membrane versus nanopillared membrane. (**b**) Phonon dispersions of uniform, nanopillared, and nanowalled membranes. (**c**) Ratio of the thermal conductivity along the y-direction to the thermal conductivity along the x-direction for nanopillared (background bars) and nanowalled (foreground bars) membranes. The anisotropy in the planar thermal conductivity is evident for nanowalled membranes. Specific values of x- and y-direction thermal conductivities for each of the configurations are provided

nanopillared cases and k_r values of 0.3157 ± 0.06368 and 0.1255 ± 0.0255 for the ΓX and ΓY nanowalled cases, respectively. The additional reduction in the thermal conductivity brought about by the nanowalls is a direct outcome of increasing the volume of the nanoresonator. In summary, replacing the nanopillars by nanowalls is shown to (1) reduce the thermal conductivity significantly (by a nearly 30% in the present example) and (2) introduce a sharp planar anisotropy in the thermal conductivity (by nearly a factor of 2 in the present example). Furthermore, adopting nanowalls might be more amenable for nanofabrication over nanopillars.

5   Thermoelectric *ZT* Figure-of-Merit Projections

Accurate prediction of thermoelectric *ZT* values for the NPM configurations investigated in this chapter requires intensive electronic structure calculations followed by implementation of techniques for obtaining the electrical properties

(e.g., see investigation by Mangold et al. (2016) for calculations and models used for thermoelectric analysis of uniform Si membranes). However, one may generate rough estimates of the ZT performance by projections. We recall that the lattice thermal conductivity of the double-nanopillar case reported in Fig. 7 is lower than a corresponding uniform membrane by a factor of 130 ± 28. We may multiply this number by 3.7 ± 0.7, which is the factor by which k_l drops for a 9.78-nm-thick uniform silicon membrane compared to the bulk form. This gives a total reduction factor of 481 ± 55. By conservatively assuming the same k_r value for $\alpha = 6$ for the same geometric configuration, a 19.55-nm-thick NPM would exhibit a lattice thermal conductivity reduction by a factor of 311 ± 85 with respect to bulk silicon. (We predict k_l of a 19.55-nm-thick uniform silicon membrane to be 2.4 ± 0.4 times lower than the bulk form.) If we assume $k_e \approx 0$ (Shanks et al. 1963), we may apply a one-to-one mapping between the lattice thermal conductivity reduction and the increase in ZT. (This assumption is supported by experimental studies of similarly sized freestanding nanomeshed silicon membranes (Yu et al. 2010) and silicon nanowires (Boukai et al. 2008). Considering the electrical conductivity in particular, the MFP of electrons in bulk silicon at room temperature has been determined experimentally to range, roughly, from 10 to 30 nm depending on the level of doping (Weber and Gmelin 1991); thus with moderate doping (which is recommended to optimize between S and σ), it is conceivable that a nanostructure with a 20-nm characteristic size would exhibit a negligible change in σ compared to the bulk form.) Given that ZT of moderately doped bulk silicon at $T = 300$ K is 0.01 at a minimum (Weber and Gmelin 1991; Stranz et al. 2013), we obtain through this mapping a projected room-temperature ZT value of 3.1 ± 0.9. (If we assume that the electrical properties also do not change for the best 9.78-nm-thick double-nanopillar case, we would obtain a room-temperature ZT value of 4.8 ± 0.6.)

In some references, the contribution of the electronic thermal conductivity in silicon is reported to be relatively significant (Lee et al. 2008; Mangold et al. 2016). This would require a modification to our projections. When accounting for k_e, we will still assume that the NPM electrical properties are bulk-like, i.e., the NPM electronic thermal conductivity is the same as the bulk electronic thermal conductivity. By considering a nonzero k_e, there is now a one-to-one mapping between the total (not lattice) thermal conductivity and ZT improvement. The predicted values of the bulk electronic thermal conductivity are taken from the studies by Lee et al. (2008) and Mangold et al. (2016). These values were obtained by first principles calculations and are shown in Fig. 9a for doping concentrations in the range of 0.2–$8 \times 10^{19} \text{cm}^{-3}$. Across these two references, there are some discrepancies between the two sets of results that would affect the ZT projections as is also shown in Fig. 9 – the k_e values predicted by Mangold et al. are generally higher than the ones reported by Lee et al. In our analysis, the k_e values are first normalized by an experimental value of the lattice thermal conductivity, which we take as $k_l = 160 \, \text{Wm}^{-1} \, \text{K}^{-1}$. This fraction is then multiplied by the EMD-GK lattice thermal conductivity of $k_l = 376 \pm 19 \, \text{Wm}^{-1}\text{K}^{-1}$ to estimate the corresponding k_e values for our models. Following these assumptions, in Fig. 9b, the NPM ZT values are projected based on the bulk ZT values of 0.01 (for carrier concentrations
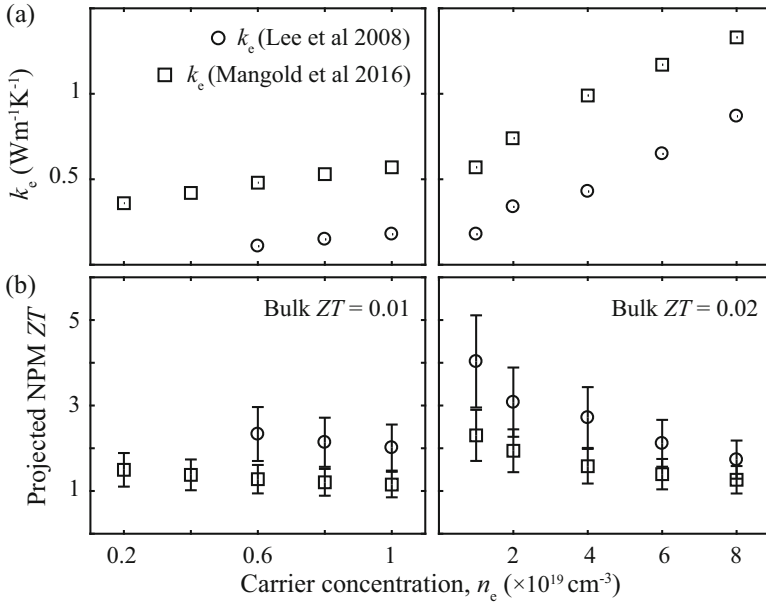
Fig. 9 (**a**) Electronic thermal conductivity k_e and (**b**) projected NPM ZT versus carrier concentration. The experimental and theoretical values of the bulk lattice thermal conductivity are chosen to be $k_l = 160\,\mathrm{Wm^{-1}K^{-1}}$ and $k_l = 376 \pm 19\,\mathrm{Wm^{-1}K^{-1}}$, respectively (Kremer et al. 2004; Honarvar and Hussein 2018)

in the range 0.2–1 $\times$ $10^{19}\mathrm{cm^{-3}}$) and 0.02 (for carrier concentrations in the range 1–8 $\times$ $10^{19}\mathrm{cm^{-3}}$) (Stranz et al. 2013). Depending on the carrier concentration and k_e values, the projected NPM average ZT varies roughly between 1 to 4. For example, for silicon that is doped with a carrier concentration of $1 \times 10^{19}\,\mathrm{cm^{-3}}$ and has a bulk ZT value of 0.02, we project a ZT value of 4 ± 1.1 for the 19.55-nm-thick double-nanopillared silicon membrane. The upper bound of this range is significantly higher than any previously reported ZT value at any temperature, for silicon and for thermoelectric materials in general.

6 Conclusions

This chapter presents elements of phonon dynamics of a recently developed class of nanostructured metamaterials, nanophononic metamaterials, that have their thermal conductivity dramatically decreased by an array of distributed resonating substructures. The substructures' numerous local resonances hybridize with the phonon modes of the underlying host, or base, material, which leads to reducing

their group velocities significantly and creating energy localization–two factors that reduce the thermal conductivity. In addition, the local resonances, also termed vibrons, reduce the phonon lifetimes at and around the hybridization regions in the phonon band diagram thus causing even more reduction in the thermal conductivity.

A particularly useful type of an NPM is in the form of an all-silicon nanopillared, or nanowalled, freestanding membrane. This configuration is advantageous because the nanoresonators are located outside the domain of the transport medium, which is the membrane, and is therefore not expected to significantly alter the electrical properties. Such outcome is ideal for thermoelectric energy conversion which requires a material with as low thermal conductivity as possible and as high Seebeck coefficient and electrical conductivity as possible.

The notion of conformity between the membrane and nanoresonator density of states is presented and analyzed. It is shown in Fig. 6 that this conformity in a nanopillared membrane increases when the size of the nanopillars increases at a higher rate than the base membrane (when considering progressively larger unit cells). This scenario, termed vibrons compensation, enables a stronger reduction in the thermal conductivity with upscaling – a trait that is highly desirable for potential integration of NPMs into TE devices.

From our calculations, the lattice thermal conductivity of a 9.78-nm-thick uniform silicon membrane is 3.7 ± 0.7 times lower than the bulk form. Multiplying this by the factor of 130 ± 28 for the double-nanopillar case reported in Fig. 7 gives a total factor of 481 ± 55, which corresponds to roughly half the thermal conductivity of bulk amorphous silicon (He et al. 2011; France-Lanord et al. 2014; Larkin and McGaughey 2014). By conservatively assuming the same k_r value for $\alpha = 6$ for the same geometric configuration, the lattice thermal conductivity for a 19.55-nm-thick NPM would be lower than for bulk silicon by a factor of 311 ± 85. Assuming that all the electrical properties retain their bulk values at a length scale on the order of 20 nm and at $T = 300$ K (see Section 5), we may estimate the effect of a decrease in the lattice thermal conductivity on the ZT value. Given that the ZT of moderately doped bulk silicon at $T = 300$ K ranges from 0.01 to 0.02 (Stranz et al. 2013), we obtain a projected room-temperature ZT value in the range of 1–4 depending on the specific level of doping. The upper bound for this projection is significantly higher than any previously reported ZT value at any temperature, not only for silicon but for thermoelectric materials in general.

References

Achaoui Y, Laude V, Benchabane S, Khelif A (2013) Local resonances in phononic crystals and in random arrangements of pillars on a surface. J Appl Phys 114:104503

Anderson J, Burns PJ, Milroy D, Ruprecht P, Hauser T, Siegel HJ (2017) Deploying RMACC summit: an HPC resource for the Rocky Mountain Region. In: Proceedings of the Practice and experience in advanced research computing 2017 on sustainability, Success and Impact, ACM p. 8

Boukai AI, Bunimovich Y, Tahir-Kheli J, Yu JK, Goddard WA III, Heath JR (2008) Silicon nanowires as efficient thermoelectric materials. Nature 451(7175):168–171

Chen G, Dresselhaus M, Dresselhaus G, Fleurial JP, Caillat T (2013) Recent developments in thermoelectric materials. Int Mater Rev 48(1):45–66

Davis BL, Hussein MI (2014) Nanophononic metamaterial: thermal conductivity reduction by local resonance. Phys Rev Lett 112:055505

Esfarjani K, Chen G, Stokes HT (2011) Heat transport in silicon from first-principles calculations. Phys Rev B 84:085204

France-Lanord EB, Albaret T, Merabia S, Lacroix D, Termentzidis K (2014) Atomistic amorphous/crystalline interface modelling for superlattices and core/shell nanowires. J Phys-Condens Mat 26:055011

He Y, Donadio D, Galli G (2011) Heat transport in amorphous silicon: interplay between morphology and disorder. Appl Phys Lett 98:144101

Honarvar H, Hussein MI (2016) Spectral energy analysis of locally resonant nanophononic metamaterials by molecular simulations. Phys Rev B 93:081412(R)

Honarvar H, Hussein MI (2018) Two orders of magnitude reduction in silicon membrane thermal conductivity by resonance hybridizations. Phys Rev B 97:195413

Honarvar H, Yang L, Hussein MI (2016) Thermal transport size effects in silicon membranes featuring nanopillars as local resonators. Appl Phys Lett 108:263101

Hussein MI, Davis BL (2016) Nanophononic metamaterials. US patent 9417465 B2

Hussein MI, Leamy MJ, Ruzzene M (2014) Dynamics of phononic materials and structures: historical origins, recent progress, and future outlook. Appl Mech Rev 66:040802

Johnson JA, Maznev AA, Cuffe J, Eliason JK, Minnich AJ, Kehoe T, Torres CMS, Chen G, Nelson KA (2013) Direct measurement of room-temperature nondiffusive thermal transport over micron distances in a silicon membrane. Phys Rev Lett 110:025901

Khammas A (2007–2018) Book of synergy. http://www.buch-der-synergie.de

Kremer R, Graf K, Cardona M, Devyatykh G, Gusev A, Gibin A, Inyushkin A, Taldenkov A, Pohl HJ (2004) Thermal conductivity of isotopically enriched ^{28}Si: revisited. Solid State Commun 131(8):499–503

Ladd AJC, Moran B, Hoover WG (1986) Lattice thermal conductivity: a comparison of molecular dynamics and anharmonic lattice dynamics. Phys Rev B 34:5058–5064

Landry ES, Hussein MI, McGaughey AJH (2008) Complex superlattice unit cell designs for reduced thermal conductivity. Phys Rev B 77:184302

Larkin JM, McGaughey AJH (2014) Thermal conductivity accumulation in amorphous silica and amorphous silicon. Phys Rev B 89:144303

Larkin JM, Turney JE, Massicotte AD, Amon CH, McGaughey AJH (2014) Comparison and evaluation of spectral energy methods for predicting phonon properties. J Comput Theor Nanos 11(1):249–256

Lee JH, Galli GA, Grossman JC (2008) Nanoporous si as an efficient thermoelectric material. Nano Lett 8:3750–3754

Liu ZY, Zhang XX, Mao YW, Zhu YY, Yang ZY, Chan CT, Sheng P (2000) Locally resonant sonic materials. Science 289(5485):1734–1736

Mangold C, Neogi S, Donadio D (2016) Optimal thickness of silicon membranes to achieve maximum thermoelectric efficiency: a first principles study. Appl Phys Lett 109:053902

McGaughey AJH, Kaviany M (2004) Thermal conductivity decomposition and analysis using molecular dynamics simulations. Part I. Lennard-Jones argon. Int J Heat Mass Tran 47(8): 1783–1798

Minnich AJ (2012) Determining phonon mean free paths from observations of quasiballistic thermal transport. Phys Rev Lett 109:205901

Monroe D (2014) Slowing heat without obstructions. Physics 7:14

Neogi S, Reparaz S, Pereira LFC, Graczykowski B, Wagner MR, Sledzinska M, Shchepetov A, Prunnila M, Ahopelto J, Torres CMS, Donadio D (2015) Tuning thermal transport in ultrathin silicon membranes by surface nanoscale engineering. ACS Nano 9:3820–3828

Pendry JB, Holden AJ, Robbins DJ, Stewart WJ (1999) Magnetism from conductors and enhanced nonlinear phenomena. IEEE Trans Microwave Theory Tech 47:2075–2084

Pennec Y, Djafari-Rouhani B, Larabi H, Vasseur JO, Hladky-Hennion AC (2008) Low-frequency gaps in a phononic crystal constituted of cylindrical dots deposited on a thin homogeneous plate. Phys Rev B 78:104105

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19

Rabiner LR, Gold B (1975) Theory and application of digital signal processing. Prentice-Hall Inc, Englewood Cliffs

Rowe DM (2005) Thermoelectrics handbook: macro to nano. CRC Press, Boca Raton

Shanks HR, Maycock PD, Sidles PH, Danielson GC (1963) Thermal conductivity of silicon from 300 to 1400°K. Phys Rev 130:1743–1748

Smith DR, Padilla WJ, Vier DC, Nemat-Nasser SC, Schultz S (2000) Composite medium with simultaneously negative permeability and permittivity. Phys Rev Lett 84:4184–4187

Snyder G, Toberer E (2008) Complex thermoelectric materials. Nat Mater 7:105–114

Stillinger FH, Weber TA (1985) Computer simulation of local order in condensed phases of silicon. Phys Rev B 31(8):5262

Stranz A, Kähler J, Waag A, Peiner E (2013) Thermoelectric properties of high-doped silicon from room temperature to 900 k. J Electron Mater 42:2381–2387

Thomas JA, Turney JE, Iutzi RM, Amon CH, McGaughey AJH (2010) Predicting phonon dispersion relations and lifetimes from the spectral energy density. Phys Rev B 81(8):081411

Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD et al (2014) Xsede: accelerating scientific discovery. Comput Sci Eng 16(5): 62–74

Vineis CJ, Shakouri A, Majumdar A, Kanatzidis MG (2010) Nanostructured thermoelectrics: big efficiency gains from small features. Adv Mater 22:3970–3980

Weber L, Gmelin E (1991) Transport properties of silicon. Appl Phys A 53:136–140

Wu TT, Huang ZG, Tsai TC, Wu TC (2008) Evidence of complete band gap and resonances in a plate with periodic stubbed surface. Appl Phys Lett 93:111902

Xiong S, Sääskilahti K, Kosevich YA, Han H, Donadio D, Volz S (2016) Blocking phonon transport by structural resonances in alloy-based nanophononic metamaterials leads to ultralow thermal conductivity. Phys Rev Lett 117:025503

Yu JK, Mitrovic S, Tham D, Varghese J, Heath JR (2010) Reduction of thermal conductivity in phononic nanomesh structures. Nat Nanotechnol 5:718–721

Zwanzig R (1965) Time-correlation functions and transport coefficients in statistical mechanics. Annu Rev Phys Chem 16:1798–1800

Modeling of Heat Transport in Polymers and Their Nanocomposites

41

Pawel Keblinski

Contents

Abstract

In this chapter, methodology to model heat flow in polymers and their nanocomposites will be presented. One focus will be on the use of molecular-level modeling techniques for the thermal conductivity determination of bulk polymers. In the case of polymer nanocomposites, the methods for molecular-level modeling of the interfacial thermal resistance will be described, and its role along with other factors, such as filler shape, orientational distributions, and dispersion on the composite thermal conductivity, will be analyzed within the homogenization analysis.

1   Introduction

Polymers, as well as their composites, constitute complex materials with a number of structural characteristics, such as polymer chain backbone chemistry and con-

---
P. Keblinski (✉)
Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
e-mail: keblip@rpi.edu

formation, side groups, and filler particle size and distribution, all of which are influencing properties, including thermal transport characteristics. The structural complexity of these materials requires an informed approach in order to model their heat transport coefficients. On one hand, one needs to capture the most important structural and bonding features to enable predictions with a reasonable level of fidelity, and, on the other hand, the model has to be tractable either analytically or computationally.

We will limit our consideration to solid polymers and their composites and thus disregard contributions to the heat flow from convection that might play a role in the liquid phase during polymer processing. Within this constraint, from a practical perspective, the most important characteristics for heat conduction are thermal conductivity, κ (W/m-K), and volumetric heat capacity, C (J/K-m^3). The thermal conductivity is defined via Fourier's law:

$$j_q = -\kappa \nabla T \tag{1}$$

where j_q is the heat flux (W/m^2) and ∇T is the temperature gradient (K/m). Accounting for energy conservation and heat sources, one can derive the diffusive heat flow equation:

$$C \frac{\partial T}{\partial t} - \nabla (\kappa \nabla T) = \frac{\partial q_V}{\partial t} \tag{2}$$

where q_V is the volumetric heat source. The above two equations are general and describe heat conduction in all materials at the macroscopic scale. Therefore the modeling of the specific materials focuses on the modeling and determination of the thermal conductivity and heat capacity, and the rest is the analytical or numerical solution of the well-known continuum problem given by Eq. 2, with appropriate boundary and initial conditions.

The determination of thermal conductivity and its relationship to the underlying structure and bonding, in addition to practical value, is motivated by scientific discovery. Such scientific knowledge exposes the underlying the mechanism of thermal transport and provides guidance toward the design of the materials with the desired thermal conductivity. In this context, modeling and simulations are uniquely positioned to contribute to such design as one can control and characterize the structure and relate it to thermal conductivity to a much larger extent than is possible in experiment.

In order to systematically present an approach to modeling of the thermal conduction of polymers and their nanocomposites, this chapter is composed of several sections. The following section provides background on thermal conduction due to vibrations of ion and atoms, i.e., phonons that are the dominant heat carriers in polymers, along with conceptual models of the thermal conductivity of polymers. The third section presents the use of molecular dynamics simulations for thermal conductivity determination, particularly in the context of polymers.

The fourth section describes molecular-level modeling of the interfacial thermal resistance associated with dissimilar constituents forming polymer nanocomposites. The fifth section provides a connection between molecular-level modeling and heat flow at larger scales either by solutions of the continuum heat flow equation or homogenization theory. The last section will provide a summary and conclusion.

2   Conceptual Models of Thermal Conduction in Polymers

To provide a fundamental context of heat conduction modeling in polymers and nanocomposites, we briefly discuss phonon (vibration)-based thermal transport kinetic theory. Within this theory, for crystalline materials, each propagating vibrational mode (phonon) contributes to thermal conductivity, and for a system of volume V, the bulk thermal conductivity is given by (Ziman 1960):

$$\kappa = \frac{1}{3V} \sum V_j(v)\lambda_i(v)c_j(v, T) \tag{3}$$

where $c_i(v, T)$ is the spectral heat capacity, i.e., a temperature dependent of heat capacity of a harmonic oscillator with a frequency v, v_i is the mode-specific group velocity (in the low frequency limit, it is the speed of sound), and λ_i is the phonon mean free path, and the summation is done over all vibrational modes. Such formulation is very successful in modeling of thermal transport in crystalline materials. In this case determining the phonon mean free path or alternatively, the phonon lifetime, poses the biggest challenge. This quantity is affected by multiple phonon scattering processes, such as anharmonicity-related phonon-phonon scattering, or phonon-defect scattering, and can be calculated from theory or evaluated in molecular-level simulations (Turney et al. 2009).

The description of phononic heat conduction provided above has some relevance to polymers in the case of polymers with a degree of crystallinity and, perhaps most prominently, in the case of polymer fibers with highly aligned polymer chains. The vast majority of polymers, however, have an amorphous structure or at most a semicrystalline structure involving a mix of crystalline and amorphous regions. This renders the kinetic theory of propagating phonons largely not applicable.

However, the heat conduction in amorphous polymers is still carried by vibrations. For such materials, as well as other nonmetallic glasses, the basic heat transport model is the so-called minimum thermal conductivity, which was first proposed by Einstein and later developed by Slack (Cahill et al. 1992). The basic heat transport model builds upon the Einstein model and borrows from the Debye model of solids to allow for different frequencies and wavelengths. The model assumes that the lifetime of the vibration is half of the period of vibrations, and summing up the contribution of all vibrations with the Debye dispersion relationships, it yields minimum thermal conductivity, $\kappa_{\min}$, via the following formula:

$$\kappa_{min} = \left(\frac{\pi}{6}\right)^{1/3} k_B n^{2/3} \sum_j V_j \left(\frac{T}{\theta_j}\right)^2 \int_0^{B,/T} \frac{x^3 e^x}{(e^x - 1)^2} dx \qquad (4)$$

where the sum is calculated over one longitudinal and two transverse phonon branches, with their respective speeds of sound, $v_{i=v_l}$ *or* $v_{i=v_t}$, and Debye temperatures, θ_i. In Eq. 1, n is the number density of atoms or ions, and k_B is the Boltzmann constant. Equation 3 correctly predicts a continuous increase of κ from zero as a function of temperature, due to increasing heat capacity, with κ saturating at higher temperatures. Examination of the κ predictions for a variety of glasses and disordered crystals indicates that this theory/estimate provides a good representation of the experimental measurements (Cahill et al. 1992).

In the case of amorphous polymers, at room temperature or above, we can take the high-temperature limit of the expression given by Eq. 3, yielding $\kappa_{\min} = 0.4 k_g n^{2/3} (V_i + 2V_t)$. This expression was tested in experiment and molecular dynamics simulations by monitoring thermal conductivity dependence on high pressure applied to amorphous poly(methyl methacrylate) (PMMA) (Hsieh et al. 2011), yielding reasonable agreement with the experimental data and even better agreement with the result of molecular dynamics simulations. Such good agreement is rather surprising because polymer materials exhibit highly nonhomogenous bonding with stiff bonds within the polymer chain backbone and side groups and much softer nonbonded interactions. Furthermore, nonbonded interactions between the chains are significantly nonlinear, and anharmonic interactions between localized vibrations, of the type invoked in "fracton hopping" models (Alexander et al. 1986), could conceivably contribute to heat transport. Evidence for the importance of anharmonicity in the thermal conductivity of amorphous polymers was recently observed in molecular dynamics (MD) simulations of polystyrene (PS) (Shenogin et al. 2009).

It is also likely that the minimum thermal conductivity model is a poor predictor of thermal conductivity in the case of amorphous polymers when polymer chains are characterized by relatively long straight sections, i.e., by a large persistence length In this case, the nature of the heat flow involves a rapid propagation of heat along the straight sections and slower propagation between sections.

The effect of such straight sections is even more pronounced for oriented polymers and can be characterized by a much larger and highly anisotropic conductivity, with a high conductivity in the alignment direction and a low conductivity in the transverse directions. While drawing *amorphous* polymers increases their conductivity by about a factor of ~3 in the draw direction (Kurabayashi 2001), drawing of semicrystalline polymers can increase their thermal conductivities 100-fold (Choy et al. 1978).

Current theoretical descriptions of heat transport in such complex polymeric materials involve phenomenological and continuum-level models. For example, recognizing that heat flow along the covalently bonded polymer backbone is more efficient than in the transverse directions, a simple model of oriented resistors was proposed to describe the anisotropy of drawn amorphous polymers (Henning 1967).

To obtain upper and lower bounds (Ward 1971) for the effective conductivities, resistor-in-series and resistor-in-parallel models are used. According to the literature, a resistor-in-series model provides a better prediction of experimental behavior. For such a model, orientational averaging gives the following expression for macroscopic conductivities, $k_{||}$ and $k_{\perp}$:

$$\frac{1}{K_{\perp}} = \frac{1}{2}\left[\left(\frac{1}{K_{ji\perp}} + \frac{1}{K_{ji||}}\right) - \left(\frac{1}{K_{ji||}} + \frac{1}{K_{ji\perp}}\right)\left\langle\cos^2\delta\right\rangle\right] \quad and$$

$$\frac{1}{K_{||}} = \frac{1}{2}\left(\frac{1}{K_{ji||}} - \frac{1}{K_{ji\perp}}\right)\left\langle\cos^2\delta\right\rangle + \frac{1}{K_{ji\perp}}$$

(5)

where δ is the angle between the chain direction and direction of heat flow and $\langle\cos^2\delta\rangle$ is the orientational average over the sample. Although the above model provides a reasonable description of the thermal conductivity of deformed amorphous polymers, the parameters of the model are obtained from fits to the data and have no fundamental microscopic justification. The much larger conductivity and anisotropy observed in aligned semicrystalline and liquid crystal polymers is described by rapid heat flow along the well-connected "continuum" network of conducting fibers of highly aligned and crystalline polymers (Ashton et al. 1969).

In these phenomenological approaches, the "microscopic" thermal conductivities are treated simply as adjustable parameters to be fitted to the experimental data. Physically, of course, the true microscopic parameters are determined by the details of the atomic-level bonding and the conformation of the polymer chains. In this instance, one has to resort to more fundamental modeling approaches, most prominently to atomic-level models of polymers and the use of molecular dynamics simulations to predict thermal conductivity that will be described in the following section.

3   Molecular Simulations of Polymers Thermal Conductivity

As already discussed above, the key goal of polymer thermal transport modeling is relating the structure and bonding to thermal transport characteristics to enable an informed design of the polymers with desired thermal conductivity and to gain insights into the fundamental mechanism of thermal transport in these materials. To achieve this goal one can resort to the use of molecular dynamics (MD) simulations.

Molecular dynamics is a well-developed classical simulation method in which Newton's equation of motion, $F = ma$, is integrated in time for a collection of atoms. The force between the atoms is prescribed in a semiempirical or empirical manner, carefully fitted to pertinent experimental data or to electronic structure calculations. The MD approach can capture all of the features of the vibrational motions of the atoms and can reproduce the phonon spectrum of the system in a quantitative or semiquantitative manner.

Molecular dynamics (MD) simulations are classical and are thus strictly only applicable above the Debye temperature. For example, for polyethylene (PE), strong covalent intra-chain bonds lead to Debye temperatures of 1650 K and 700 K for longitudinal and transverse along-the-chain (one-dimensional) modes, respectively. Much weaker nonbonded, van der Waals interactions lead to corresponding temperatures of 223 K and 134 K (Tucker and Reese 1967). Since for practical applications room temperature behavior is the most relevant, at first glance this would appear to produce an insuperable obstacle to classical simulations in which all vibrations are fully excited, regardless of the temperature. However, thermal energy is mostly carried by low-frequency and intermediate-frequency vibrations. By contrast, high-frequency phonons have rather low group velocities and are easy to scatter; thus, although overrepresented in the classical simulation, they do not contribute significantly to thermal transport. As an approximation to real quantum behavior, a temperature renormalization technique has been developed, where the temperature at which the classical simulation is performed actually corresponds physically to a somewhat higher temperature for a quantum system of the same mean kinetic energy (Wang et al. 1990). However, the dynamics of such a system are still classical, and, by definition, it is assumed that all excited phonons contribute equally to the heat transport. Goddard and his coworkers performed semiclassical analysis that took into account quantum effects (Che et al. 2000). Remarkably, they found that quantum corrections, even for diamond (Debye temperature is ~2200 K), do not significantly affect thermal transport at room temperature. Given the similar strong bonding in polymeric materials, a purely classical approach is both completely suitable and computationally the most efficient for the description of polymer's thermal transport properties even at room temperature.

We also note that full electronic-level approaches for MD simulations, i.e., ab initio MD or quantum-mechanics-based tight-binding MD simulations, do not provide significant advantages and have their own limitations. While in these methods, the forces between atoms and ions are calculated more accurately using quantum-mechanics-based calculations, the computational load is very significant and often prohibitive, and the dynamics of atomic/ionic motion is still classical.

One challenge in using MD simulations for polymers lies in the fact that the structural relaxation time even in polymer melts is much longer than that accessible to molecular dynamics. Therefore it is critical to construct outright model structures representative of real polymers. The amorphous structures can be created using a method first proposed by Suter and collaborators (Muller et al. 2001). This method (the software is named PolyPack) consists of a heuristic search algorithm in the space of torsion angles, which automatically delivers the correct conformational statistics of the chains maintaining so-called rotational isomeric states (RIS). Such structures can be later deformed to mimic drawing (see Fig. 1). Hybrid approaches need to be employed to create models or semicrystalline polymers that involve "composites" of crystalline and amorphous regions (Lee and Rutledge 2011).

Once model structures are generated, the forces acting between ions and atoms need to be prescribed. For simulations of thermal transport polymers, there are two levels of molecular modeling. The first is united atom (UA) description,

Fig. 1 Fully atomistic
polycarbonate structure
prepared with PolyPack
software and deformed under
constant load in tension. The
color code represents the
local strain
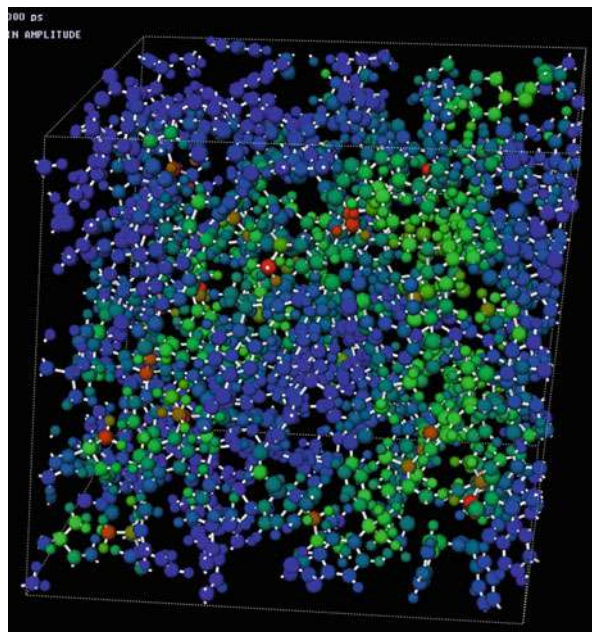
where hydrogen atoms are not represented explicitly, but are "united" with carbon
or other heavy atoms. This simplification is not expected to significantly affect
thermal energy propagation, since hydrogen atoms introduce very-high-frequency
local vibrations that are very weakly coupled to vibrations responsible for heat
transport. For example, the UA description used in simulations of thermal transport
in crystalline and liquid paraffin (see Fig. 2) predicted thermal conductivity in
excellent agreement with experiment (Babaei et al. 2013). The chief benefit of the
UA approach is the ability to access one order of magnitude longer simulation times
and larger structures at the same computational cost.

Effects associated with the explicit presence of H atoms are expected to play a
role associated with altered nonbonded interactions. For example, it is known that
explicit atom models are needed to correctly reproduce the crystalline PE structure
(Duffour and Malfreyt 2002) or more accurate structures of more complex polymers.
For this purpose one of the most widely used descriptions of atomic interactions is
the polymer consistent force field (PCFF) (Sun 1998), a second-generation force
field, derived from ab initio models and is an extension of CFF91 force field, which
was parameterized against a wide range of experimental observables for organic
compounds.

The most popular descriptions of interaction for both united and all-atom models
are given in the following general form:

$$V = V_2\left(r_{ij}\right) + V_3\left(\theta_{ijk}\right) + V_4\left(\phi_{ijkl}\right) + V_{nb}\left(r_{ij}\right) \qquad (6)$$
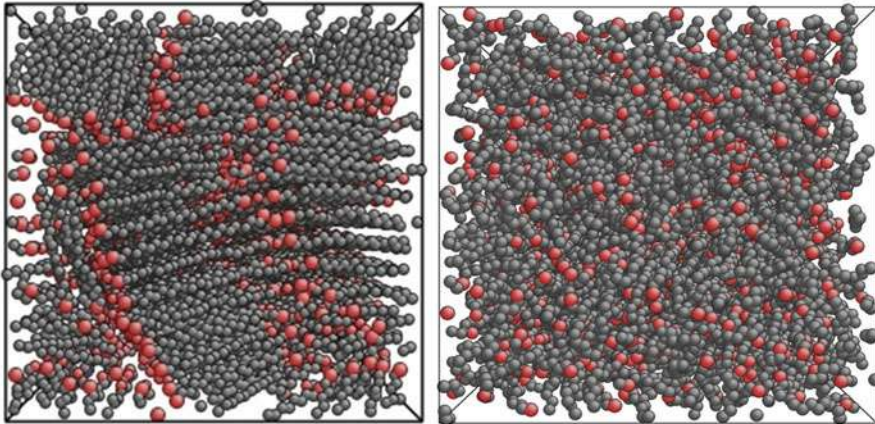
Fig. 2 United atom models of crystalline and liquid paraffin. Red "united atoms" are end chain CH_3 groups, and gray atoms represent CH_2. (From Babaei et al. 2013)

where the stretching of the covalent bonds is described by a pair potential V_2, preferably accounting for anharmonicity, and thus capable of capturing bond-stretching-related contributions to the phonon-phonon scattering process. We note, however, that even with a harmonic bond-stretching potential, due to coupling with bending, the system is not harmonic with respect to atomic displacements. Furthermore, except for highly oriented polymers, phonon scattering is dominated by structural scattering. The bond-bending potential, V_3, describes interactions between three consecutive bonded atoms and depends on the equilibrium angle and the spring constant parameterized to the frequencies of bending. The torsional potential, V_4, describes the different energetics of trans and gauche configurations (Duffour and Malfreyt 2002). Finally, the nonbonded interactions (i.e., those between united or explicit atoms on different chains and united atoms on widely separated parts of the same chain) use pair potentials to describe hard-core repulsion (preventing atomic overlap), van der Waals attraction, and Coulomb interactions, if applicable. Once the model structures are prepared, and atomic interactions described, the proper MD simulations can be executed.

The most straightforward method for thermal conductivity determination via the MD simulations is the nonequilibrium molecular dynamics (NEMD) method. The most popular NEMD approach is analogous to a simple experimental setup with a temperature gradient induced by a planar heat source and heat sink. This method is referred to as the direct method or heat source and sink method (Schelling et al. 2002). In this method the temperature of the sink and the source is kept constant by the application of local thermostats, or heat energy is added to the source and removed for the sink at a constant rate. In both approaches, heat power can be determined, and thus the heat flux can be evaluated. At the same time, after a transient period, when the system reaches a steady state, the temperature profile between the source and the sink is determined, which in the this simple planar
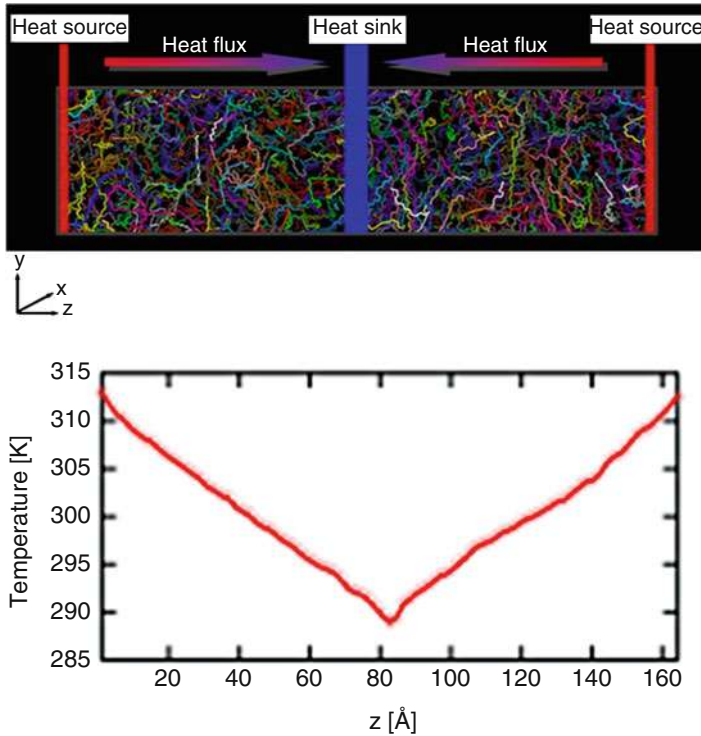
Fig. 3 Top: The simulation cell containing amorphous PE subjected to a steady-state heat flow. Bottom: The corresponding temperature profile. (From Kikugawa et al. 2013)

geometry is simply a straight line (see Fig. 3). Subsequently, with the known heat flux and temperature gradient, the thermal conductivity can be determined from Eq. 1.

In the direct methods simulations, the molecular-level formula for the heat flux allows one to look into more detail into the path and channels of the heat flow. This formula is typically given for the so-called heat current, $J_q = V j_q$., where j_q is the heat flux form (see Eq. 1) and V is the volume considered. In the case of pairwise atomic interactions, the heat current is given by:

$$\mathbf{J}_q = \sum_i \mathbf{v}_j \varepsilon_i + \frac{1}{2} \sum_{ij \; i \neq j} \mathbf{r}_{ij} \left(\mathbf{F}_{ij} \cdot \mathbf{v}_i \right) \tag{7}$$

where the sum is over all atoms, or atomic pairs in the given volume, $\mathbf{F}_{ij}$ is the force on atom i due to atom j, ε_i is the energy of atom i, and $\mathbf{v}$s are atomic velocities. The above formulation can be generalized in the case of many body interactions, such as three-body bond-bending and four-body torsional interactions (Torii et al. 2008). Using the above formula, one can create a spatial map of the heat current density

or evaluate contributions to the heat flow separately from bonded and nonbonded interactions – such analyses were performed for regular and crossed-linked PE (Kikugawa et al. 2013).

The thermal conductivity determination from the direct method assumes that Fourier's law is valid. The validity of Fourier's law relies of the diffusive heat flow mechanism, i.e., on the fact that the mean free path of the majority of the heat-carrying vibrations is much shorter than the distance between the heat source and heat sink. If this is not the case, the heat-carrying vibrations can propagate ballistically, and Fourier's s law is not valid. To assess such possibility, one has to execute several simulations using simulation cells of varying lengths. If the thermal conductivity is length-independent, the results indeed represent bulk thermal conductivity. Otherwise one can try to extrapolate thermal conductivity to an infinite system size by plotting the inverse of the conductivity as function of the inverse of the system size (Schelling et al. 2002). In the case of amorphous polymers, in which the phonon mean free path is short, size effects are negligible, and the above-described method is accurate and practical. On the other hand, in the case of highly aligned chains, the phonons can propagate by large distances, and the convergence of the direct methods can be challenging and in some cases prohibitive.

The convergence challenge that the direct method suffers in some cases can be alleviated by the use of the equilibrium MD (EMD) (Schelling et al. 2002) techniques based on Green-Kubo formalism. In particular, the Green-Kubo approach is an EMD method that uses fluctuations in the heat current components, $J_{x,y,x}$ (t), to compute the whole thermal conductivity tensor via the fluctuation-dissipation theorem:

$$k_{\mu v}\left(\tau_m\right) = \frac{1}{V k_B T^2} \int_0^\tau \left\langle J_\mu\left(\tau\right) J_v(0)\right\rangle d\tau \tag{8}$$

where τ_m is the upper integral limit that has to be large enough to reach the convergence of the integral and the angular brackets denote a time average of the heat current autocorrelation function. The analytic form of the heat current is determined for the interatomic interactions (see Eq. 7). In this method, there are no heat sources and sinks, and with the use of the periodic boundary conditions, vibrations can propagate across the simulation cell multiple times without scattering, thus eliminating the associated size effects present in the direct method. Also the total thermal conductivity can be broken down into contributions from the autocorrelations of each of the various terms in Eq. 7 plus cross terms involving, e.g., correlations between bond stretching and bond bending. The correlation function in Eq. 8 typically decays exponentially with a time constant yielding the relaxation time of the dominant phonon. Furthermore, in the EMD method has a significant conceptual advantage over the NEMD method since there is no temperature gradient imposed, which eliminates the need to use heat sinks and sources. This leads to a practical advantage that smaller system sizes can be used, albeit typically for
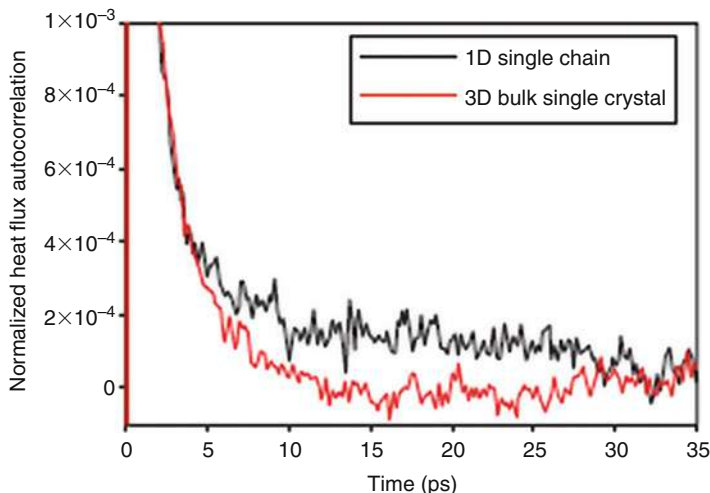
Fig. 4 Heat current autocorrelation function for the perfect crystal PE (red curve) and a single straight PE chain (black curve). While the former function converges to zero, the latter shows persistent positive values indicating non-convergent thermal conductivity given by the integral of the function in Eq. 7. (From Shen et al. 2010)

significantly longer simulation times. Also the fact that a single simulation can provide the whole thermal conductivity tensor is useful for anisotropic materials.

In the case of highly oriented/crystalline polymer fibers, the EMD method is perhaps the only practical method. For example, the MD simulations of Henry and collaborators demonstrated the thermal conductivity of a perfect crystal PE chain can be over ∼100 W/m-K (Shen et al. 2010). However, in the case of single, isolated straight PE chains, even EMD faces convergence problems (Henry and Chen (2008) (see Fig. 4).

4   Interfacial Thermal Resistance

The heat conduction of polymer nanocomposites, or for that matter of any material with high density of interfaces, is significantly affected by the interfacial thermal resistance and known as the thermal boundary resistance, R_{Bd}, or Kapitza resistance. This resistance represents an additional resistance to the heat flow, beyond that due to bulk material. The inverse of R_{Bd} is interfacial thermal conductance, G_k. G_k is an interfacial thermal transport parameter and is quantified in via an equation analogous to Fourier's law (Eq.1):

$$j_q = -G_K \Delta T \qquad (9)$$
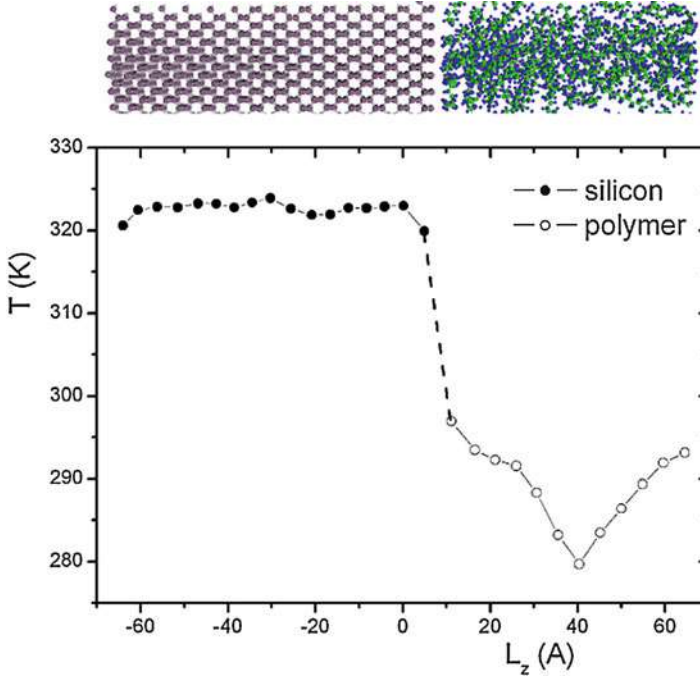
Fig. 5 Temperature profile across silicon-amorphous PE interface – the temperature drop at the interface is due to thermal boundary resistance. (From Hu et al. 2007)

However, the role of the temperature gradient is replaced with the discontinuous temperature at the interface, ΔT (see Fig. 5).

As an alternative to the steady-state heat flow simulations across the interface, one can determine the interfacial conductance by mimicking a laser pump-probe experimental measurement (Shenogin et al. 2004). In this case, one material forming the interface that is nanoscopic in size, such as nanoparticle or nanoscopic film, is subjected to a pulse heating. The observation of the thermal relaxation/equilibration of the system allows to extract the interfacial conductance. As shown in Fig. 6, the temperature difference between the two materials forming the interface exhibits exponential decay. The exponential decay is the result of the heat flow being limited predominantly by the thermal resistance at the interface. Under such conditions, the time constant of the decay, τ, depends on the heat capacity, C, of the nanoscopic component and the interfacial thermal conductance via:

$$\tau = \frac{C}{G_K A},$$ (10)

where A is the interfacial area.

Fig. 6 The exponential decay of the temperature difference between a graphene layer and polymer matrix as a function of time in thermal relaxation molecular dynamics simulations. (From Hu et al. 2011a)
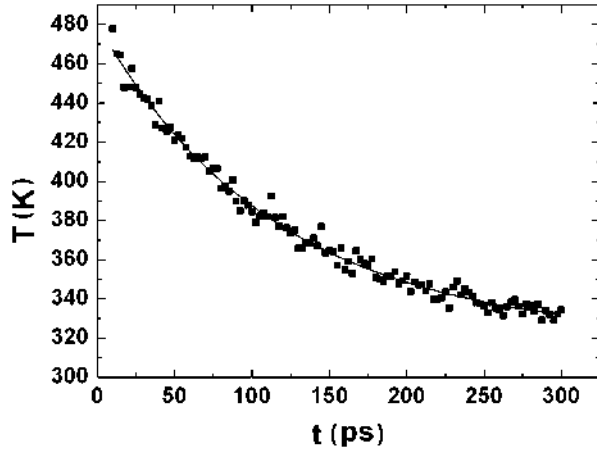
While the heat source-sink method and thermal relaxation method allow for the determination of the value of the conductance, a more detailed understanding of the thermal coupling across the interface can be obtained from equilibrium MD simulations (Barrat and Chiaruttini 2003). Specifically, an interface-specific Green-Kubo relationship is employed, which relates the fluctuating cross interfacial heat power, $P(t)$, to the interfacial conductance via:

$$G_K = \frac{1}{A k_B T^2} \int_0^\infty \langle P(t) P(0) \rangle \, dt, \tag{11}$$

where $\langle P(t)P(0) \rangle$ is the interfacial heat power autocorrelation function and A is the interfacial area. The interfacial heat power can be easily obtained by monitoring the total energy of the material on the one side of the interface and then taking its time derivative. Also, the upper limit of the integral given by Eq. 11 is in practice finite as the autocorrelation function decays to zero. In addition to the value of G_K, the full functional form of $\langle P(t)P(0) \rangle$ provides dynamical information about coupling between molecular vibrations in the two phases forming the interface. As pointed out by Barrat and Chiaruttini, the Green-Kubo relation in Eq. 11 is rigorously valid for an infinite system where the heat capacity of both materials forming the interface is infinite. When the system size is small, the effects of the finite heat capacity need to be accounted for (Barrat and Chiaruttini).

In addition to the finite heat capacity effects, in the limit of extremely small nanoscopic object, such as single-walled carbon nanotube or few atomic layer graphene, more fundamental issue affects the determination of the interfacial thermal resistance. This is illustrated in Fig. 7. When in steady-state direct method simulations, the heat source is placed on graphene, the temperature drop at the interface is very large, and the associated G_K is $\sim$20 MW/m^2-K. Essentially the same value of the G_K is obtained in thermal relaxation simulations (Hu et al. 2011a).
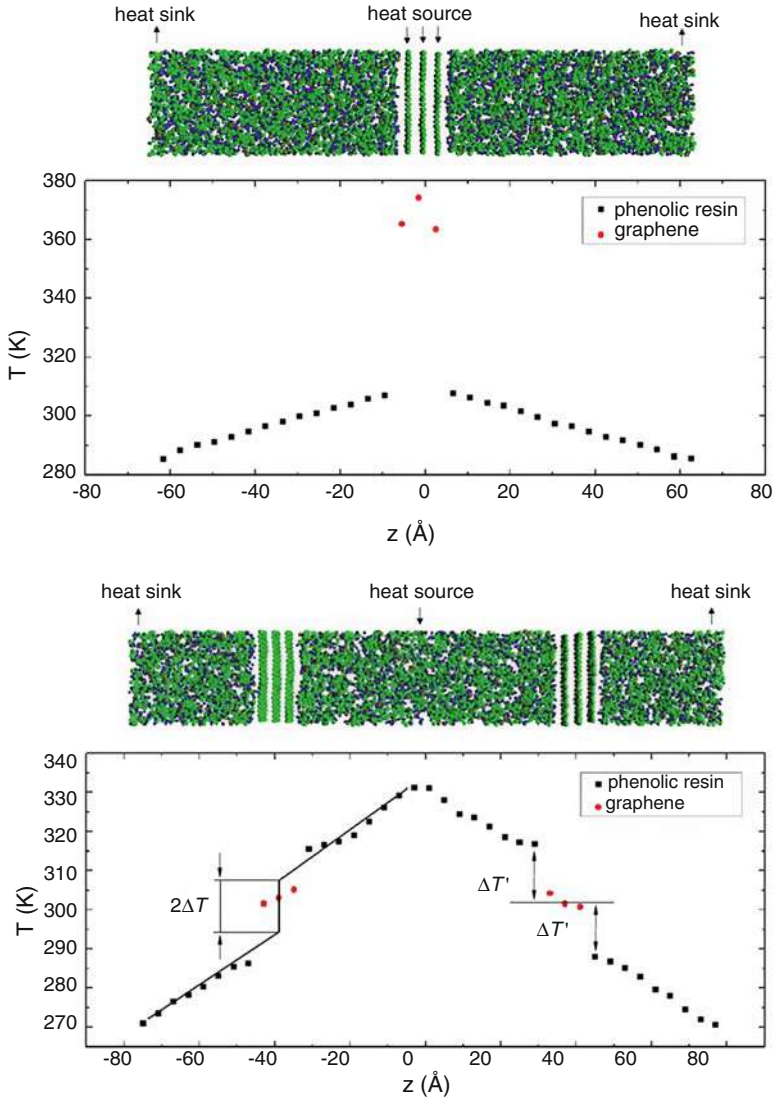
Fig. 7 Left panel: Model structure with a graphene layer in polymeric material along with the steady-state temperature profile obtained from the heat source-sink method, with graphene constituting the heat source. Right panel: Similar to the left panel, but with heat sources and sinks outside graphene. (From Hu et al. 2011a)

By contrast, when the heat sources and sinks are outside graphene, the temperature drop at the interface is much smaller (see Fig. 7), and the associated G_K $\sim$170 MW/m^2-K is much larger.

The above difference is due to the fact that low-frequency vibrations dominate thermal transport across the weakly bonded interface. Consequently, in the simulations where the graphene is the heat source (including thermal relaxation simulations), the interfacial resistance has two main contributors: (i) the internal to graphene resistance between high-frequency modes (that have the majority of thermal energy) and low-frequency modes and (ii) the external resistance between low-frequency modes in graphene (that carry most heat across the interface) and the matrix. When the heat sources and sinks are outside graphene, the internal resistance plays no role in interfacial thermal transport, and in effect, the interfacial conductance is much higher.

The significance of the above finding lies in the fact that the relaxation method that mimics the experimental laser based pump-probe measurements of the interfacial thermal resistance might not provide relevant information about the interfacial thermal transport in terms of predicting and understanding the thermal conductivity of nanocomposites.

Yet another value of the effective interfacial thermal conductance, of about $30\,\mathrm{MW/m^2}$-K, was calculated for the case of heat entering a high aspect ratio carbon nanofiber in order to propagate along the fiber (Hu et al. 2011b) (see Fig. 8). In this case, the heat enters the fiber predominantly via low-frequency modes involving vibrations normal to the interface. However, rapid transport along the carbon fiber is facilitated by acoustic modes with little or none of the normal to the interface displacement. Consequently, the heat flow in this configuration involves energy redistribution over modes within the fiber and is thus affected by associated internal to the fiber phonon-phonon coupling.

All the cases described above show that interfacial thermal transport can involve a complicated process in which the phonons carrying most of the heat across the interface are not necessarily the phonons carrying the heat in the bulk or along the fiber, which in turn makes phonon-phonon coupling within each material forming the interface an important factor. The associated effects are not just artifacts of the simulations, but are present in the pump-probe transient experimental measurements and are relevant to the heat conduction, particularly in the case of high aspect ratio fiber composites.

5   Thermal Conductivity of Polymer Composites

The thermal conductivity of polymer composite is affected by a number of factors, including bulk thermal conductivity of the polymer matrix, interfacial thermal resistance, conductivity of the filler, filler loading, filler shape, and the spatial and orientational distributions of the fillers. Molecular simulation-based methods can provide quality data on bulk and interfacial thermal conductivities, as described above, as well as on the heat flow around individual small nanofillers. However, even such fillers as multiwalled carbon nanotubes (MWCNs) with diameters of tens of nanometers and much larger lengths are much too large to simulate with molecular dynamics.
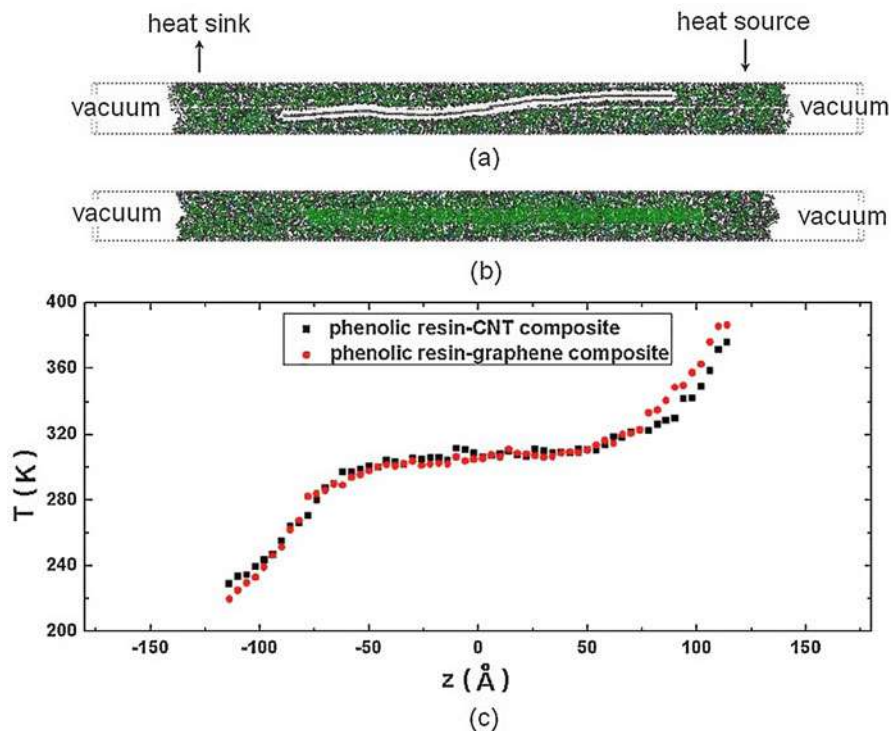
Fig. 8 (**a**) Model structure of resin/graphene composite. (**b**) Model structure of resin/CNT composite. (**c**) The steady-state temperature profile obtained from the nonequilibrium, heat source-sink method. (From Hu et al. 2011b)

To treat larger-scale fillers and assemblies of multiple fillers, one can resort to the solution of the diffusive heat flow equation (Eq. 2). In amorphous polymer cases, the continuum equation holds all the way to the atomic scale as mean free paths of vibrations are very short. For highly conductive fillers, the mean free path can be longer than the filler size, so in principle, the diffusive equation for the heat flow might not be valid. However, in these cases, it might be enough to use the continuum equations, with the stipulation that filler conductivity be much higher than the matrix, as far as the heat flow is mainly limited by the matrix and interfacial thermal resistances.

An example of finite element method (FEM) analysis of the 90° junction between two MWCNs is shown in Fig. 9. This analysis accounted for the interfacial thermal resistance between the tubes and the matrix using surface elements (Shenogina et al. 2005). Interestingly, the above analysis demonstrated that as far as the distance between tubes is much smaller than their length, the separation between the tubes does not affect the rate of the heat flow between the tubes noticeably. Even if tubes are in physical contact, the rate of heat flow is not increased noticeably when contact

Fig. 9 A picture of the meshed model used in an out FEM analysis of two tubes crossed at 90° in a cylinder-shaped calculation box (Shenogina et al. 2005)
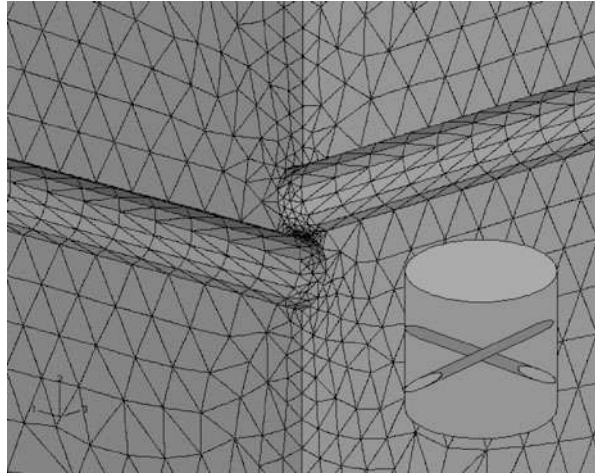
resistance is present. The situation would be different when the contact resistance is negligible, such as the case when the carbon nanotubes are covalently bonded.

The numerical solution of the diffusive heat flow equation can be applied to large filler and its environment and also to multiple filler ensembles. However, a simulation volume that faithfully represents the bulk nanocomposite with a number of associated fillers and their distributions can be prohibitively expensive to treat even with a numerical solution of the continuum heat flow equation. In this case, one can resort to the homogenization treatments, where, e.g., a case of heat flow around a single ellipsoidal inclusion is solved analytically and also analytically averaged over distributions of filler orientations.

To this end, the homogenization treatment due to Nan et al. (1997) is particularly useful in the case of polymer nanocomposites as it accounts for the interfacial thermal resistance. In this treatment, the effective thermal conductivity of the composite along a specific direction, K_{33}^*, is given by:

$$K_{33}^* = K_m \frac{1 + f \left[\beta_{11} \left(1 - L_{11} \right) \left(1 - \langle \cos^2\theta \rangle \right) + \beta_{33} \left(1 - L_{33} \right) \langle \cos^2\theta \rangle \right]}{1 - f \left[\beta_{11} L_{11} \left(1 - \langle \cos^2\theta \rangle \right) + \beta_{33} L_{33} \langle \cos^2\theta \rangle \right]}, \quad (12)$$

with

$$\beta_{ii} = \frac{K_{ii}^c - K_m}{K_m + L_{ii} \left(K_{ii}^c - K_m \right)} \quad (13)$$

$$L_{11} = L_{22}$$

$$= \begin{cases} \frac{p^2}{2(p^2-1)} - \frac{p}{2(p^2-1)^{3/2}} \cosh^{-1} p, & \text{for } p > 1, \\ \frac{p^2}{2(p^2-1)} + \frac{p}{2(1-p^2)^{3/2}} \cos^{-1} p, & \text{for } p < 1, \end{cases} \quad (14)$$

$$K_{ii}^c = K_p / \left(1 + \gamma L_{ii} K_p / K_m\right), \tag{15}$$

with

$$\gamma = \begin{cases} (2 + 1/p)\,\alpha, \text{ for } p \geq 1 \\ (1 + 2p)\,\alpha, \text{ for } p \leq 1 \end{cases}.$$

and

$$\alpha = \begin{cases} a_k/a_1, \text{ for } p \geq 1, \\ a_k/a_3, \text{ for } p \leq 1, \end{cases} \tag{16}$$

where K_m is the thermal conductivity of the matrix, K_p is the thermal conductivity of the ellipsoidal filler particle, and L_{33} are geometrical factors dependent on the particle shape. Other parameters are given as follows: $p = a_3/a_1$ is the aspect ratio of the graphene given by the ratio of longest to shortest radii of the ellipsoid; $\gamma = (2 + 1/p)\alpha$, $a_k = R_{\mathrm{gd}} K_m = K_m/G_k$ is the Kapitza length, i.e., the length over which the temperature drop in the matrix is the same as at the interface; and f is the particle (filler) volume fraction. Finally the triangular brackets in $\langle \cos^2 \theta \rangle$ indicate the orientational average over a specific distribution of fiber orientations, for example, for fully randomly oriented fibers $\langle \cos^2 \theta \rangle = 1/3$ and for fully aligned fibers, along the alignment direction $\langle \cos^2 \theta \rangle = 1$.

Using Eqs. 12, 13, 14, 15, and 16 and the results of molecular dynamics simulations on graphene fiber composites (see Fig. 8), Hu and collaborators (Hu et al. 2011b) extracted the effective interfacial thermal conductance, G_K, for different fiber lengths and obtained consistent value of about 30 MW/m²-K. These results indicate the effective medium theory is indeed capable of an accurate estimate of the composite conductivity for a wide range of aspect ratios, filler volume fractions, and filler orientations.

The effective interfacial thermal conductance of ~30 MW/m²-K is relatively low, and for a typical polymer thermal conductivity, the corresponding Kapitza length is equal to about 10 nm. Even more important for the thermal conductivity of fiber composites is the estimate of the "critical" filler length, L_C, over which the heat flown in to the fiber is dissipated to the matrix, given by (Huxtable et al. 2004):

$$L_c = \sqrt{\frac{rK_p}{G_k}} \tag{17}$$

where K_p and r are the filler conductivity and thickness, respectively. Using $K_p = 1000$ W/K-m, $r = 0.3$ nm, and $G_K = 30$ MW/m²-K, Eq. 17 yields L_C of 100 nm, with the corresponding aspect ratio of about 300. The importance of L_C lies in the fact that for fillers significantly longer than L_C, the interfacial resistance plays a minor role on thermal conductivity, while for fillers shorter than L_C, the conductivity is severely limited by the interfacial resistance. This finding indicates

that for fillers with large enough aspect ratios, the limitation associated with the interfacial resistance can be largely alleviated.

When filler aggregation leads to effectively lower aspect ratio fillers, it is detrimental to thermal transport. To illustrate this point, Fig. 10 shows the effective thermal conductivity for randomly oriented carbon graphene fibers – polymer nanocomposites predicted by Eqs. 12, 13, 14, 15, and 16. In this case, the thermal conductivity is isotropic. Using $K_p = 1000$ W/K-m and $G_K = 30$ MW/m^2-K, the conductivity of the composite at 1% of the graphene volume fraction, as a function of the diameter of the graphene platelets, is shown in Fig. 10 for two cases: (i) single graphene-layer platelets and (ii) five graphene-layer platelets corresponding to ∼2 nm platelet thickness. In both cases, conductivity increases rapidly with an increasing diameter and converges to the common upper limit value given by $1/3K_P f + (1 - f) K_M$ (Nan et al. 1997). However, the upper limit is reached at smaller diameters for single graphene-layer platelets, emphasizing the benefit of good dispersion, resulting in a higher effective aspect ratio.

As discussed above for the very high aspect ratio (∼>500), the interfacial resistance plays no role, and for random filler orientation, the effective thermal conductivity is $1/3K_P f + (1-f) K_M$, while for aligned fillers in the alignment direction, the effective thermal conductivity is $K_P f + (1-f) K_M$. These results are also applicable when the filler forms a well-connected network, i.e., a continuum phase.

By contrast, in the case of spherical and well-dispersed fillers, the homogenization theory in the limit of low volume fractions of nanoparticles, and with $K_p \gg K_m$, predicts:

$$\frac{K_{eff}}{K_p} = 1 + 3f \frac{a_K/r - 1}{a_K/r + 2}, \tag{18}$$
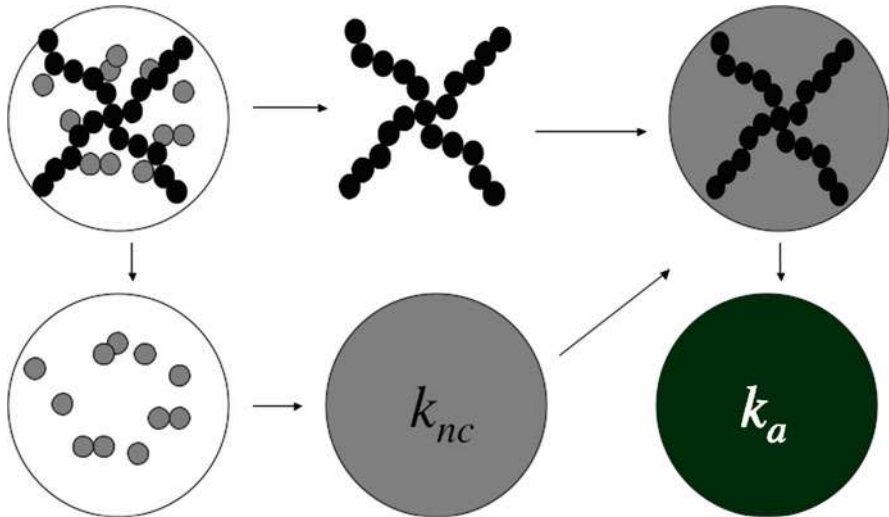
Fig. 11 Schematic of a homogenization approach for fractal particle clusters. The aggregate is decomposed into dead ends with the fluid and the backbone. The effective thermal conductivity of the aggregate with only particles belonging to the dead ends is first calculated using homogenization theory. In the next level of homogenization, linear chains embedded inside a medium with effective conductivity of k_{nc} are considered. (From Prasher et al. 2006)

where a_K is the Kapitza length as before and r is the particle radius. According to Eq. 18, when the particle radius becomes equal to the Kapitza length, there is no thermal conductivity enhancement at all, while for larger interfacial resistance or smaller particle radius, the addition of particles actually decreases the thermal conductivity of the composite.

Homogenization theories can also be applied to more complex situations where small aspect ratio nanoparticles agglomerate, but into sparse fractal patterns. This way they create fillers with an effectively much larger aspect ratio. Such situations can occur in suspensions of solid nanoparticles in fluids (so-called nanofluids) and possibly in polymer resin before curing. Prasher and collaborators (2006) introduced an interesting homogenization treatment for such systems depicted in Fig. 11. According to Fig. 11, a fractal cluster is embedded within a sphere of radius equal to Rg and is composed of a few approximately linear chains, which span the whole cluster (aggregate) and side chains. The linear chains are treated as high aspect ratio fillers, while side chains are treated as well-dispersed particles. Such an approach enables the prediction of thermal conductivity for a variety of particle agglomeration states.

6 Summary and Discussion

In this chapter, an approach to model the heat conduction in polymers and their nanocomposites was presented. Classical molecular dynamics simulations were

presented as the method of choice for the thermal conductivity determination of bulk polymers which combine the computational efficiency and the relative accuracy of the description of atomic structure, chemistry, polymer chain conformation, and bonding. This allows studying the relationship between these structural characteristics and thermal conductivity. The MD-based methods for determining the interfacial thermal conductance were introduced and described in the context of inorganic-polymer interfaces present in the polymer nanocomposites. The incorporation of such determined thermal transport parameters including thermal conductivity of bulk polymers and interfacial thermal resistance into homogenization theories was discussed. These theories allow predicting the role of these parameters, along particle size, shape, volume fraction, distribution of particle orientations, and agglomeration state on thermal conductivity.

While the presented approach constitutes a quality predictive tool for the heat flow analysis in polymers and their composites, a number of challenges and limitations need to be recognized. A standard challenge in MD simulations is the accuracy of describing the interatomic forces. In the case of bulk amorphous polymers, the situation can be quite favorable. The interatomic potentials along with the Monte Carlo-based algorithms allow the preparation and the simulation of the polymer structures which accurately represent real materials. The thermal conductivity predictions for these materials compare favorably with their experimental values. This agreement might be partially due to the fact that some details of the models and atomic interactions might be not that consequential for thermal transport properties. Most likely and good account for the chain conformation (trans-gauche statistics), polymer density and elastic modulus might be sufficient for accurate thermal conductivity predictions.

The situation is more challenging in the cases of highly oriented polymers that exhibit a multifold higher thermal conductivity. The first challenge here is the creation of the quality model structures. Also, in the case of highly aligned and crystalline polymer fibers, the phonon scattering can be very weak and dominated by phonon-phonon scattering, instead of structural scattering. In this case, it is imperative to use both high-accuracy model structures and atomic interaction that would also account for the anharmonic nature of the bonding. However, these bonding parameters are typically not considered when the force fields are designed and parameterized. Nevertheless, the MD simulation studies of the role of chain orientation and their alignments are feasible, particularly to develop correlations, rather than to determine the absolute values of the thermal conductivity.

References

Alexander S, Entinwohlman O, Orbach R (1986) Phonon-fracton interactions: thermal conductivity of amorphous materials. Phys Rev B 34:2726–2734

Ashton JE, Halpin JC, Petit PH (1969) Primer on composite materials: analysis. Technomic, Stamford

Babaei H, Keblinski P, Khodadadi JM (2013) Thermal conductivity enhancement of paraffins by increasing the alignment of molecules through adding CNT/graphene. Inter J Heat Mass Trans 58:209–216

Barrat J-L, Chiaruttini F (2003) Kapitza resistance at the liquid – solid interface. Mol Phys 101:1605–1610

Cahill DG, Watson SK, Pohl RO (1992) Lower limit to the thermal conductivity of disordered crystals. Phys Rev B 46:6131–6140

Che J, Cagin T, Deng W, Goddard W (2000) Thermal conductivity of diamond and related materials from molecular dynamics simulations. J Chem Phys 113:6888–6900

Choy LC, Luk WH, Chen FC (1978) Thermal conductivity of highly oriented polyethylene. Polymer 19:155–162

Duffour E, Malfreyt P (2002) Structure and thermodynamic properties from molecular dynamics simulations of the polyethylene crystal. Polymer 43:6341–6349

Henning J (1967) Anisotropy and structure in uniaxially stretched amorphous high polymers. J Polym Sci C 16:2751–2761

Henry A, Chen G (2008) High thermal conductivity of single polyethylene chains using molecular dynamics simulations. Phys Rev Lett 101:235502

Hsieh W-P, Losego MD, Braun PV, Shenogin S, Keblinski P, Cahill DG (2011) Testing the minimum thermal conductivity model for amorphous polymers with high pressure. Phys Rev B 83:174205

Hu M, Shenogin S, Keblinski P (2007) Molecular dynamics simulation of interfacial thermal conductance between silicon and amorphous polyethylene. Appl Phys Lett 91:241910

Hu L, Desai TG, Keblinski P (2011a) Determination of interfacial thermal resistance at nanoscale. Phys Rev B 83:195423

Hu L, Desai TG, Keblinski P (2011b) Thermal transport in graphene-based nanocomposite. J Appl Phys 110:033517

Huxtable ST, Cahill DG, Phinney LM (2004) Thermal contact conductance of adhered microcantilevers. J Appl Phys 95:2102–2108

Kikugawa G, Desai TG, Keblinski P, Ohara T (2013) Effect of crosslink formation on heat conduction in amorphous polymers. J Appl Phys 114:034302

Kurabayashi K (2001) Anisotropic thermal properties of solid polymers. Inter J Thermophys 22:277–288

Lee S, Rutledge GC (2011) Plastic deformation of semicrystalline polyethylene by molecular simulation. Macromolecules 44:3096–3108

Muller M, Nievergelt J, Santos S, Suter UW (2001) A novel geometric embedding algorithm for efficiently generating dense polymer structures. J Chem Phys 114:9764–9771

Nan CW, Birringer R, Clarke DR, Gleiter H (1997) Effective thermal conductivity of particulate composites with interfacial thermal resistance. J Appl Phys 81:6692–6699

Prasher R, Evans W, Fish J, Meakin P, Phelan P, Keblinski P (2006) Effect of aggregation on thermal conduction in colloidal nanofluids. Appl Phys Lett 89:143119

Schelling PK, Phillpot SR, Keblinski P (2002) Comparison of atomic-level simulation methods for computing thermal conductivity. Phys Rev B 65:144306

Shen S, Henry A, Tong J, Zheng R, Chen G (2010) Polyethylene nanofibres with very high thermal conductivities. Nat Nanotech 5:251–255

Shenogin S, L X, Ozisik R, Cahill DG, Keblinski P (2004) Role of thermal boundary resistance on the heat flow in carbon-nanotube composites. J Appl Phys 95:8136–8144

Shenogin S, Bodapati A, Keblinski P, McGaughey A (2009) Predicting the thermal conductivity of inorganic and polymeric glasses: the role of anharmonicity. J Appl Phys 105:034906

Shenogina N, Shenogin S, Xue L, Keblinski P (2005) On the lack of thermal percolation in carbon-nanotube composites. Appl Phys Lett 87:133106

Sun H (1998) COMPASS: an ab initio force-field optimized for condensed-phase applications-overview with details on alkane and benzene compounds. J Phys Chem B 102:7338–7364

Torii D, Nakano T, Ohara T (2008) Contribution of inter- and intramolecular energy transfers to heat conduction in liquids. J Chem Phys 128:044504

Tucker JE, Reese W (1967) Heat capacity of polyethylene from 2.5 to 30K. J Chem Phys 46:1388

Turney JE, Landry ES, McGaughey AJH, Amon CH (2009) Predicting phonon properties and thermal conductivity from anharmonic lattice dynamics calculations and molecular dynamics simulations. Phys Rev B 80:165304

Wang CZ, Chan CT, Ho KM (1990) Tight-binding molecular dynamics study of phonon anharmonic effects in silicon and diamond. Phys Rev B 42:11276–11283

Ward IM (1971) Mechanical properties of polymers. Wiley, New York, pp 253–262

Ziman JM (1960) Electrons and phonons. Oxford University Press, London

Part VII
Oxides in Energy and Information Technologies

Challenges and Opportunities in Modeling Oxides for Energy and Information Devices

42

Bilge Yildiz, Cesare Franchini, and Jing Yang

Contents

Abstract

The growth of computational resources has enabled investigations of large-scale and highly correlated problems by using first principles computational techniques such as density functional theory (DFT). In context of oxide materials, these problems include oxide surface reconstructions (Diebold et al. 2010), diffusion and reaction barriers in heterogeneous systems (Chizallet and Raybaud 2014; Aksyonov et al. 2018), phase diagrams for transition metal oxides (Park et al. 2014; Leonov 2015), and point defects as well as extended defects (Youssef and Yildiz 2012; Sun et al. 2015). These developments have opened up new

B. Yildiz (✉)
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: byildiz@mit.edu

C. Franchini
Faculty of Physics and Center for Computational Materials Science, University of Vienna, Vienna, Austria
e-mail: cesare.franchini@univie.ac.at

J. Yang
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

opportunities for predicting not only the bulk crystal properties of oxides, but also the effect of complex microstructures such as associated point defects (Hu et al. 2013; Liu et al. 2012; Zhang et al. 2014; T-Thienprasert et al. 2012), grain boundaries (Polfus et al. 2012; McKenna and Shluger 2009; Hojo et al. 2010), dislocations (Sun et al. 2015; Hojo et al. 2011; McKenna 2013), and surfaces (Lee and Morgan 2015; Freysoldt and Neugebauer 2018; Bajdich et al. 2015) under thermodynamic drivers. These developments can ultimately allow for ab initio prediction of realistic device performance. Yet, challenges remain on both the theoretical and algorithmic level to accurately predict oxide materials properties on a complex potential energy surface. Here we summarize several growing fields in addressing these challenges and present our perspectives on future directions that these methods will enable.

1   Methods: Overview and Perspectives

Developing functionals towards chemical accuracy: Developing functionals that accurately describe the exchange and correlation energy has been a continuing interest ever since the development of Kohn-Sham (KS) DFT (Mattsson 2002). For oxides, the use of standard KS-DFT is known to give errors in properties such as electronic structures (Li et al. 2013), formation energies (Kirklin et al. 2015), and surface adsorption energy and reaction barriers (Lousada et al. 2013). These errors hinder the accurate prediction of phase diagrams (Heinemann et al. 2013; Hautier et al. 2012), redox potentials (Chevrier et al. 2010), surface reactivities (Nørskov 2011) for many materials of interest for energy storage and conversion and for information processing. Increasing the accuracy of functionals by moving from local to nonlocal descriptions is called climbing the "Jacob's ladder" (Perdew et al. 2005). To date, five rungs of Jacob's ladder have been extensively investigated and readily applied to oxide solid state systems: local density approximation (LDA), generalized gradient approximation (GGA), meta-GGA, hybrid functionals, and random phase approximation (RPA) (Tran et al. 2016; Perdew and Schmidt 2001; Tao et al. 2003; He and Franchini 2017). Despite the increasing accuracy in treating the exchange-correlation energies and the computational cost paid going up on the ladder, the concurrent DFT methods still fail to accurately treat short-range, intermediate-range, and long-range accurately at the same time (Becke 2014). One of the consequences is that systems with diverse bond types (covalent, metallic as well as ionic, hydrogen, and van der Waals bonding) are hard for DFT prediction. One notorious example is the prediction of molecule adsorption energies on oxide surfaces. Bajdich et al. showed that on alkaline-earth metal oxides surfaces, the best-behaving functional gives errors of up to 0.1 eV compared to the experimentally measured value, and this error is significant for weakly bound molecules (Bajdich et al. 2015). It can be foreseen that improving the accuracy of DFT functionals will still be a major field of interest in the coming years, both by climbing up the Jacob's ladder (Bechstedt 2018) and by developing problem-specific methods, such as DFT + U (Anisimov et al. 1991), SCAN-meta-GGA, and van der Waals schemes (Kulik 2015; Sun et al. 2016; Klimeš and Michaelides 2012).

Treating excited electronic states: Excited state calculations are of interest in many photonic, optoelectronic, and photochemistry processes. In particular, for oxide materials, photoluminescence (Chizallet et al. 2008), carrier transport (Han et al. 2016), and photocatalytic activity (Abe et al. 2013) are important in determining the performance of the material in applications ranging from transparent conducting oxides (Varley et al. 2017) to solar cells (Makhal et al. 2010; Fattori et al. 2010), and photoelectrochemical cells (Zhao et al. 2015). First-principles-based calculation of spectroscopic features which are accessible experimentally, such as in X-ray absorption, photoemission, and electron energy loss spectroscopy, has been helpful in uncovering the photo- and electro-dynamics of chemical bonds and electronic transitions as they take place (di Valentin et al. 2014). Such studies have been demonstrated on various device applications such as photovoltaics (Labat et al. 2012), photocatalysis (Suzuki et al. 2012), and light emitting diodes (Kuttipillai et al. 2016). However, due to challenges in selecting functionals (Leang et al. 2012; Laurent and Jacquemin 2013), convergence on a complex potential energy surface (Plasser et al. 2012), and high computational cost, excited state calculations have been so far limited mainly to simple systems. An archetypal case is the band gap of semiconductor and insulator, a key quantity for virtually all applications, which is not accurately accounted for within standard local and semilocal functionals. Recent advances in quasiparticles approaches in the framework of the GW approximation (Hedin 1965) have proven to be a viable route for a precise prediction of band gaps even in complex oxides (Klimeš et al. 2014; Ergönenc et al. 2018). Another challenging example is that in modeling dye-sensitized solar cells, predicted energy level alignment between TiO_2 and organic dye can vary up to 1 eV with different functionals and thus, cannot lead to a conclusive quantitative result (Pastore et al. 2013). Further method development in this direction is needed, and the most promising approaches for assessing optical and electronic excitations at present include the time-dependent DFT (TD-DFT) (Runge and Gross 1984; Petersilka et al. 1996; Stratmann et al. 1998) and many body Green's function approaches such as solving the Bethe–Salpeter equation (BSE) (Salpeter and Bethe 1951; Onida et al. 2002; Reining et al. 2002). Developments in this direction can allow for applying excited states calculations to obtain optical and electronic properties of surface complexes, hetero-interfaces, and defects, opening up opportunities to tune oxide-based electronic and electrochemical devices by defect and interface engineering (Pentcheva and Pickett 2010; Mannhart and Schlom 2010). An alternative approach to describe the many-body physics of correlated materials is the dynamical mean field approximation (DMFT) (Georges et al. 1996), that in recent years has demonstrated its validity in interpreting optical spectra and optical transitions and is becoming applicable to realistic problems in materials science.

Embedding methods for large-scale problems: In many atomistic modeling problems, a fraction of the system is of special interest and worth being treated with a higher-level theory. The rest of the system serves as its environment and can be treated rather at a lower level. Problems of interest include chemical and electrochemical reactions on oxide catalysts or oxide electrodes that depend on the surface morphology or particle shape, such as corners, facet edges, kink,

surface-terminated dislocations (McKenna 2013; Zheng et al. 2015; Cao et al. 2010; Kleijn et al. 2014; Meyer and Reuter 2014; Stecher et al. 2016) Methods that combine an expansive, more accurate method and a computationally cheap, low-level method are called embedding methods (Sun and Chan 2016). These methods allow for treating large-scale problem while maintaining accuracy for the active region of interest. Depending on the choice of the two levels of methods, embedding methods come with many flavors, such as quantum mechanics/molecular mechanics (QM/MM) (Lin and Truhlar 2006), which combines classical force field with quantum chemistry methods, DFT embedding (Libisch et al. 2014) (different levels of DFT or DFT with many-electron wave function methods), Green's function embedding (Chibani et al. 2016) (DFT with Green's function). Besides the challenges for individual methods as mentioned previously, embedding methods bring about more challenges in terms of developing the embedding potential and coupling the two methods in a reliable and computationally efficient way across their border region.

Including finite temperature effects: Temperature effect is important to be taken into consideration to accurately predict oxide properties in real applications. Accounting for finite-temperature effects is important in many properties of materials, such as heat conduction, phase transitions, and transport properties (Togo and Tanaka 2015; Mankowsky et al. 2014; Serrano et al. 2010; Lee and Rabe 2010). Introduction of temperature effects to couple with first principles calculations involves two lines of effort: (1) explicit free energy terms with all relevant excitation mechanisms such as electronic, quasiharmonic, and anharmonic vibrational terms and (2) explicit account for defect equilibria at finite temperature and their effect on material properties. On the first, calculating anharmonic vibrational contribution generally requires sufficient high-quality DFT force data and is limited mainly to bulk compounds at present (Grabowski et al. 2009; Duff et al. 2015; Zhou et al. 2014; Tadano and Tsuneyuki 2015). Phonon properties and phonon-electron coupling at hetero-interfaces remains a question of interest as well (Sadasivam et al. 2017; Wang et al. 2016). On the second aspect, a successful case is the prediction of point defect equilibria in bulk oxide materials by combining first-principles and statistical mechanics calculations (Youssef and Yildiz 2012; Freysoldt et al. 2014; Youssef et al. 2016). On the other hand, equilibrium structures of extended defects, including surfaces, grain boundaries, and interfaces, and their effects remain largely unexamined for oxide materials. Scaling up of DFT methods, development in advanced sampling techniques (Chua et al. 2010; Harmer 2011), and including the embedding methods discussed above can enable more progress in this context of taking into account temperature effects at extended defects of oxides.

2   Highlights from the Part

Below, we highlight key points of the chapters covered in this section, each demonstrating the state of the art at present in addressing the challenges summarized above. These works include the modeling of charged point defects and hetero-interfaces

from an electronic structure perspective and extending these studies to continuum level equilibria and transport kinetics, ferroelectric, and multiferroic properties, and transport and charge transfer properties in battery and solid oxide fuel cell materials in bulk and at surfaces.

The chapters by Shluger, Reticcioli et al., and Youssef et al. address the modeling of point defects in oxides as relevant to electronic and electrochemical properties. Shluger in ▶ Chap. 43, "Defects in Oxides in Electronic Devices," provides an overview of established models of some of the typical defects in oxides used in microelectronic devices. Defects in oxide dielectric films are often electrically active, capturing and emitting charge carriers from the interfacing materials. Such charge trapping and emitting defects strongly affect the device characteristics, sometimes leading to device failure. This chapter highlights some recent results of theoretical modeling as well as existing problems of computational methods in handling such defects. The properties considered concern mainly nonstoichiometry and disorder of oxides as well as the role of hydrogen. These recent results demonstrate that theoretical modeling and simulation of complex and often defective and disordered thin oxide films is a powerful tool of materials discovery. Importantly, the chapter brings together calculations and results from a range of oxide materials, provides an opportunity to compare models of similar defect species in different materials, and identifies interesting trends that are not obvious from disparate data available in different research fields. The chapter points to the challenges related to lack of transferability of the found defect models from crystalline bulk structures to disordered structures or to surfaces and interfaces. The chapter also points to the lack of direct links from the present static calculations to the dynamics going on in electronic devices, calling for more method development in addressing spatial and time scales that are reachable with first principles methods.

In ▶ Chap. 44, "Small Polarons in Transition Metal Oxides" Reticcioli et al. address the physics of polarons, a quasiparticle defect ubiquitous in polar semiconductors and transition metal oxides which affects the fundamental properties and functionalities of the hosting materials. The chapter begins with a thorough introduction on the concept of polaron, including theoretical accounts (from the seminal model descriptions of Fröhlich and Holstein to a modern view of polarons based on first principles theories) and experimental aspects (how to detect and measure polaron properties). Then, according to the aims of this handbook, the authors describe different types of computational procedures and prescriptions to model the formation, stability, and dynamics of polarons (including charge-trapping techniques, molecular dynamics, statistical analysis and energetics) and provide evidence that this array of schemes is capable to disclose and interpret the role of polarons in the paradigmatic polaronic material TiO_2.

The chapter by Youssef et al. takes the defect studies from single defect properties to establishing equilibria of a multitude of point defects in the bulk and near interfaces of oxides, in ▶ Chap. 45, "Defect Equilibria and Kinetics in Crystalline Insulating Oxides: Bulk and Hetero-interfaces." This is an important step in predicting and controlling the properties of metal oxides under different thermodynamic conditions. This chapter introduces the approach of coupling electronic

structure methods, thermodynamics, electrostatics, and transport theory, to predict defect equilibria and transport kinetics in metal oxides. Equilibria in the bulk is captured by combining first principles calculations of defect formation energies with statistical mechanics, that self-consistently solves for concentration of defects and the established electron chemical potential as a function of temperature and oxygen chemical potential. The framework is extended to interfaces, by calculating defect energies at the interface as well, and connecting the interface to bulk by solving for the Poisson-Boltzmann formalism. The kinetics of defect transport is captured by using concentrations and migration barriers found from first principles in random walk and kinetic Monte Carlo calculations. The demonstrations of this framework are shown on ZrO_2 as a model system for insulating oxides. In the end, the chapter points to challenges in implementing this framework to metallic oxides which are typically strongly correlated systems whose electronic structure and thus defect properties are difficult to capture accurately. In addition, in more complex oxides with multiple elements and a large range of defect types, automated algorithms would be beneficial to systematically analyze the possible defects and pinpoint the most important ones. Finally, the chapter calls for studying the effect of other thermodynamic forces, including electromagnetic fields and nonhydrostatic stresses.

▶ Chapter 46, "Oxide Heterostructures from a Realistic Many-Body Perspective" by Lechermann et al., focuses on materials design from a many-body perspective. The materials at stake are oxide heterostructures and the computational methodology is based on DMFT. The creation of oxide heterostructures represents a relatively new paradigm in condensed matter research which enables a controlled design of new materials and the onset of novel quantum phases arising from (i) the interplay between the diverse electronic characteristics of the bulk constituents and (ii) the presence of interface regions. In this chapter, Lechermann shows that this complex class of materials can be reliably described using DMFT. The authors offer a focused presentation of the concept of strong electronic correlation in solids, conveys a concise and well-documented introduction to DMFT, and discusses the applicability of DMFT to two exemplary interfaces: band-band insulators and Mott-band insulators.

The importance of interface effects is also addressed in ▶ Chap. 47, "First-Principles Modeling of Interface Effects in Oxides" by Demkov et al., in the context of single-particle DFT. The authors review several important aspects related to the first principles modeling of interfaces (supercell approach, band alignment, construction of the structural models, dipole corrections, defect energy-levels) and validate the reliability of these schemes to technologically relevant interfaces based on transition metal oxides materials as well as Si-based structures. Here the focus is to provide a proper account of several measured properties such as electron energy loss spectra, scanning transmission electron microscopy images, core-level x-ray photoemission spectra.

In ▶ Chap. 48, "Design of New Multiferroic Oxides" Lu and Reticcioli discuss a fundamentally and technologically relevant class of materials, multiferroics. After providing a classification of multiferroic systems in terms of the specific

mechanism enabling the coupling among different ferroic orders (spontaneous ferroelectricity, ferromagnetism, and ferroelasticity), the authors focus their presentation on theoretical and computational protocols necessary to design novel materials with enhanced ferroic properties. These design principles are based on chemical considerations (crystal stricture and composition), electronic properties (the system must exhibit an insulating regime), lattice dynamical analysis (polar phonons), and spin properties (magnetic exchange interactions). The authors show that all these aspects can be assessed in the framework of DFT (suitably extended to account for electronic correlation effects), which provides a computationally reliable scheme to interpret the coexisting ferroic responses and phase stability in realistic situations.

The chapter by Wang et al addresses larger scale simulations by phase field method, which can input thermodynamic data deduced from first principles calculations, in a multiscale manner. The chapter takes ferroelectric domains, ▶ Chap. 49, "Strain Control of Domain Structures in Ferroelectric Thin Films: Applications of Phase-Field Method." Phase-field method has been extensively applied to understanding the underlying physics of the experimentally observed domain structures and predicting their responses to external electrical, mechanical, thermal, and chemical stimuli. This chapter reviews the fundamentals of the thin-film phase-field method and its applications in predicting the effects of strains on the phase transitions, domain structures, and the domain switching. The chapter also presents the prospects of using the phase-field method in microstructure design and property optimization for ferroelectric thin films. Of particular interest to note here is that, this approach has only been applied to a limited number of materials due to the lack of the thermodynamic parameters including the Landau coefficients, electrostrictive coefficients, elastic constants, and gradient energy coefficients. Obtaining Landau coefficients, the polarization, and dielectric responses as a function of temperature experimentally is challenging. First-principles or atomistic calculations can fill this gap, providing the needed input to phase field calculations to address a broader range of ferroelectric oxide materials.

The chapters by Chu et al. and by Kozinsky review the advances brought to energy storage materials by first principles modeling techniques. Ong and Meng and coauthors, in ▶ Chap. 50, "Battery Electrodes, Electrolytes, and Their Interfaces" introduce how the well-established ab initio modeling methods were implemented widely in studying and predicting battery materials, electrodes, electrolytes and their interfaces, from both the thermodynamic and kinetics aspects. The thermodynamic approaches include the computations of equilibrium voltage and voltage profiles, electronic structure, and stability analyses. The kinetics approaches cover methods for ionic diffusion studies, including transition state theory and nudged elastic band method, and ab initio molecular dynamics. This chapter presents an illustrative demonstration of the need for close-loop interaction of first-principles calculations with experimental design and data. The authors conclude by defining their perspective on the pressing challenges in modeling noncrystalline systems, e.g., the presence of intrinsic defects and the electrolyte/electrode interfaces, and the design of high-voltage high-capacity

cathode materials. They call for developments in the areas of multiscale approaches to bridge the multiple time and length scales, and automation software and machine learning tools to facilitate and accelerate the studies of battery materials.

In line with the challenge and need described in ▶ Chap. 51, "Transport in Frustrated and Disordered Solid Electrolytes," Kozinsky addresses ion transport studies in highly defective and disordered materials for Li-ion conduction in solid state. The aim is to reveal the descriptors or design principles for enabling high ionic conductivity in inorganic materials at room temperature. This chapter assesses the effects of strong ion-ion interaction, geometric frustration, disorder, and collective motion in ionic motion and illustrates applications of atomistic computational tools in combination with techniques of statistical mechanics, group theory, and computational geometry for analyzing dynamics and transport mechanisms in the context of Li-ion conducting oxides. In particular, the chapter demonstrates that ab initio molecular dynamics in combination with coarse graining is a very useful tool in resolving ionic transport mechanisms in ceramic electrolytes. A key result is that the overall ability of the Li sublattice configuration to rearrange with low transition barriers in the oxide is found as the necessary ingredient of high ionic conductivity. This requires symmetry breaking and maximizing the amount of geometric frustration in the Li sublattice, preventing the system from settling into a low-entropy arrangement. The chapter calls for systematic consideration of the complex regimes of high Li-ion concentration and strong interaction in different classes of Li-ion conductors.

In ▶ Chap. 52, "Solid Oxide Fuel Cell Materials and Interfaces" Hess et al. present computational studies of the bulk and surface chemistry of perovskite oxides that are widely used as catalysts and electrocatalysts in energy conversion devices. Here the specific focus has been the operating conditions of solid oxide fuel cells. For these materials, while the ion and electron transport properties in their bulk are generally well understood, the surfaces remain much less explored and challenging. Depending on the temperature, oxygen pressure, and electrochemical potential, the surface and near-surface region of doped perovskite oxides can significantly deviate from bulk, in terms of chemistry, atomic structure, and even phase composition. It is known in the field that such deviations are detrimental to surface catalytic properties under operating conditions. The chapter first presents a short description of the bulk defect chemistry of perovskite oxides and highlights the challenges for both theory and experiment in treating these materials. The authors then present the work from the field in investigating surface instabilities and their potential reasons, including electronic reconstruction, ionic segregation, and the formation of stable AO surface terminations. Finally, the chapter examines oxygen adsorption and dissociation reactions on different terminations of selected perovskite oxides. The outlook in this field is to advance the computational frameworks to capture both the equilibrium and dynamic structure of surfaces of complex oxides, a yet unresolved field but immensely important for a range of catalytic and energy conversion applications.

References

Abe R et al (2013) Visible-light-induced water splitting based on two-step photoexcitation between dye-sensitized layered niobate and tungsten oxide photocatalysts in the presence of a triiodide/iodide shuttle redox mediator. J Am Chem Soc 135(45):16872–16884

Aksyonov DA et al (2018) Understanding migration barriers for monovalent ion insertion in transition metal oxide and phosphate based cathode materials: a DFT study. Comput Mater Sci 154:449–458

Anisimov VI, Zaanen J, Andersen OK (1991) Band theory and Mott insulators: Hubbard U instead of Stoner I. Phys Rev B 44(3):943–954

Bajdich M, Nørskov JK, Vojvodic A (2015) Surface energetics of alkaline-earth metal oxides: trends in stability and adsorption of small molecules. Phys Rev B 91(15):155401

Bechstedt F (2018) Correlation beyond the random phase approximation: a consistent many-body perturbation theory approach. Phys Rev B 97(24):241109

Becke AD (2014) Perspective: fifty years of density-functional theory in chemical physics. J Chem Phys 140(18):18A301

Cao A, Lu R, Veser G (2010) Stabilizing metal nanoparticles for heterogeneous catalysis. Phys Chem Chem Phys 12(41):13499–13510

Chevrier VL et al (2010) Hybrid density functional calculations of redox potentials and formation energies of transition metal compounds. Phys Rev B 82(7):075122

Chibani W et al (2016) Self-consistent Green's function embedding for advanced electronic structure methods based on a dynamical mean-field concept. Phys Rev B 93(16):165106

Chizallet C, Raybaud P (2014) Density functional theory simulations of complex catalytic materials in reactive environments: beyond the ideal surface at low coverage. Catal Sci Technol 4(9):2797–2813

Chizallet C et al (2008) Assignment of photoluminescence spectra of MgO powders: TD-DFT cluster calculations combined to experiments. Part I: structure effects on dehydroxylated surfaces. J Phys Chem C 112(42):16629–16637

Chua ALS et al (2010) A genetic algorithm for predicting the structures of interfaces in multicomponent systems. Nat Mater 9:418

Diebold U, Li S-C, Schmid M (2010) Oxide surface science. Annu Rev Phys Chem 61(1):129–148

di Valentin C, Botti S, Cococcioni M (Eds) (2014) First principles approaches to spectroscopic properties of complex materials, Top Curr Chem vol 347. Springer, Berlin/Heidelberg

Duff AI et al (2015) Improved method of calculating ab initio high-temperature thermodynamic properties with application to ZrC. Phys Rev B 91(21):214311

Ergönenc Z et al (2018) Converged GW quasiparticle energies for transition metal oxide perovskites. Phys Rev Mater 2(2):024601

Fattori A et al (2010) Fast hole surface conduction observed for indoline sensitizer dyes immobilized at fluorine-doped tin oxide–TiO_2 surfaces. J Phys Chem C 114(27):11822–11828

Freysoldt C, Neugebauer J (2018) First-principles calculations for charged defects at surfaces, interfaces, and two-dimensional materials in the presence of electric fields. Phys Rev B 97(20):205425

Freysoldt C et al (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86(1):253–305

Georges A et al (1996) Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. Rev Mod Phys 68(1):13–125

Grabowski B et al (2009) Ab initio up to the melting point: anharmonicity and vacancies in aluminum. Phys Rev B 79(13):134106

Han D et al (2016) Phonon-enabled carrier transport of localized states at non-polar semiconductor surfaces: a first-principles-based prediction. J Phys Chem Lett 7(18):3548–3553

Harmer MP (2011) The phase behavior of interfaces. Science 332(6026):182–183

Hautier G et al (2012) Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability. Phys Rev B 85(15):155208

He J, Franchini C (2017) Assessing the performance of self-consistent hybrid functional for band gap calculation in oxide semiconductors. J Phys Condens Matter 29(45):454004

Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139(3A):A796–A823

Heinemann M, Eifert B, Heiliger C (2013) Band structure and phase stability of the copper oxides Cu_2O, CuO, and Cu_4O_3. Phys Rev B 87(11):115111

Hojo H et al (2010) Atomic structure of a CeO_2 grain boundary: the role of oxygen vacancies. Nano Lett 10(11):4668–4672

Hojo H et al (2011) Atomic structure and strain field of threading dislocations in CeO_2 thin films on yttria-stabilized ZrO_2. Appl Phys Lett 98(15):153104

Hu W et al (2013) Electron-pinned defect-dipoles for high-performance colossal permittivity materials. Nat Mater 12:821

Kirklin S et al (2015) The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. Npj Comput Mater 1:15010

Kleijn SEF et al (2014) Electrochemistry of nanoparticles. Angew Chem Int Ed 53(14):3558–3586

Klimeš J, Michaelides A (2012) Perspective: advances and challenges in treating Van Der Waals dispersion forces in density functional theory. J Chem Phys 137(12):120901

Klimeš J, Kaltak M, Kresse G (2014) Predictive GW calculations using plane waves and pseudopotentials. Phys Rev B 90(7):075125

Kulik HJ (2015) Perspective: treating electron over-delocalization with the DFT+U method. J Chem Phys 142(24):240901

Kuttipillai PS et al (2016) Phosphorescent nanocluster light-emitting diodes. Adv Mater 28(2):320–326

Labat F et al (2012) First-principles modeling of dye-sensitized solar cells: challenges and perspectives. Acc Chem Res 45(8):1268–1277

Laurent AD, Jacquemin D (2013) TD-DFT benchmarks: a review. Int J Quantum Chem 113(17):2019–2039

Leang SS, Zahariev F, Gordon MS (2012) Benchmarking the performance of time-dependent density functional methods. J Chem Phys 136(10):104101

Lee Y-L, Morgan D (2015) Ab initio defect energetics of perovskite (001) surfaces for solid oxide fuel cells: a comparative study of $LaMnO_3$ versus $SrTiO_3$ and $LaAlO_3$. Phys Rev B 91(19):195430

Lee JH, Rabe KM (2010) Epitaxial-strain-induced multiferroicity in $SrMnO_3$ from first principles. Phys Rev Lett 104(20):207204

Leonov I (2015) Metal-insulator transition and local-moment collapse in FeO under pressure. Phys Rev B 92(8):085142

Li W et al (2013) Density functional theory and beyond for band-gap screening: performance for transition-metal oxides and dichalcogenides. J Chem Theory Comput 9(7):2950–2958

Libisch F, Huang C, Carter EA (2014) Embedded correlated wavefunction schemes: theory and applications. Acc Chem Res 47(9):2768–2775

Lin H, Truhlar DG (2006) QM/MM: what have we learned, where are we, and where do we go from here? Theor Chem Accounts 117(2):185

Liu L et al (2012) p-Type conductivity in N-doped ZnO: the role of the NZn-VO complex. Phys Rev Lett 108(21):215501

Lousada CM et al (2013) Reactivity of metal oxide clusters with hydrogen peroxide and water – a DFT study evaluating the performance of different exchange–correlation functionals. Phys Chem Chem Phys 15(15):5539–5552

Makhal A et al (2010) Role of resonance energy transfer in light harvesting of zinc oxide-based dye-sensitized solar cells. J Phys Chem C 114(23):10390–10395

Mankowsky R et al (2014) Nonlinear lattice dynamics as a basis for enhanced superconductivity in YBa2Cu3O6.5. Nature 516:71

Mannhart J, Schlom DG (2010) Oxide interfaces – an opportunity for electronics. Science 327(5973):1607–1611

Mattsson AE (2002) In pursuit of the "divine" functional. Science 298(5594):759–760

McKenna KP (2013) Electronic and chemical properties of a surface-terminated screw dislocation in MgO. J Am Chem Soc 135(50):18859–18865

McKenna KP, Shluger AL (2009) First-principles calculations of defects near a grain boundary in MgO. Phys Rev B 79(22):224116

Meyer J, Reuter K (2014) Modeling heat dissipation at the nanoscale: an embedding approach for chemical reaction dynamics on metal surfaces. Angew Chem Int Ed 53(18):4721–4724

Nørskov JK et al (2011) Density functional theory in surface chemistry and catalysis. Proc Natl Acad Sci 108(3):937–943. https://doi.org/10.1073/pnas.1006652108

Onida G, Reining L, Rubio A (2002) Electronic excitations: density-functional versus many-body Green's-function approaches. Rev Mod Phys 74(2):601–659

Park H, Millis AJ, Marianetti CA (2014) Total energy calculations using DFT+DMFT: computing the pressure phase diagram of the rare earth nickelates. Phys Rev B 89(24):245133

Pastore M, Fantacci S, De Angelis F (2013) Modeling excited states and alignment of energy levels in dye-sensitized solar cells: successes, failures, and challenges. J Phys Chem C 117(8):3685–3700

Perdew JP, Schmidt K (2001) Jacob's ladder of density functional approximations for the exchange-correlation energy. AIP Conf Proc 577(1):1–20

Perdew JP et al (2005) Prescription for the design and selection of density functional approximations: more constraint satisfaction with fewer fits. J Chem Phys 123(6):062201

Petersilka M, Gossmann UJ, Gross EKU (1996) Excitation energies from time-dependent density-functional theory. Phys Rev Lett 76(8):1212–1215

Plasser F et al (2012) Electronically excited states and photodynamics: a continuing challenge. Theor Chem Accounts 131(1):1073

Polfus JM et al (2012) Defect chemistry of a BaZrO3 Σ3(111) grain boundary by first principles calculations and space-charge theory. Phys Chem Chem Phys 14(35):12339–12346

Reining L et al (2002) Excitonic effects in solids described by time-dependent density-functional theory. Phys Rev Lett 88(6):066404

Pentcheva R, Pickett WE (2010) Electronic phenomena at complex oxide interfaces: insights from first principles. J Phys Condens Matter 22(4):043001

Runge E, Gross EKU (1984) Density-functional theory for time-dependent systems. Phys Rev Lett 52(12):997–1000

Sadasivam S et al (2017) Thermal transport across metal silicide–silicon interfaces: first-principles calculations and Green's function transport simulations. Phys Rev B 95(8):085310

Salpeter EE, Bethe HA (1951) A relativistic equation for bound-state problems. Phys Rev 84(6):1232–1242

Serrano J et al (2010) Phonon dispersion relations of zinc oxide: inelastic neutron scattering and ab initio calculations. Phys Rev B 81(17):174304

Stecher T, Reuter K, Oberhofer H (2016) First-principles free-energy barriers for photoelectrochemical surface reactions: proton abstraction at TiO2 (110). Phys Rev Lett 117(27):276001

Stratmann RE, Scuseria GE, Frisch MJ (1998) An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules. J Chem Phys 109(19):8218–8224

Sun Q, Chan GK-L (2016) Quantum embedding theories. Acc Chem Res 49(12):2705–2712

Sun L, Marrocchelli D, Yildiz B (2015) Edge dislocation slows down oxide ion diffusion in doped CeO2 by segregation of charged defects. Nat Commun 6:6294

Sun J et al (2016) Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. Nat Chem 8:831

Suzuki S, Tsuneda T, Hirao K (2012) A theoretical investigation on photocatalytic oxidation on the TiO2 surface. J Chem Phys 136(2):024706

Tadano T, Tsuneyuki S (2015) Self-consistent phonon calculations of lattice dynamical properties in cubic SrTiO3 with first-principles anharmonic force constants. Phys Rev B 92(5):054301

Tao J et al (2003) Climbing the density functional ladder: nonempirical meta – generalized gradient approximation designed for molecules and solids. Phys Rev Lett 91(14):146401

Togo A, Tanaka I (2015) First principles phonon calculations in materials science. Scr Mater 108:1–5

Tran F, Stelzl J, Blaha P (2016) Rungs 1 to 4 of DFT Jacob's ladder: extensive test on the lattice constant, bulk modulus, and cohesive energy of solids. J Chem Phys 144(20):204120

T-Thienprasert J et al (2012) Identification of hydrogen defects in SrTiO3 by first-principles local vibration mode calculations. Phys Rev B 85(12):125205

Varley JB et al (2017) High-throughput design of non-oxide p-type transparent conducting materials: data mining, search strategy, and identification of boron phosphide. Chem Mater 29(6):2568–2573

Wang X, Zebarjadi M, Esfarjani K (2016) First principles calculations of solid-state thermionic transport in layered Van Der Waals heterostructures. Nanoscale 8(31):14695–14704

Youssef M, Yildiz B (2012) Intrinsic point-defect equilibria in tetragonal ZrO_2: density functional theory analysis with finite-temperature effects. Phys Rev B 86(14):144109

Youssef M, Yang M, Yildiz B (2016) Doping in the valley of hydrogen solubility: a route to designing hydrogen-resistant zirconium alloys. Phys Rev Appl 5(1):014008

Zhang G, Lu Y, Wang X (2014) Hydrogen interactions with intrinsic point defects in hydrogen permeation barrier of α-Al_2O_3: a first-principles study. Phys Chem Chem Phys 16(33):17523–17530

Zhao Y et al (2015) Understanding the effect of monomeric iridium(III/IV) aquo complexes on the photoelectrochemistry of $IrO_x \cdot nH_2O$-catalyzed water-splitting systems. J Am Chem Soc 137(27):8749–8757

Zheng Y et al (2015) Advancing the electrochemistry of the hydrogen-evolution reaction through combining experiment and theory. Angew Chem Int Ed 54(1):52–65

Zhou F et al (2014) Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations. Phys Rev Lett 113(18):185501

Defects in Oxides in Electronic Devices

43

Alexander Shluger

Contents

Abstract

This chapter provides an overview of established models of some of the typical defects in oxides used in microelectronic devices, highlighting some recent results of theoretical modeling as well as existing problems of computational methods. In numerous electrical measurements, defects in oxide dielectric films have been demonstrated to be electrically active, meaning that they can capture and emit charge carriers from the substrate as well as the gate. These defects, commonly referred to as charge traps, can strongly affect the device characteristics or even impair their functionality causing device failure. Predictive theoretical modeling and simulation of growth and properties of complex and often disordered thin oxide films became a powerful tool of materials discovery where new sophisticated methods of computer experiments

A. Shluger (✉)
Department of Physics and Astronomy, University College London, London, UK
e-mail: a.shluger@ucl.ac.uk

play an important part in designing and screening new materials and studying and predicting their properties.

1   Introduction

Defects play a crucial role in the performance and reliability of microelectronic devices, optical fibers, transparent conductors, catalysts, sensors, and many other systems. Achieving a better understanding of defect structures and processes has been a driving force behind the development of many powerful theoretical concepts and methods. For example, extensive studies of color centers in the 1960s driven by radiation physics and chemistry helped to shape point defect, polaron, and exciton theories and to develop a wide range of scintillators, laser materials, and dosimeters. Doping of semiconductors, which proved to be vital for microelectronics industry, has long been an inspiration for developing computational solid-state methods. Problems related to the mechanisms of silicon oxidation, reliability of microelectronic devices, Bragg grating formation in optical fibers, luminescence of porous silicon, and others promoted extensive studies of amorphous materials. More recent examples concern the development of new high-k dielectrics for microelectronics applications, transparent conductive oxides, materials exhibiting magneto-resistant properties, high-T_c superconductors, and various applications of quantum dots. Again, most of the work is performed in experimental labs where various combinations of materials are tried and their properties tested. At the same time, predictive theoretical modeling and simulation of growth and properties of complex and often disordered thin oxide films became a powerful tool of materials discovery where new sophisticated methods of computer experiments play an important part in designing and screening new materials and studying and predicting their properties.

A significant component of this research, close to the interests of the authors of this chapter, is related to defects in oxides. They are responsible for leakage current, breakdown and degradation of microelectronic devices, Bragg grating formation, conductivity of electro-active insulators, chemical sensing, and many other processes. Defects are likely to play even more important roles as devices continue to scale and reach dimensions where single defects can affect their performance.

In this chapter we provide an overview of established models of some of the typical defects in oxides, highlighting some recent results of theoretical modeling as well as existing problems of computational methods. We will focus mainly on oxides relevant for applications in microelectronics. This choice is mainly determined by our assertion that these materials are likely to remain in focus of active research for the next few years. However, we hope that this review will provide a useful source of models and understanding of theoretical modeling in a wider area of wide-gap oxides.

Successful development of new materials for gate oxides and transparent conducting oxides (TCOs) poses difficult problems for theoretical modeling. Some of the typical problems include predicting stable structures and electronic properties of Si/dielectric systems. This implies calculations of band offsets; possible interface

defects, such as dangling bonds, Si–Si bonds, hydrogen defects, etc.; studies of the structure and properties of grain boundaries (they can, e.g., accumulate vacancies); and electronic states due to disorder if the dielectric is amorphous. Another important issue concerns predicting the structure and energy levels of possible defects induced by growth and post-deposition annealing in different ambient conditions, for example, due to incorporation of boron, chlorine, oxygen, nitrogen, hydrogen, and other species. Determining the most stable charge states of interface and bulk defect species and their possible role in static charging of as-grown and annealed oxides has long been one of the main challenges of theoretical modeling. An even more difficult problem concerns transient electron and hole traps responsible for leakage current and degradation of oxide films and possible ways of eliminating them. This includes predicting defect diffusion and re-charging in applied electric fields. Defect-induced conductivity is the key issue for TCOs, such as indium tin oxide (ITO). Theory should also help in identifying defects by predicting their vibrational, EPR, optical, and other properties.

Figure 1 gives an overview of some of the processes relevant to gate oxide films: creation of defects, incorporation from the environment and creation during growth; charge transfer, defects can change their charge state due to electron and hole trapping; reactions, defects react with each other to form different defects; and defect diffusion. Theoretical models discussed in this chapter may help to
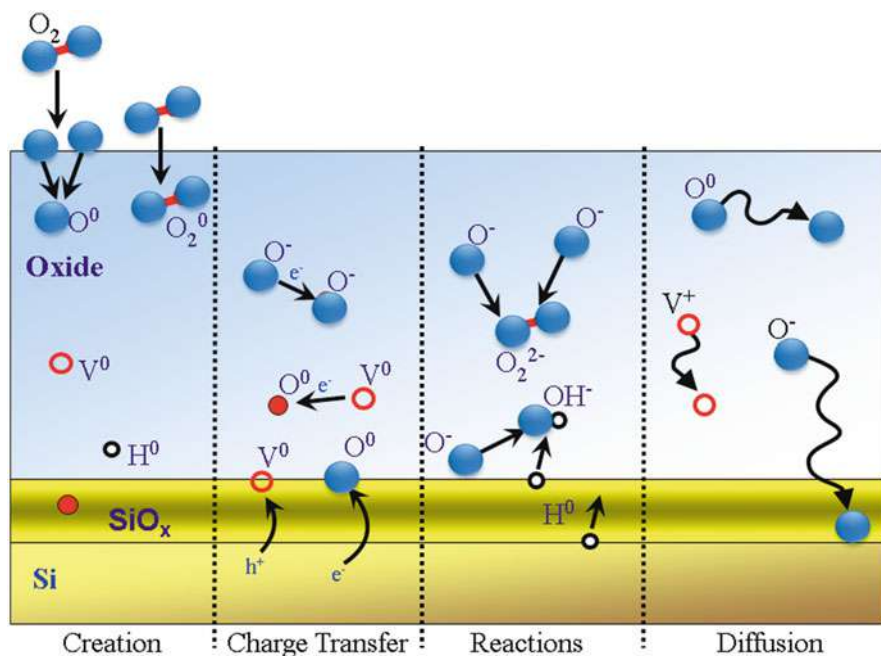
Fig. 1 Diagram showing the different defect processes which can be important to the properties to gate oxide films. These include defect creation during growth and post-deposition anneal, electron transfer processes between defects and from/to silicon, reactions between defects leading to formation of more complex species and annihilation of existing defects, and defect diffusion

understand the mechanisms of these processes as well as the structure and electronic properties of defects as a function of oxide composition.

The initial stage of studying defects in a new material concerns building structural models of defect species. For example, hydrogen, oxygen, and nitrogen species can incorporate into the lattice during growth or annealing of oxide films in atomic and molecular forms in different charge states. Another class of problems is related to energy levels of vacancies and incorporated species. As-grown films are almost always nonstoichiometric. In relatively narrow-gap oxides, such as In_2O_3, SnO_2, and $Sr(Ba)TiO_3$, oxygen vacancies are the source of electron doping, which makes them n-type conductors. On the other hand, electron and hole trapping by interstitial oxygen and oxygen vacancies from Si in gate oxides may lead to oxide charging and affect leakage current. Charged defects can create strong electric fields and affect band offsets. They can contribute to dielectric loss and their diffusion in an electric field is equivalent to electrolysis. Building a complete picture of defect processes also includes defect reactions, requiring the parameters of electron transfer between defects, calculations of defect diffusion, and recombination. More complex modeling concerns the dynamics of defect processes, mechanisms of energy dissipation, and ultimately absolute rates of these processes.

In many cases, reliable and predictive theoretical screening of complex oxide materials takes much longer than the decision regarding their irrelevance for particular applications reached experimentally. Therefore computer experiments to investigate characteristics related to basic electronic properties, processing, and operational performance of these materials are carried out mainly on prototype systems. In the case of high-k dielectrics, these include ZrO_2, HfO_2, and $HfSiO_4$. The best-studied prototype TCOs include SnO_2 and In_2O_3. SiO_2 has long served and will still continue to serve for some time as a gate dielectric and the material of optical fibers and optical components in lithography. Therefore many of our examples will concern basic defects in this material. Finally, there are many relevant studies of electron and hole defects in the bulk and at surfaces of MgO, some ferroelectrics and in other oxide materials. They allow one to outline some general defect models and trends from their properties. For this chapter we will focus mainly on oxygen- and hydrogen-related defects in these materials due to their ubiquitous nature and general relevance.

Below we will give a brief overview of defects and defect processes relevant for electronic devices. Many of these defects and related processes have been previously studied in bulk, crystalline oxides. Therefore we will discuss the transferability of defect models developed in model oxides to other more complex materials and will consider the effect of disorder in amorphous materials on defect properties.

This chapter inevitably provides only a very specific slice of an extensive and vibrant area of research. Recent excellent reviews related to the topics outlined here may help to broaden the picture. In particular, defects in cubic oxides have been considered in Henderson (1980), Hayes and Stoneham (1985), and Kotomin and Popov (1998). Extensive reviews of defects in SiO_2 in relation to Bragg grating formation in optical fibers and in gate dielectrics can be found in Pacchioni et al. (2000), Lenahan and Conley (1998), and Fleetwood et al. (2008). Advances in

the theory of oxide-semiconductor interfaces have been discussed in Edwards and Fowler (1999) and more recently in extensive collections of papers (Massoud et al. 2000; Jupille and Thornton 2015; Netzer and Fortunelli 2016). The role of oxide defects in reliability of microelectronic devices has recently been discussed in Fleetwood et al. (2008) and Grasser (2014). Defects in wide bandgap II-VI crystals and in transition metal oxides have been reviewed in Neumark (1997), Pacchioni (2000), and Hofmann et al. (2007). Models of radiation-induced point defects and oxygen vacancies in MgO and several ferroelectric perovskites have been considered in Kotomin and Popov (1998), Islam (2000), Kotomin et al. (2002), and Popov et al. (2010). Methods and challenges for atomic-scale modeling of electronic materials and high-k dielectrics have been reviewed in Zunger (1998), Kawamoto et al. (2000, 2002), and Freysoldt et al. (2014). A significant part of research related to defects in oxides is carried out in relation to diverse applications of oxide surfaces in catalysis and in solid electrolyte fuel cells as reviewed in Gellings and Bouwmeester (2000) and Jupille and Thornton (2015).

2 Calculation of Defect Properties

2.1 Formation Energies

In thermodynamic equilibrium, the probability of creating an individual defect D in the charge state q is determined by the defect formation energy Freysoldt et al. (2014):

$$E_{for}(D) = E_D^q - E_0 - \sum_{i=1} n_i \mu_i + q E_F + E_{corr}, \tag{1}$$

where $E_{for}(D)$ is calculated as the total energy difference between the system with the defect E_D^q and the system without the defect E_0. The integer $n_i > 0$ indicates the number of atoms of type i, which can be the host atoms missing or impurity atoms added to the system to form the defect, and the μ_i are the corresponding chemical potentials of these species. The Fermi energy E_F is analog of the chemical potential of the electrons. E_{corr} is a correction term that accounts for finite size of the periodic cell used for calculating the system with the defect.

Defining the chemical potentials of the atomic species in a compound crystal is impossible without further information. Usually, the atomic chemical potentials are set to reflect the experimental conditions, which define the properties of the reservoirs of atoms in equilibrium with the system. If, for example, the sample of MgO has been annealed in an oxygen atmosphere, the chemical potential of the oxygen atoms is the same as in gaseous molecular oxygen at the experimental partial pressure and temperature. The chemical potential of the oxygen atoms in well-annealed oxide in thermodynamic equilibrium is thus strongly dependent on temperature and pressure. However, most calculations do not consider the *process* of formation of a vacancy or incorporation of an atom or molecule. For example,

during annealing of an oxide film in oxygen atmosphere, oxygen molecules can dissociate into atoms at the surface or due to irradiation. However, in most cases calculations are concerned only with final possible state of an atom or molecule in the bulk of material in thermodynamic equilibrium. If atoms are available directly from an oxygen atomic source, then an atomic reference for energy E_O is used. If an O_2 molecule can dissociate into atoms in the process of incorporation, then half of the molecular energy is taken as a reference to account for the dissociation costs. Thus the formation energy gives just a reference for the difficulty of forming a defect in the oxide – the higher the energy, the more difficult it is to create.

2.2   Electrical Levels

Yet another issue, specific for device applications, is that the oxide of interest can be in contact with the electrode(s) – semiconducting (e.g., silicon) or metallic – which set a Fermi level E_F in the system and provide a source of extra electrons. Also an oxide film can be placed in the electric field between the two electrodes, which may alter the defect charge states. To define which charge state of an intrinsic defect (e.g., vacancy) or incorporated defect specie is favored, assuming the common case of oxide film grown on silicon, the chemical potential of the electron can be chosen at the bottom of Si conduction band or at the silicon midgap energy (Blöchl 2000). Electrons can tunnel elastically or inelastically from/to these states to/from defect states in the oxide and change their occupation creating charged defects. Since the interface is not explicitly included in most calculations, one needs to use the experimental information (Alay and Hirose 1997; Mihaychuk et al. 1999) (e.g., 4.6 eV for the valence band offset at the Si/SiO$_2$ interface, 1.1 eV for the bandgap of Si) to estimate the energy of an electron at the bottom of the conduction band of Si with respect to the theoretical zero energy level (see Fig. 2). The calculated energies are readily adjusted if doping, temperature, or bias favors another source of electrons.

To study the dependence of the defect charge state on the position of the system Fermi energy, it is useful to define the thermodynamic charge transition level, E_{th}. The $E_{th}(q_1/q_2)$ is defined as the Fermi-level position for which the formation energies of charge states q_1 and q_2 are equal. Both charge states can coexist near the charge transition level energy.

Once the charge transition levels are defined, the dependence of the defect charge state on the electrode Fermi energy can be introduced via a correction term qE_F, where E_F is varied in the energy range from the top of the oxide valence band to the bottom of the conduction band (providing the bandgap is correctly reproduced). The results are presented as plots of $E_{for}(D)$ vs. E_F calculated for different values of E_F. The crossing points of the corresponding lines determine the E_F where one charge state of a defect becomes more favorable than the other, i.e., charge transition levels. Although the exact position of the Fermi energy can be unknown in some experiments, such diagrams allow one to identify some critical cases and help in the analysis of the experimental data. An example of such a plot is shown below in Fig. 5.
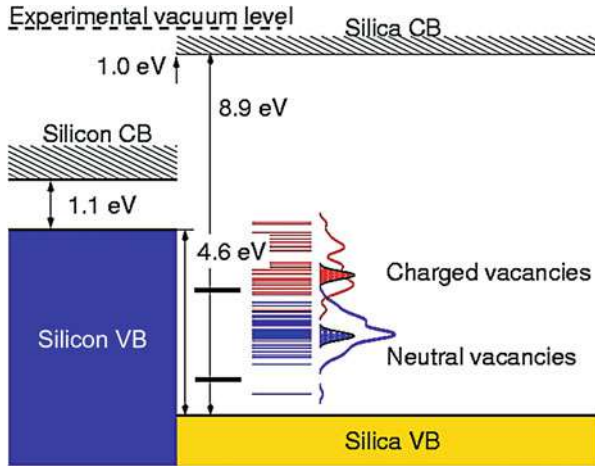
Fig. 2 Schematic band offset diagram showing the relative energies of the top of the valence bands and the bottom of the conduction bands of the bulk materials in contact (for the Si/SiO$_2$ interface Alay and Hirose 1997; Mihaychuk et al. 1999). The figure shows two discrete black levels in the silica bandgap, which correspond to the positions of occupied states of neutral and positively charged oxygen vacancy in α-quartz. To the right from these levels are schematically shown distributions of levels of neutral and positively charged vacancies in a-SiO$_2$: discrete levels correspond to the finite number of vacancies in a periodic cell and continuous distributions to a macroscopic sample. Note that the zero energy level in theoretical calculations of bulk materials is poorly defined and can be even below the top of the calculated valence band

3   Oxygen Vacancies

Oxide films in electronic devices are often nonstoichiometric on oxygen, and their structure includes oxygen vacancies. Oxygen vacancies are likely to be the most prevalent defects in many oxides and may have several charge states. They are characterized by formation energies, position of defect levels with respect to the valence and conduction bands, structural models, spectroscopic properties (including optical absorption and luminescence energies, parameters of EPR spectra, and IR and Raman spectra), mechanisms of diffusion, electron transfer and recombination with other defects, and other characteristics. The study of vacancies in bulk oxides has a long history (e.g., see reviews in Pacchioni 2000, Popov et al. 2010, Hofmann et al. 2007), and here we will try to only highlight those aspects which are of specific interest. First, we will introduce some general models and will illustrate them by several examples. Then we will consider in more detail a case study: oxygen vacancies in hafnia.

Dependent on material and particular application, vacancies can be classified as electron donors or hole traps. Their most stable charge state depends on whether the oxide is in contact with Si or other electron source and on applied voltage. Another type of classification is based on structural models. It provides a rough but very

simple cross section through different oxides in terms of two common types: F, centers, and $Me\bullet$, centers. The main difference between the two structural models discussed below is in the character of localization of electrons left by removed oxygen and of the lattice relaxation.

F-like centers are known to form in cubic alkali halides and oxides, such as MgO, CaO, and BaO (Pacchioni 2000). In these ionic materials, the oxygen site is surrounded by six equivalent metal ions. A classical F-center will have one (alkali halides) or two (cubic oxides) electrons in an s-type deep state in the bandgap. Formation of this defect is accompanied by small and fully symmetric (O_h point group) relaxation of the surrounding ions. A significant amount of the electron density is localized in the vacancy with considerable and equal contributions of neighboring metal ions (see Fig. 3a). Ionization of the F-center will lead to characteristic strong and, again, symmetric distortion of surrounding ions, where the nearest metal ions are moving outward from the vacancy and the 12 next nearest-neighbor anions move inward.

MgO represents perhaps the most studied of oxides, and therefore we will use it as a reference for other oxide materials. In the context of electronic devices, it is used, e.g., in magnetic tunnel junction devices. The properties of neutral and charged oxygen vacancies in MgO have been studied experimentally in great detail (Chen et al. 1969; Hofmann et al. 2007). It is well established that vacancies in MgO can be described by an F-center model (Hayes and Stoneham 1985; Pacchioni 2000) (see also Fig. 3a). As discussed previously, a neutral vacancy possessing two electrons is called an F-center, and a positively charged vacancy, which keeps only one electron, is called an F^+ center, respectively. The paramagnetic F^+ center has been well described using ESR and ENDOR (Henderson and Wertz 1977). The unpaired electron of the F^+ center is shown to be well-localized in the vacancy
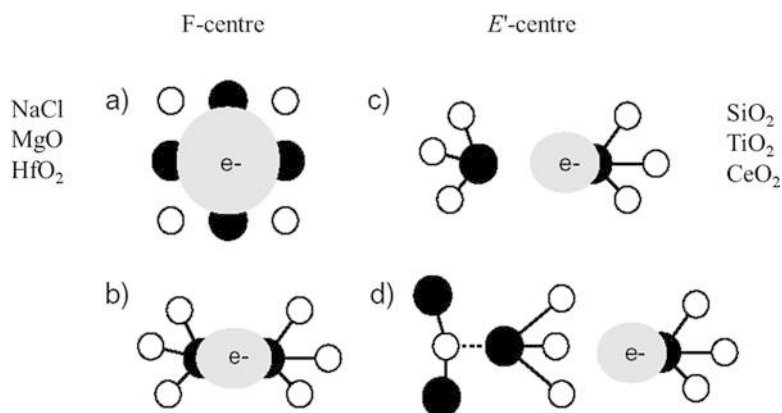
Fig. 3 Schematic of different types of oxygen vacancies in insulators. Cations are shown in black and anions in white. (**a**) An F-like center defect characterized by symmetric electron distribution centered in the vacancy (shown in gray); (**b–d**) various types of $Me\bullet$ – center models (see text for discussion). Again electron localization is shown schematically in gray

with only small hyperfine interaction with the surrounding Mg nuclei (Henderson and Wertz 1977). Both centers have close optical absorption energies -4.92 and 5.01 eV for the F and F^+ center, respectively (Chen et al. 1969). Surface analogues of bulk F-centers have long been considered to be part of various surface processes (see, e.g., the recent review Pacchioni 2000).

Another characteristic oxygen vacancy type is represented by what may be called a $Me\bullet$ model. In this case the electrons left in the vacancy reside on the neighboring Me ions rather in the vacancy, in contrast with the F-center model. Oxygen vacancies of this type are formed in different phases of SiO_2, GeO_2, and in silicates, where each oxygen ion is surrounded by only two nearest neighbor positive ions. Formation of a neutral oxygen vacancy in these materials is accompanied by the creation of a $Me--Me$ bond and by strong displacements of these two Me ions toward each other (see Fig. 3b). In TiO_2 and CeO_2 electrons occupy Me ions in the vicinity of the vacant site, but there is no $Me--Me$ bond formation.

The model of a neutral vacancy in SiO_2 is in marked contrast with cubic oxides, such as MgO and the more complex HfO_2 discussed below. In those materials, despite the substantial ionic polarization, the formation of a neutral oxygen vacancy leads to only very small displacements of the surrounding metal atoms. In the case of MgO, ionic displacements are isotropic and at a distance of 8.0 Å from the neutral vacancy are less than 0.002 Å. A double positively charged anion vacancy in MgO produces a much stronger perturbation, but again, the displacements of the lattice ions decrease relatively rapidly and do not exceed 0.02 Å at a distance 8.0 Å from the vacancy. This results from the different structure and chemical bonding in these materials. In particular, in MgO, unlike α-quartz, the oxygen site is the center of inversion. The peculiar structure of α-quartz is also manifested in the strong asymmetry of the relaxation of the two Si atoms neighboring the vacancy (see Sulimov et al. 2002).

The E'-center is perhaps the most famous defect related to oxygen deficiency in silica, which also plays a very important role in many applications of silica in microelectronics and fiber optics. The E'-center is thought to be a positively charged oxygen vacancy. Ionization of the neutral vacancy leaves a relatively weak bond between two silicon atoms. Flexibility of the amorphous structure gives room to formation of several centers. In particular, the E'_γ-center identified by EPR in a-SiO_2 (Griscom et al. 1974; Griscom 2000) has an unpaired spin localized on a dangling bond of a silicon atom (see Fig. 3c, d). This defect has an optical absorption band at 5.85 eV with line-width (FWHM) of 0.8 eV and an oscillator strength f $= 0.14\pm0.1$. This absorption yields no luminescence. The EPR spectrum of E'_1-center in α-quartz is similar to that of the E'_γ center in amorphous SiO_2, as shown in Weeks (1956) and Weeks (1994). However, in spite of some analogies between models of vacancy defects in α- quartz and amorphous SiO_2, it is becoming increasingly clear that there are significant differences too. Some of them stem from the fact that, in contrast with crystalline α-quartz, there are various types of a-SiO_2, which differ by preparation, treatment, water content, and other parameters.

Having discussed the "classical" F-center and $Me\bullet$ case, we will now briefly review models of oxygen vacancies in two other oxides widely used in devices.
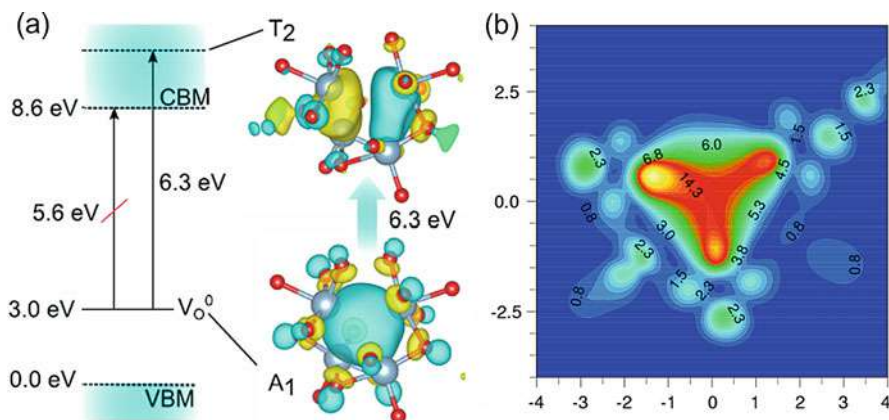
Fig. 4 Structural models and spin density distribution in O vacancies in α-Al$_2$O$_3$ and monoclinic HfO$_2$. (**a**) An energy level diagram for the neutral vacancy in α-Al$_2$O$_3$ showing the symmetry forbidden and allowed optical transitions and the molecular orbitals of the A$_1$ and T$_2$ states involved in the symmetry allowed excitations. (**b**) Spin density in a plane containing three Hf ions neighboring the three-coordinated oxygen vacancy (V) with one electron removed from the system. Black arrows show directions of displacements of Hf ions neighboring the vacancy resulting from the electron removal. Charge density is in 0.1 e/Å^3 and all distances are in Å

Sapphire or α-Al$_2$O$_3$ is a representative of a wide class of ceramic materials important for applications in microelectronics and laser optics, as a substrate, and in many other areas. Each Al atom in this material is bonded to six oxygen ions, and each oxygen ion is surrounded by four nearest neighbor Al ions in the form of distorted tetrahedron. Recent ab initio calculations (Hine et al. 2009; Weber et al. 2011) demonstrated that oxygen vacancies in α-Al$_2$O$_3$ form an F-center type defect (see Fig. 4a). However, the relatively low symmetry of the vacancy and mixed ionic-covalent bonding of alumina means that there is significant relaxation (up to 16% of the original bond lengths) of the Al and O ions around the vacancy. The neutral vacancy, F-center, introduces a doubly occupied deep level of about 3.0 eV above the valence band edge, whereas the position of the F^+-center state is predicted to be only 0.25 eV lower. Analysis of the electron density shows that the ground state of the F-center is strongly localized inside the vacancy with large contributions on the surrounding Al and O ions (see Fig. 4a). An energy level diagram for the neutral vacancy in α-Al$_2$O$_3$ in Fig. 4a shows the symmetry forbidden and allowed optical transitions, and the molecular orbitals of the A$_1$ s-type and T$_2$ p-type states involved in the symmetry allowed optical excitations.

Hafnium oxide, HfO$_2$, is widely used in microelectronics as a gate dielectric as well as in memory cells and other applications. Together with ZrO$_2$ they represent the case of mixed ionic-covalent type of chemical bonding and materials with a relatively high ($\approx$20) dielectric constant. Oxygen vacancies in monoclinic hafnia can form at either a four-coordinated tetragonally or three-coordinated trigonally bonded lattice oxygen site. The spin density distribution inside the positively

Fig. 5 Calculated formation energies of the singular O vacancy in monoclinic HfO$_2$ as a function of the electron chemical potential with the origin corresponding to the VB maximum. Charge states are indicated in the inset
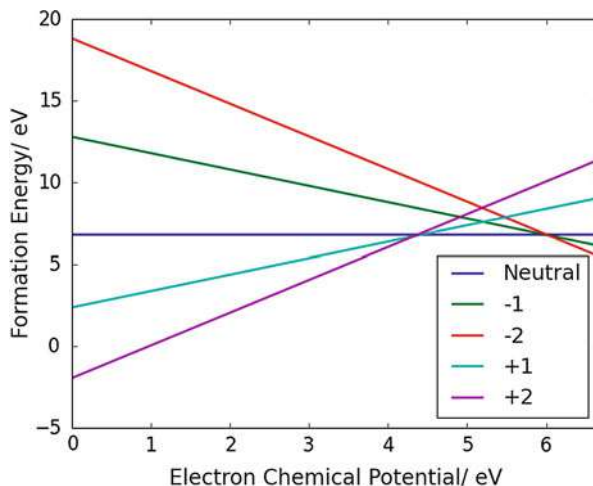
charged vacancy in Fig. 4b shows strong electron localization inside the vacancy and on the d–orbitals of Hf ions surrounding the vacancy. Due to the high dielectric constant and d-character of the conduction band they can exist in both positively and negatively charge states (Muñoz Ramo 2007). The formation energy diagram as a function of Fermi-level position in Fig. 5 shows that vacancies prefer to be positively charged at Fermi-level position below 4.3 eV and may become negative at Fermi-level position close to the bottom of the conduction band. The lattice relaxation around neutral vacancies involves small displacements of the nearest neighbor Hf ions. The displacements of these ions for both types of vacancies are in the range of 0.01–0.02 Å, which corresponds to 0.5–1.0% of the HfO bond length. Such small displacements are characteristic for the F-center-type defects. In the negatively charge state, extra one or two electrons are localized on the Hf ions surrounding the vacancy (Muñoz Ramo 2007).

These results provide examples of relatively complex systems where periodic DFT calculations allow one to routinely investigate models and ground state electronic properties of defects in different charge states.

4   Defects at Interfaces

Devices are systems of interfaces between different materials. Electronic properties of thin oxide films grown on Si and other substrates vary widely with their morphology and with the morphology of the interface. The oxide film morphology, in turn, is very sensitive to small variations in chemical composition and to deposition and post-deposition processing. Most as-deposited thin oxide films are amorphous, and annealing at high temperature leads to their partial or full crystallization and formation of polycrystalline films. Structural disorder in amorphous films and grain

boundaries in polycrystalline films help to reduce strain developing at the interface due to misfit of structural parameters of interfacing materials.

Over the years atomistic modeling proved to be instrumental in understanding the structure and defect properties of silicon/oxide and metal/oxide interfaces. The earlier works on metal/oxide interfaces emphasized the role of image interactions, which has been later included in atomistic simulations. More recent work is using DFT and molecular dynamics to predict interface structures. There are broadly three approaches employed for constructing semiconductor/oxide and metal/oxide interfaces. All of them employ a periodic model along the interface plane and therefore are subject to constraints when trying to accommodate two materials in a periodic cell of certain dimension.

One of the most common approaches combines lattice matching with anneal using molecular dynamics. This mimics a growth process and gives additional freedom for the interface atoms to relax. Powered by several types of classical interatomic potentials, it has been used to construct models including more than ten million atoms (Bachlechner et al. 1998). Reactive force fields have been developed and applied to modeling interfaces between crystalline Si and amorphous SiO_2 (van Duin et al. 2003), often accompanied by DFT simulations. Similar approach using DFT and molecular dynamics to melt and quench a-SiO_2 layer sandwiched between two crystalline Si substrates has been used to model a $Si/SiO_2/Si$ systems and more recently $Ge/GeO_2/Ge$ and $Si/SiO_2/HfO_2$ stacks.

The most popular approach to constructing metal/oxide interfaces is using static DFT calculations for systems constructed by lattice matching (Tse et al. 2010) for different values of oxygen pressure (Netzer and Fortunelli 2016). The main criterion of the interface stability is the interface free energy as a function of oxygen pressure (Zeng et al. 2014). In particular, static DFT calculations have been performed to model the formation of the interface between silicon and strontium titanate and its atomic structure (Forst et al. 2004). In this case the order of deposition is important, and this study shows that atomic control of the interfacial structure by altering the chemical environment can dramatically improve the electronic properties of the interface to meet technological requirements.

Another approach is based on simulating the film growth via atomic layer deposition (ALD) or chemical vapor deposition (Dkhissi et al. 2008, 2009). Other studies, focused on the interface growth kinetics, suggested that HfO_2 film growth on Si may be substantially non-epitaxial and would strongly depend on precursor chemistry.

Although atomistic modeling to date had established several important structure-property dependencies, the main challenge remains in identification and characterization of the realistic interface models as opposed to simplified (and possibly even unstable) structures and characterization of interface defects. Electrically active interface defect states may lead to degradation of device characteristics, such as the threshold voltage, the on-current, or the surface carrier mobility. Creation of an interface naturally leads to breaking some of the bonds in the interfacing materials. To further improve the interface quality, the number of dangling bonds can be further reduced by annealing the system in forming gas with hydrogen atoms. The purpose
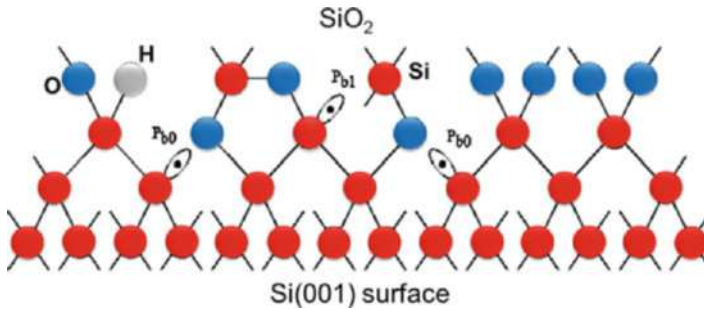
Fig. 6 Schematic showing P_{b0} and P_{b1} defects located at the interface between an amorphous SiO_2 film and Si substrate with the (100) orientation. The P_{b0} center is back-bonded to three silicon atoms and electrically very similar to the P_b center. The silicon atom of the P_{b1} center is back-bonded to two other silicon atoms and an oxygen atom. Both traps are electrically active as their energy lies in the silicon bandgap making their charge state Fermi-level dependent

of annealing is to reduce the number of structural defects formed during the film growth and to passivate dangling bonds.

Figure 6 shows a schematic of Si/SiO_2 interface. The structure of the interface depends on the Si surface exposed to the oxidation (Massoud et al. 2000; Fleetwood et al. 2008; Grasser 2014). P_b centers are dangling Si bonds remaining at the interface after the oxidation. P_{b0} and P_{b1} centers in Fig. 6 are located at the interface between an amorphous SiO_2 film and the (100) Si substrate. The P_{b0} center is back-bonded to three Si atoms. The silicon atom of the P_{b1} center is back-bonded to two other silicon atoms and an oxygen atom. Both centers are electrically active as their energies lie in the silicon bandgap making their charge state Fermi-level dependent. These dangling silicon bonds are usually passivated by hydrogen after anneal, forming Si-H bonds. With this treatment the amount of electrically active interface states can be reduced to $\approx 10^{10}\,cm^{-2}\,eV^{-1}$. The exact properties of the interface defects, which are trivalent silicon atoms with one unpaired valence electron $Si_3 \equiv Si^{\bullet}$, or $Si_2O \equiv Si^{\bullet}$, depend on the atomic configuration and on the orientation of the substrate. While P_b centers are formed on (111) oriented substrates, P_{b0} and P_{b1} centers exist when Si (100) substrates are used.

It is instructive to consider an example of a more complex $Si/SiO_2/HfO_2$ interface shown in Fig. 7. To construct this periodic cell, a Si/SiO_2 interface was generated in a large supercell of $15 \times 15 \times 40\,\text{Å}$. The bottom silicon (001) layer was passivated by hydrogen and kept constrained during subsequent geometry optimizations using DFT calculations (Gavartin and Shluger 2007). The rest of the system, including the fully hydroxylated SiO_2 surface, was relaxed using simulated temperature annealing followed by static relaxation. Then a block of the monoclinic HfO_2 was added at the top of the structure, such that each hydrogen atom at the SiOH surface was replaced by the Hf at the bottom of HfO_2. The top surface of the HfO_2 block was also hydroxylated to eliminate dangling bonds. Subsequent simulated high temperature annealing leads to a dramatic relaxation of the interface manifested

Fig. 7 The atomistic
structure of a periodic cell
used to calculate the
electronic structure and defect
properties in HfO$_2$/SiO$_2$/Si
stack. In DFT calculations,
the cell is translated in the
direction along the HfO$_2$
surface. Formation energies
of 43 neutral oxygen
vacancies at various positions
between the Si and HfO$_2$
parts of the stack are shown
in the graph below the
structure. The rightmost
defects are directly at the
HfO$_2$/SiO$_2$ interface and the
leftmost directly at the
SiO$_2$/Si interface
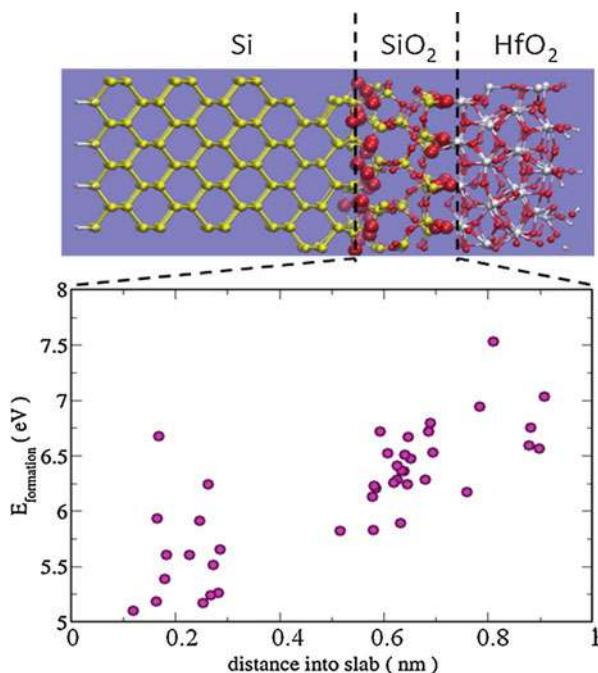
by highly correlated oxygen diffusion from the bulk HfO$_2$ toward the interface.
Then the number of Si layers was increased with respect to the original structure to
converge the bandgap of the Si substrate to 1.2 eV. The geometric structure of the
whole stack has been optimized using the hybrid density functional. The thickness
of the SiO$_2$ layer is about 0.65 nm and that of the HfO$_2$ layer is about 1 nm, close to
the technologically relevant structures.

Defects in the SiO$_2$ and HfO$_2$ layers affect reliability of devices and have been
studied extensively. However, the *nm* thickness of oxide layers, confinement, and
strain at interfaces change the defect characteristics with respect to the bulk oxide.
In addition, disorder in the amorphous SiO$_2$ layer leads to a site dependence and
distribution of defect properties. This is illustrated in Fig. 7 where the formation
energies of a sample of O vacancy defects at sites throughout the SiO$_2$ part of the
model interface were calculated. There is a considerable spread in the formation
energies of oxygen vacancies at different sites within the stack. Importantly, O
vacancies tend to segregate at the Si/SiO$_2$ interface.

5   Hydrogen Defects

The lightest element, hydrogen, is probably also the most abundant, as well as
the least detectable impurity in the dielectric oxides. It is also believed to be the
most damaging for the dielectric properties of the gate stack – leading to interface

charging (Afanas'ev and Stesmans 1998), stress-induced leakage currents, and ultimately an electrical breakdown of the dielectric films (Stathis 2002; Fleetwood 2002; Degreave et al. 1999; Fleetwood et al. 2008; Grasser 2014). However, in traditionally used Si/SiO$_2$ systems, hydrogen also plays a positive role in passivating Si dangling bonds, as discussed in the previous section.

Abundance of hydrogen in a dielectric film arises from it being present throughout the manufacturing process. For example, hydrogen is abundantly available during atomic layer deposition of HfO$_2$ and subsequent device processing. It can also be injected into the oxide during device operation and electrical stress from gate stack or from the spacer regions and remains trapped in the film in various stable forms, not all of which are easily detectable. For example, molecular hydrogen does not usually interact strongly with the dielectric lattice and may be identified by Raman spectroscopy. Neutral hydrogen atoms can be detected in the EPR (electron paramagnetic resonance) measurements, but the complications arise from a high mobility of hydrogen atoms. In contrast, various hydroxide ions (OH)$^-$ and water molecules can be probed using the FTIR (Fourier-transform infrared spectroscopy) techniques, although an ambiguity often arises as to where from in the sample (device) an IR signal originates.

Due to difficulties in experimental detection, most of the evidence of the involvement of hydrogen in the electric degradation of CMOS devices is indirect (Fleetwood et al. 2008; Grasser 2014). Though technology primarily aims at passivation of existing defects by hydrogen to improve the device performance, it may also induce new defects or charges detrimental to the reliability. For example, most of the charging phenomena in much better studied insulating a-SiO$_2$ films are associated with hydrogen-related network fragments. Oxygen vacancies in a-SiO$_2$ can effectively trap hydrogen resulting in, e.g., a hydrogen bridge center where hydrogen replaces the bridging oxygen. As discussed earlier, anion vacancies have a large electron affinity, and so does the hydrogen bridge center.

The interaction of hydrogen with transition metal oxides is important for a wide range of applications and has also been studied extensively. DFT investigations of thermodynamic stability and charge transition levels of hydrogen in stoichiometric monoclinic HfO$_2$ and ZrO$_2$ as a function of the Fermi-level position using nonlocal DFT functionals concluded that H is a negative-U system, with the $+1$ or -1 charge states lower in energy for all relevant Fermi-level values (Godet et al. 2007; Li and Robertson 2014; Lyons et al. 2011). In the positive charge state, the hydrogen interstitial is bonded to a threefold-coordinated oxygen atom, with a bond length of 0.98 Å. Negatively charged H$^-$ ion sits interstitially and is coordinated by two (Li and Robertson 2014) or three Hf atoms (Lyons et al. 2011). Such behavior is characteristic to many other crystalline wide-gap oxides (Li and Robertson 2014).

The dynamics of interaction of hydrogen with crystalline oxides is complex, and its understanding has been aided tremendously by the so-called oxide muonics, where muonium is used as a pseudoisotope of hydrogen, as summarized in Cox et al. (2006) and Cox (2003). In a nutshell, hydrogen atom injected into in wide-gap perfect bulk oxide can be (meta)stable and diffuse through the lattice. If the so-called $+/$ pinning level (Li and Robertson 2014) is above the bottom of the conduction

band or the bandgap is smaller than about 5.5 eV, atomic H will eventually donate
an electron into the conduction band and become a proton and bind to the lattice
oxygen ion. If there are defects in bulk material, such as oxygen vacancies, grain
boundaries, and misfit dislocations, atomic hydrogen can donate its electron to these
defects and again form an OH bond at one of the lattice sites. If the Fermi level set
by external electrodes or dopants is high enough, it will eventually trap an extra
electron and become H^-.

The final result depends on the character and concentration of electron trapping
sites in a particular system. For example, atomic hydrogen does not react with
terraces at the MgO (001) surface, but donates electrons to three-coordinated Mg
sites at corners and kinks at that surface, creating new $[Mg^+\!\!-O\text{-}H]$-type centers
(Ricci et al. 2003). The recent study of amorphous silica has concluded that,
unlike in α-quartz, atomic H can break some strained Si-O bonds and form a
new thermodynamically stable defect, termed hydroxyl E' center (El-Sayed et al.
2015). In this case, the electron is localized on a Si atom and a proton forms
an O-H bond nearby. However, the proton is loosely bound, and the volatility
observed in electrical measurements of SiO_2-based CMOS devices can be explained
by the motion of the proton from one O ion to another (Wimmer et al. 2016).
In amorphous HfO_2, hydrogen is also negative-U, with charged states being the
most thermodynamically stable at all Fermi-level positions. However, metastable
atomic hydrogen can share an electron with intrinsic electron trapping precursor
sites forming a $[Hf^{3+}\!\!-OH]$ type center (Kaviani et al. 2017), which is lower
in energy on average by about 0.2 eV. These electron trapping sites can affect both
the dynamics and thermodynamics of the interaction of hydrogen with a-HfO_2 and
the electrical behavior of amorphous hafnia films in CMOS devices. Thus structural
disorder adds important new features to the interaction of hydrogen with materials.

6   Defects in Amorphous Oxides

Amorphous oxides are produced using different methods: metal oxidation, film
growth on semiconductor, and metal substrates using different techniques, such
as atomic layer deposition (ALD), pulse laser deposition (PLD), chemical vapor
deposition (CVD), molecular-beam epitaxy (MBE), rf-sputtering, sol-gel dip coat-
ing, chemical synthesis of powders, and other methods. Densities and atomic
structures of produced samples strongly depend on the method and temperature
of growth, anneal, and other treatments. In this chapter we mainly consider oxide
films grown on metal and semiconductor substrates, which are better studied due
to their wide applications. Still, different communities use different methods of
preparation and propose different ideas regarding the structure of these films. For
example, the thermodynamic approach suggests that an amorphous structure of
oxide overgrowths on metals may be stabilized with respect to the corresponding
crystalline oxide, up to a certain critical oxide-film thickness. This results from
the absence of mismatch strain for the amorphous oxide film (in contrast to an
epitaxial or semi-coherent crystalline oxide film). The value of critical oxide-film

thickness when initial amorphous oxide film transforms into a crystalline oxide film depends on the substrate orientation, temperature, and metal-oxygen system under study (Reichel et al. 2008). However, density, thickness, and morphology of oxide films strongly depend on the substrate and method of deposition, and amorphous films well exceeding a critical thickness are grown on many substrates. They have a varied degree of crystallization depending on the growth conditions, anneal temperature, as well as other factors, such as oxygen pressure. Low oxygen pressure leads to the formation of oxygen vacancies which may induce a decrease of the valence state of the metallic ions or a phase separation reaction in the oxide (Millon 2013). These results highlight the fact that, unlike bulk crystal phase, the morphology and atomistic structure of amorphous films are extremely varied dependent on the sample preparation, which makes simulation of these structures challenging (Medvedeva et al. 2017; Pandey et al. 2016).

By far the most popular method of creating atomistic models of bulk amorphous oxides is based on melting crystalline samples using periodic boundary conditions and then cooling the melt to low temperatures. This method is based on the assumption that all liquids can be frozen into amorphous state by fast cooling. Application of 3D periodic boundary conditions helps to eliminate nucleation of crystal phases unless crystalline nuclei are deliberately introduced into melt. Cooling is usually unrealistically fast due to computer limitations and the cooling rate is rarely slower than 1 K/ps. In many cases, especially when quantum mechanical methods are employed, the cooling rate can be as high as 500 K/ps. Since many oxides are so-called intermediate glass formers, they tend to crystallize at low cooling rates. Further optimization of the volume and geometry of these structures and calculations of defect states are often performed using DFT and periodic models of a system.

The biggest complication for modeling defects in amorphous oxides is related to the variations in their local environment due to disorder. Defects in the bulk crystalline phase formed at equivalent lattice sites have equivalent properties. However, in amorphous structures, all sites are different, and any comparison of model parameters with experiment must involve statistical analysis of different configurations. Using periodic cells is required in order to avoid border effects which may affect defect characteristics. However, this makes the amorphous structures quasiperiodic and induces constraints on the structural relaxation accompanying defect creation. Taken together these factors imply that simulations should be performed in the largest periodic cells feasible for DFT calculations. To credibly predict distributions of properties, calculations at many sites in amorphous structure and of many models are required.

Some of the effects of disorder have been mentioned above when discussing the spread in the formation energies of oxygen vacancies at different sites within the $Si/SiO_2/HfO_2$ stack in Fig. 7. In the previous section, we mentioned that hydrogen can react differently with certain sites in amorphous oxides creating specific defects which do not exist in the crystalline phase. Perhaps the most straightforward manifestation of structural disorder is distribution of defect one-electron states and charge transition levels illustrated in Fig. 2. The figure shows

two discrete black levels in the silica bandgap, which correspond to the positions of occupied states of neutral and positively charged oxygen vacancy in α-quartz. To the right from these levels are schematically shown distributions of levels of neutral and positively charged vacancies in a-SiO$_2$. To calculate such a distribution, one would usually position vacancies at different sites in a periodic cell simulating an amorphous structure. Discrete levels correspond to the finite number of vacancies in a periodic cell and continuous distributions to a macroscopic sample. Different levels correspond to vacancies formed at different sites in an amorphous structure. The distribution of levels leads to inhomogeneous broadening of optical spectra and gives rise to a large variety of possible combinations of relaxation energies and energy barriers when considering defect diffusion and electron capture and emission processes (Wimmer et al. 2016).

7   Outlook

We reviewed some of the main issues pertaining to calculations of point defects and defect processes in oxides for microelectronics. To some extent we followed empirical approaches, which rely on knowledge of other materials in similar states to those to be used in devices. Therefore we have drawn some analogies with other oxides, such as alumina and MgO. The properties we considered concern mainly non-stoichiometry and disorder of oxides as well as related issues, such as the role of hydrogen.

Theoretical modeling is a source of reliable models of defects and defect processes. They are impossible to develop fully from the experimental data alone. A comparison of models of similar species in different materials points to interesting trends, which are not obvious from disparate data available in different research areas. Particularly striking are similarities in models of structure and diffusion of interstitial atomic oxygen species in different oxides. They form peroxy linkage-like species symmetric with respect to the original oxygen site and usually diffuse by the exchange (interstitialcy)-type mechanism (Foster et al. 2002a, b). On the other hand, oxygen vacancies, which are intrinsic structural defects, form a range of models depending on the ionicity of the material and coordination of oxygen sites. In ionic materials with high coordination, such as MgO, Al$_2$O$_3$, and HfO$_2$, they form F-center-like defects and deep states in the bandgap. In more covalent materials, such as SnO$_2$ and CeO$_2$, electrons left in the vacancy tend to be more localized on surrounding metal atoms. In SiO$_2$ where oxygen sites are only two-coordinated, neutral vacancies can be viewed as a Me-Me chemical bond.

The direct relevance of this to technological applications is perhaps not immediately evident. Certainly defect models discussed above have been developed as a result of analysis of large amount of experimental data. In return, they helped to rationalize that data and predict models are transferable between different materials. Still specific problems pertaining to microelectronics are different: defects initiate degradation; they cause charge trapping and provide fixed charge; dielectric performance will vary in time, as the defects or charge state populations evolve;

thin film materials will often be inhomogeneous because of the different epitaxial constraints at the two interfaces; finally, there will be effects on interface properties, including interface dipoles. The mechanisms of these processes remain largely unresolved although the number of works trying to bridge the gap between defect modeling and real processes is growing fast. Establishing a direct link between defect models and properties and their role in processes leading to degradation of microelectronic devices remains one of the main challenges of both theoretical modeling and experiment.

Among other challenges one can note computational studies of the effect of structural disorder on the structure and properties of defect in oxides, which have only started to appear, with the main emphasis on amorphous silica. They demonstrate the limited transferability of defect models developed in crystals within disordered amorphous materials. One can also expect limited transferability of defect models onto thin films and interfaces. Direct modeling of defects in these systems will become extremely important with the ever-decreasing thickness of oxide films and increasing computer resources. Linking the static properties of these defects with real processes in microelectronic devices requires more effort on the front of modeling dynamics of defect reactions. In particular, calculations of the cross sections of defect states with respect to charge transfer processes, studies of effectiveness and rates of defect reactions, and the effect of energy dissipation are bound to become extremely important. This should begin to bridge the size and timescale gap existing in most of the calculations discussed above.

References

Afanas'ev VV, Stesmans A (1998) Hydrogen-induced valence alternation state at SiO_2 interfaces. Phys Rev Lett 80:5176–5179

Alay JL, Hirose M (1997) The valence band alignment at ultrathin SiO_2/Si interfaces. J Appl Phys 81:1606

Bachlechner ME, Omeltchenko A, Nakano A, Kalia RK, Vashishta P (1998) Multimillion-atom molecular dynamics simulation of atomic level stresses in $Si(111)/Si_3N_4(0001)$ nanopixels. Appl Phys Lett 72:1969–1971

Blöchl PE (2000) First-principles calculations of defects in oxygen-deficient silica exposed to hydrogen. Phys Rev B 62:6158–6179

Chen Y, Kolopus JL, Sibley WA (1969) Luminescence of the F^+ center in MgO. Phys Rev 186:865

Cox SFJ (2003) The shallow-to-deep instability of hydrogen and muonium in II–VI and III–V semiconductors. J Phys Condens Matter 15(46):R1727

Cox SFJ, Gavartin JL, Lord JS, Cottrell SP, Gil JM, Alberto HV, Piroto Duarte J, Vilao RC, Ayres De Campos N, Keeble DJ et al (2006) Oxide muonics: II. Modelling the electrical activity of hydrogen in wide-gap and high-permittivity dielectrics. J Phys Condens Matter 18(3):1079

Degreave R, Kaczer B, Groeseneken G (1999) Degradation and breakdown in thin oxide layers: mechanisms, models and reliability prediction. Microelectron Reliab 39:1445–1460

Dkhissi A, Este've A, Mastail C, Mazaleyrat G, Olivier S, Jeloaica L, Djafari Rouhani M (2008) Multiscale modeling of the atomic layer deposition of HfO_2 thin film grown on silicon: how to deal with a kinetic Monte Carlo procedure. J Chem Theory Comput 4:1915–1927

Dkhissi A, Mazaleyrat G, Este've A, Djafari Rouhani M (2009) Nucleation and growth of atomic layer deposition of HfO_2 gate dielectric layers on silicon oxide: a multiscale modelling investigation. Phys Chem Chem Phys 11:3701–3709

Edwards AH, Fowler WB (1999) Recent advances in the theory of oxide-semiconductor interfaces. Microel Reliab 39:3–14

El-Sayed Al-M, Watkins MB, Grasser T, Afanas'ev VV, Shluger AL (2015) Hydrogen-induced rupture of strained Si-O bonds in amorphous silicon dioxide. Phys Rev Lett 114(11):115503

Fleetwood DM (2002) Effects of hydrogen transport and reactions on miscroelectronics radiation response and reliability. Microelectron Reliab 42:523–541

Fleetwood DM, Pantelides ST, Schrimpf RD (eds) (2008) Defects in microelectronic materials and devices. CRC Press, Boca Raton

Forst CJ, Ashman CR, Schwarz K, Blochl PE (2004) The interface between silicon and a high-k oxide. Nature 427:53–56

Foster AS, Lopez Gejo F, Shluger AL, Nieminen RM (2002a) Vacancy and interstitial defects in hafnia. Phys Rev B 65:174117

Foster AS, Shluger AL, Nieminen RM (2002b) Mechanism of interstitial oxygen diffusion in hafnia. Phys Rev Lett 89:225901–1–225901–4

Freysoldt C, Grabowski B, Hickel T, Janotti A, Neugebauer J, Kresse G, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86:254–305

Gavartin JL, Shluger AL (2007) Modeling $HfO_2/SiO_2/Si$ interface. Microel Eng 84:2412–2415

Gellings PJ, Bouwmeester HJM (2000) Solid state aspects of oxidation catalysis. Catal Today 58:1–53

Godet J, Broqvist P, Pasquarello A (2007) Hydrogen in Si (100)-SiO_2-HfO_2 gate stacks: relevant charge states and their location. Appl Phys Lett 91(26):262901

Grasser T (ed) (2014) Bias temperature instability for devices and circuits, Springer, New York

Griscom DL (2000) The natures of point defects in amorphous silicon dioxide. In: Pacchioni G, Skuja L, Griscom DL (eds). Defects in SiO_2 and related dielectrics: science and technology. NATO science series, pp 117–161. Kluwer, Dordrecht

Griscom DL, Friebele EJ, Jr Sigel GH (1974) Observation and analysis of the primary ^{29}Si hyperfine structure of the E' center in non-crystalline SiO_2. Sol State Sommun 15:479

Hayes W, Stoneham AM (1985) Defects and defect processes in non-metallic solids. Wiley, New York

Henderson B (1980) Anion vacancy centers in alkaline earth oxides. CRC Critical Rev Sol St Mater Sci 9:1–60

Henderson B, Wertz JE (1977) Defects in the alkaline earth oxides with applications to radiation damage and catalysis, Taylor and Francis, London

Hine NDM, Frensch K, Foulkes WMC, Finnis MW (2009) Supercell size scaling of density functional theory formation energies of charged defects. Phys Rev B 79:024112

Hofmann A, Ganduglia-Pirovano MV, Sauer J (2007) Oxygen vacancies in transition metal and rare earth oxides: current state of understanding and remaining challenges. Surf Sci Rep 62:219–270

Islam MS (2000) Ionic transport in ABO_3 perovskite oxides: a computer modelling tour. J Mater Chem 10:1027–1038

Jupille J, Thornton G (eds) (2015) Defects at oxide surfaces. Springer, Cham

Kaviani M, Afanasev VV, Shluger AL (2017) Interactions of hydrogen with amorphous hafnium oxide. Phys Rev B 95(7):075117

Kawamoto A, Jameson J, Cho K, Dutton RW (2000) Challenges for atomic scale modelling in alternative gate stack engineering. IEEE Trans Electr Dev 47(10):1787–1794

Kawamoto A, Cho K, Dutton R (2002) Perspectives paper: first principles modelling of high-k dielectrics. J Comp Aided Mater Design 8:39–57

Kotomin EA, Popov AI (1998) Radiation-induced point defects in simple oxides. Nucl Instr Meth Phys B 141:1

Kotomin EE, Maier J, Eglitis RI, Borstel G (2002) Calculation of radiation-induced point defect, polarons and excitons in ferroelectric perovskites. Nuclear Instr Meth Phys Res B 191:22

Lenahan PM, Jr Conley JF (1998) What can electronic paramagnetic resonance tell us about the Si/SiO$_2$ system. J Vac Sci Technol 16(4):2134–2153

Li H, Robertson J (2014)   Behaviour of hydrogen in wide band gap oxides.   J Appl Phys 115(20):203708

Lyons JL, Janotti A, Van de Walle CG (2011) The role of oxygen-related defects and hydrogen impurities in HfO$_2$ and ZrO$_2$. Microelectron Eng 88(7):1452–1456

Massoud HZ, Baumvol IJR, Hirose M, Poindexter EH (eds) (2000) The physics and chemistry of SiO$_2$ and the Si-SiO$_2$ interface, vol 4. The Electrochemical Society, Pennington, NJ

Medvedeva JE, Buchholz DB, Chang RPH (2017) Recent advances in understanding the structure and properties of amorphous oxide semiconductors. Adv Electron Mater 3:1700082–1–17

Mihaychuk JG, Shamir N, van Driel HM (1999) Multiphonon photoemission and electric-field-induced optical second- harmonic generation as probes of charge transfer across the Si/SiO$_2$ interface. Phys Rev B 59:2164

Millon E (2013) Advanced functional oxide thin films grown by pulsed-laser deposition. Appl Surf Sci 278(7):2–6

Muñoz Ramo D, Gavartin JL, Shluger AL, Bersuker G (2007) Spectroscopic properties of oxygen vacancies in monoclinic HfO$_2$ calculated with periodic and embedded cluster density functional theory. Phys Rev B 75:205336

Netzer FP, Fortunelli A (eds) (2016) Oxide materials at the two-dimensional limit. Springer, Cham

Neumark GF (1997) Defects in wide band gap II-VI crystals. Mater Sci Eng Reports 21:1–46

Pacchioni G (2000) Ab initio theory of point defects in oxide materials: structure, properties, chemical reactivity. Solid State Sci 2:161–179

Pacchioni G (2000) Theory of point defects at the MgO surface, In: Woodruff DP (ed) The chemical physics of solid surfaces vol 9 - oxide surfaces, pp 94–135. Elsevier, Amsterdam

Pacchioni G, Skuja L, Griscom DL (eds) (2000) Defects in SiO$_2$ and related dielectrics: science and technology. NATO science series. Kluwer, Dordrecht

Pandey A, Biswas P, Drabold DA (2016) Inversion of diffraction data for amorphous materials. Sci Rep 6:33731

Popov AI, Kotomin EA, Maier J (2010) Basic properties of the F-type centers in halides, oxides and perovskites. Nucl Instr Meth Phys Res B 268:2084–3089

Reichel F, Jeurgens LPH, Mittemeijer EJ (2008) The thermodynamic stability of amorphous oxide overgrowths on metals. Acta Mater 56(7):659–674

Ricci D, Di Valentin C, Pacchioni G, Sushko PV, Shluger AL, Giamello E (2003) Paramagnetic defect centers at the MgO surface. An alternative model to oxygen vacancies. J Am Chem Soc 125(3):738–747

Stathis JH (2002) Reliability limits for the gate insulator in CMOS technology. IBM J Res Dev 46:265–286

Sulimov VB, Sushko PV, Edwards AH, Shluger AL, Stoneham AM (2002) Asymmetry and long-range character of lattice deformation by neutral oxygen vacancy in α-quartz. Phys Rev B 66:024108

Tse KY, Liu D, Robertson J (2010) Electronic and atomic structure of metal-HfO$_2$ interfaces. Phys Rev B 81:035325

van Duin ACT, Strachan A, Stewman S, Zhang Q, Xu X, III Goddard WA (2003) Reaxff$_{SiO}$ reactive force field for silicon and silicon oxide systems. J Phys Chem A 107:3803–3811

Weber JR, Janotti A, Van de Walle CG (2011) Native defects in Al$_2$O$_3$ and their impact on III-V/Al$_2$O$_3$ metal-oxide semiconductor-based devices. J Appl Phys 109:033715

Weeks RA (1956) Paramagnetic resonance of lattice defects in irradiated quartz. J Appl Phys 27:1376–1381

Weeks RA (1994) The many varieties of E' centers: a review. J Non-Cryst Solids 179:1–9

Wimmer Y, El-Sayed Al-M, Gös W, Grasser T, Shluger AL (2016) Role of hydrogen in volatile behaviour of defects in SiO$_2$-based electronic devices. In: Proceedings of the Royal Society A 472:20160009. The Royal Society

Zeng Z, Hangaard Hansen M, Greeley JP, Rossmeisl J, Bjrketun ME (2014) Ab initio thermo-dynamic modeling of electrified metal-oxide interfaces: consistent treatment of electronic and ionic chemical potentials. J Phys Chem C 118:22663–22671

Zunger A (1998) Theoretical predictions of electronic materials and their properties. Current Opinion Sol St Mater Sci 3:32–37

Small Polarons in Transition Metal Oxides

44

Michele Reticcioli, Ulrike Diebold, Georg Kresse,
and Cesare Franchini

Contents

Abstract

The formation of polarons is a pervasive phenomenon in transition metal oxide compounds, with a strong impact on the physical properties and functionalities of the hosting materials. In its original formulation, the polaron problem considers a single charge carrier in a polar crystal interacting with its surrounding lattice.

M. Reticcioli · G. Kresse
University of Vienna, Vienna, Austria
e-mail: michele.reticcioli@univie.ac.at; georg.kresse@univie.ac.at

U. Diebold
Institute of Applied Physics, Technische Universität Wien, Vienna, Austria
e-mail: diebold@iap.tuwien.ac.at

C. Franchini (✉)
Faculty of Physics and Center for Computational Materials Science, University of Vienna, Vienna, Austria
e-mail: cesare.franchini@univie.ac.at

Depending on the spatial extension of the polaron quasiparticle, originating from the coupling between the excess charge and the phonon field, one speaks of small or large polarons. This chapter discusses the modeling of small polarons in real materials, with a particular focus on the archetypal polaron material TiO_2. After an introductory part, surveying the fundamental theoretical and experimental aspects of the physics of polarons, the chapter examines how to model small polarons using first-principles schemes in order to predict, understand, and interpret a variety of polaron properties in bulk phases and surfaces. Following the spirit of this handbook, different types of computational procedures and prescriptions are presented with specific instructions on the setup required to model polaron effects.

1   Introduction

Electrons in perfect crystals are adequately described in terms of periodic wave functions. However, rather than being constituted by a perfect lattice, real materials are often characterized by the presence of defects, such as point defects (vacancies or interstitials), impurities, grain boundaries, and dislocations (Ashcroft and Mermin 1976; Crawford and Slifkin 1972). Defects in the crystal break the periodicity of the electronic charge density. They can lead to the formation of localized states (Freysoldt et al. 2014; Pacchioni 2015), which affect the materials properties and, thus, their performance in applications. Less intuitively, such localized states could exist also in the absence of any defects in the crystal, i.e., they can form in perfect lattices. This phenomenon is referred to as polarons. A polaron is a quasiparticle originating from the interactions between charge carriers (i.e., electrons or holes) and lattice ion vibrations (Holstein 1959). More precisely, due to Coulomb forces, the excess charge displaces the ions in its neighborhood creating a polarization cloud that follows the charge carrier as it moves through the crystal (see Fig. 1). Such quasiparticles differ from lattice defect states own peculiar properties and are described by well-developed quantum field theories based on effective Hamiltonians (Fröhlich 1954; Holstein 1959).

Formation of polarons is particularly favorable in polar semiconductors and transition metal oxides owing to the strength of the electron-phonon interaction and is further promoted in the vicinity of a surface, where the crystal lattice is more flexible and the necessary lattice relaxations cost less energy (Deskins et al. 2009, 2011; Kowalski et al. 2010; Setvin et al. 2014a, b). Polarons play a decisive role in a wide range of phenomena, including electron transport (Crevecoeur and De Wit 1970; Moser et al. 2013), optical absorption, and chemical reactivity, and have crucial implications in high-T_c superconductivity (Salje et al. 2005), colossal magnetoresistance (Teresa et al. 1997; Rønnow et al. 2006), thermoelectricity (Wang et al. 2014a), photoemission (Cortecchia et al. 2017), and photochemistry (Linsebigler et al. 1995).

This chapter is mostly focused on the first-principles modelling of small polarons (Sect. 2) and its application to TiO_2 (Sect. 3), an archetypal and widely studied

Fig. 1 Pictorial view of a polaron. An excess charge is trapped in a lattice site (bright) and distorts the surrounding lattice
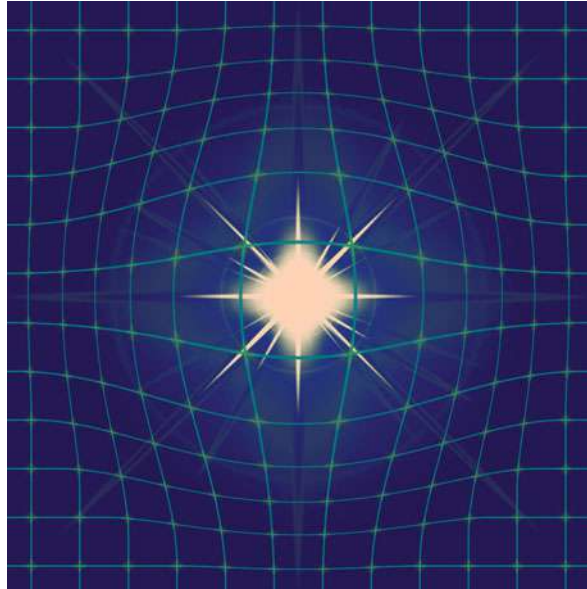
polaron material. This introduction is complemented with a brief summary of the fundamental theoretical aspects defining the physics of polarons and with an overview of the experimental techniques used to detect polaron features in materials.

1.1 Theoretical Background

The emergence of polaron theory can be traced back to 1933 when Landau elaborated on the possibility for lattice distortions to trap electrons by means of an intrinsic modification of the lattice phonon field induced by the electron itself (Landau 1933). The resulting electron-phonon quasiparticle was later called polaron, a coupled electron-phonon system in which the polarization generated by the lattice distortions acts back on the electron renormalizing the electronic properties, for instance, the effective mass. An analogous discussion is valid for holes rather than electrons. A first formal description of the polaron problem was published in 1946 by Pekar (1948), who considered a free electron interacting with lattice deformations in the continuum approximation. Therefore, Pekar's study is limited to the case of polarons with a size larger than the lattice constant, so that the atomic discreteness is negligible (Alexandrov and Devreese 2010). With the ionic lattice described as a polarizable dielectric continuum, Landau and Pekar showed that the polaron mass m^* is enhanced with respect to the "bare" electron mass (Landau and Pekar 1965), an important result that set the basis for all subsequent theories.

During the 1950s, the second quantization formalism was used to refine the description of the polaron problem in terms of quantum effective Hamiltonians including an electronic term (H_e), a phonon term (H_{ph}), and the electron-phonon coupling term (H_{e-ph}). The $e - ph$ term is of fundamental importance for understanding the polaronic states, since it takes into account the type (e.g., short- or long-range) and intensity (weak to strong) of the mutual interactions between charge carriers and lattice vibrations. However, a complete analytic solution of the polaron Hamiltonian cannot be achieved, and approximations are needed. Fröhlich et al. (1950) and Holstein (1959) separately addressed different aspects of the problem and proposed mathematically more tractable formulations. Fröhlich theory relies on the continuum approximation and assumes long-range forces (large polarons), whereas Holstein theory takes into account explicitly the lattice periodicity and treats the electron-phonon coupling as a short-range interaction (small polaron). The distinction between small and large polaron is defined according to the strength of the electron-phonon coupling (weak/strong) and the extension of the lattice distortion around the electron (large/small) (Alexandrov and Devreese 2010; Rashba 2005; Hahn et al. 2018). The distinct features of small and large polarons are discussed at the end of this section.

1.1.1 Fröhlich Hamiltonian

Fröhlich theory treats the situation of long-range coupling between electrons and phonons, which leads to the formation of large polarons (Fröhlich et al. 1950; Fröhlich 1954; Alexandrov and Devreese 2010) and is formalized by the Hamiltonian:

$$H = \sum_{\mathbf{k}} \frac{k^2}{2} c_{\mathbf{k}}^\dagger c_{\mathbf{k}} + \sum_{\mathbf{q}} b_{\mathbf{q}}^\dagger b_{\mathbf{q}} + \sum_{\mathbf{k},\mathbf{q}} \left[V_d(\mathbf{q}) b_{\mathbf{q}}^\dagger c_{\mathbf{k}-\mathbf{q}}^\dagger c_{\mathbf{k}} + V_d^\dagger(\mathbf{q}) b_{\mathbf{q}} c_{\mathbf{k}+\mathbf{q}}^\dagger c_{\mathbf{q}} \right] \quad (1)$$

where $c_{\mathbf{k}}$ and $b_{\mathbf{q}}$ are annihilation operators for a particle with wave vector $\mathbf{k}$ and a phonon with wave vector $\mathbf{q}$, respectively, and $V_d(\mathbf{q})$ is the electron-phonon coupling function for a system in d dimensions, in three dimensions:

$$V_3(\mathbf{q}) = i \left(\frac{2\sqrt{2}\pi\alpha}{V} \right)^{\frac{1}{2}} \frac{1}{q}. \quad (2)$$

Here V is the volume of the system, and α is the dimensionless Fröhlich electron-phonon coupling constant defined in terms of the reduced Planck constant $\hbar$, the charge carrier charge e, the phonon frequency ω of the long-wavelength optical longitudinal phonon, and the material's static and high-frequency dielectric constants ε_0 and ε_∞, respectively (ε_0 includes ionic relaxation effects, whereas the ion-clamped ε_∞ accounts only for electronic contributions):

$$\alpha = \frac{e^2}{\hbar} \sqrt{\frac{m}{2\hbar\omega}} \left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0} \right). \tag{3}$$

In real materials, α ranges between ≈ 0.02 (InSb) and ≈ 3.8 (SrTiO$_3$). In the weak-coupling regime (small α), the Fröhlich Hamiltonian can be solved using perturbation theory, and the Feynman's path integral approach provides accurate results for all coupling strengths (Feynman 1955; Rosenfelder and Schreiber 2001). In the large polaron Fröhlich picture, the spatial extension of the large polaron is bigger than the lattice constant, and the dressed electron is accompanied by a phonon cloud, whose density determines the size of the polaron effective mass m^* and the polaron energy (Devreese 1996; Emin 2013). Recently, the Fröhlich problem has been recast within a first-principle perspective by F. Giustino, which allows for predictive nonempirical calculations of material's specific electron-phonon properties (Giustino 2017).

1.1.2   Holstein Hamiltonian

Holstein considered the polaron determined by a short-range strong-coupling regime, the so-called small polaron, spatially confined within a radius of the order of the lattice constant, described by the general Holstein Hamiltonian (Holstein 1959; Ramakumar and Das 2004):

$$H = -t \sum_{<ij>} c_i^\dagger c_j + \omega \sum_i b_i^\dagger b_i + g \sum_i n_i (b_i^\dagger + b_i) \tag{4}$$

where t is the hopping amplitude between neighboring sites, $c_i(b_i)$ and $c_i^\dagger(b_i^\dagger)$ are fermionic (bosonic) creation and annihilation operators acting on the site i, ω the phonon frequency, g the electron-phonon coupling strength, and $n_i = c_i^\dagger c_i$ is the fermionic particle number operator. As opposed to the Fröhlich model, the Holstein model only considers electron-phonon interactions at a single lattice site, greatly simplifying the model. In the simplified case of a polaron trapped in a linear molecule, hopping between different molecular sites (Holstein 1959, 2000), the Hamiltonian in Eq. 4 reduces to a two-site Hamiltonian describing the electron hopping between two sites, interacting with an ion placed in between via its vibrational mode (Alexandrov and Devreese 2010):

$$H = t \left(c_1^\dagger c_2 + c_2^\dagger c_1 \right) - \frac{1}{2M} \frac{\partial^2}{\partial x^2} + \frac{\gamma x^2}{2} + g x \left(c_1^\dagger c_1 + c_2^\dagger c_2 \right) \tag{5}$$

with M the ion mass, $\gamma = M\omega^2$ the spring constant and x the linear ion displacement from the equilibrium position. The state $|\psi\rangle$ with an extra electron introduced into the unperturbed system $|0\rangle$ and localized at the site 1 or 2 with amplitude probability $u(x)$ and $v(x)$, respectively, is written as

$$|\psi\rangle = \left[u(x) c_1^\dagger + v(x) c_2^\dagger \right] |0\rangle . \tag{6}$$

Despite the simplicity of this model, an analytic solution for the Schrödinger equation $H\psi = E\psi$ can be found only in special cases (e.g., ionic vibration perpendicular to the hopping direction), and one has to rely on numerical approximations or restrict the study to the nonadiabatic ($t \ll \omega$) or adiabatic ($t \gg \omega$) limits. The modelling of small polaron within a first-principles picture is addressed in Sect. 2.

At this point, the distinction between polarons and electrons trapped at lattice defect sites stands out clearly by inspecting the Fröhlich's and Holstein's analysis. First of all, as already mentioned above, polaron formation occurs even in absence of defects, i.e., in the perfect crystal. Moreover, polarons are not stuck to a specific lattice site; rather they can move around in the system. Both these characteristics, i.e., spontaneous localization and mobility, were pointed out by Holstein, who summarized the concepts in few brilliant sentences (Holstein 1959):

> Let us imagine that an electron is momentarily fixed at some point of the crystal. As a result of electron-lattice interaction, the surrounding lattice particles will be displaced to new equilibrium positions [...] such as to provide a potential well for the electron. If the well is sufficiently deep, the electron will occupy a bound state, unable to move unless accompanied by the well, that is to say, by the induced lattice deformation.

At variance, the physics of an electron attached to a defect is rather different, as it is typically not mobile and its characteristics strongly depend on the type of defect (Freytag et al. 2016).

1.1.3   A Few Additional Considerations on Small and Large Polarons

Many other theoretical studies have further developed the ideas of Fröhlich and Holstein, leading to two alternative (and to some extent more advanced) descriptions of large and small polarons (Rashba 2005). Table 1 summarizes the main distinctions between the large and the small polaron (Emin 2013).

As already mentioned, their names reflect the length scale of the spatial localization: while the small polaron electronic charge is usually confined in the primitive cell (a few Å), the large polaron extends over several lattice sites (typically about 20 Å). Also, the two types of polaronic states exhibit distinct energy scales. The large polaron is usually a shallow state, few tens of meV below the conduction band minimum (CBM). Conversely, the small polaron is firmly trapped by local distortions in a stronger potential well, which determines the formation of a deeper in-gap state well localized around 1 eV below the CBM (see Fig. 2).

Table 1 General properties of large and small polarons

Large polaron	Small polaron
Polaron radius $\gg$ lattice parameter	Polaron radius $\simeq$ lattice parameter
Shallow state ($\sim$10 meV below CBM)	In-gap state ($\sim$1 eV below CBM)
Coherent motion (scattered occasionally by phonons)	Incoherent motion (assisted by phonons)
Mobility $\mu \gg 1\,\mathrm{cm^2/Vs}$	Mobility $\mu \ll 1\,\mathrm{cm^2/Vs}$
Decreasing mobility with increasing temperature	Increasing mobility with increasing temperature
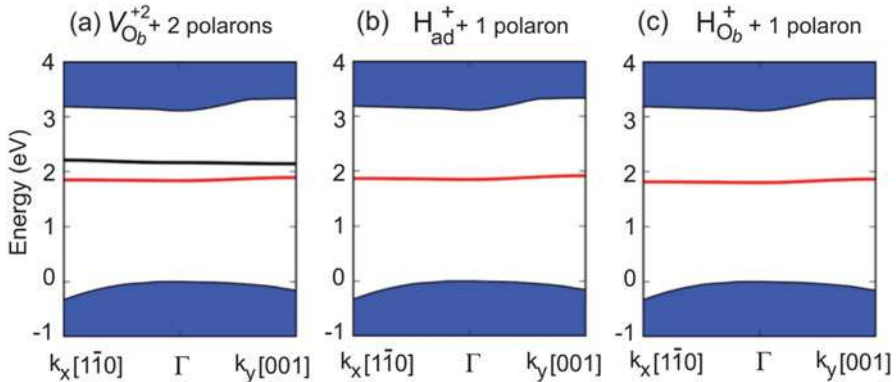
Fig. 2   Polaron in-gap level. Small polaron band-gap states in the surface of rutile TiO$_2$ formed by excess electrons donated by so-called bridge-bonded oxygen vacancies (V_{O_b}), hydrogen adsorbates (H$_{ad}$), or hydrogen hydroxyls (H$_{O_b}$). Results based on first-principles calculations within the hybrid-DFT approach. (This figure is taken from Moses et al. 2016)

The localization process is also different: a charge carrier introduced into a system can quickly equilibrate with the lattice and form a large polaron, whereas the formation of a small polaron occurs only after overcoming an energy barrier (Mott and Stoneham 1977; Lany 2015). Finally, the two types of polarons are characterized by very different transport properties. The large polaron, a heavy quasiparticle, is weakly scattered by phonons. This weak scattering can compensate the large mass, resulting in a large mobility for the large polaron. A small polaron instead hops between trapping sites with a lower mobility. Since its motion is assisted by phonons, the mobility of small polarons increases with increasing temperature. Both large and small polarons have been observed in several experiments, as reported in the next section, and studied by simulations and computational techniques. Historically, large polarons have been investigated mostly via effective Hamiltonians, in particular by means of variational treatments solved by Feynman's path integral techniques and by diagrammatic Monte Carlo (Feynman 1955; Prokof'ev and Svistunov 1998; Mishchenko et al. 2000; Rosenfelder and Schreiber 2001; Gerlach et al. 2003; Gerlach and Smondyrev 2008; Hahn et al. 2018) approaches. First-principles techniques are more suitable for the description of the small polaron, but successful attempts to address the large polaron case have been presented in the last few years (Setvin et al. 2014a; Verdi et al. 2017; Miyata et al. 2017).

1.1.4   How First-Principles Modeling Can be Used

The Fröhlich and Holstein model Hamiltonians are simplifications of the more general electron-phonon Hamiltonian that is discussed in great detail in a recent review of Giustino (2017). Giustino also explains how the individual terms can be derived by first-principles density functional theory, as briefly recapitulated below: (i) The first term describes the one-electron band structure, and that can obviously be calculated and parameterized using density functional theory, hybrid functionals

or the GW approximation. The original Fröhlich and Holstein Hamiltonian involve only a single band, with a parabolic dispersion in the Fröhlich case, and a tight binding form in the Holstein case. In the case of electron (hole) polarons, one thus needs to parameterize the conduction (valence) band. (ii) The second term describes the vibrational frequencies of the optically active mode. The frequencies as well as the eigenmodes can be calculated by density functional perturbation theory. The Fröhlich Hamiltonian involves only a single dispersion free phonon mode, which couples with a wave vector dependent coupling term q. The functional form used by Fröhlich is exact in the long-wavelength limit. The Holstein model involves a single local mode, where the local displacement couples (somehow) to the electronic bands. (iii) The coupling between the electronic and phononic subsystem are described by the last term. The matrix elements in Fröhlich model can be exactly calculated using density functional perturbation theory (Verdi et al. 2017). The Holstein model is somewhat too simplistic and cannot obviously be mapped onto first-principles calculations.

The main issue with both approaches is twofold: (i) In many cases, the restriction to a single band and phonon mode is a too severe simplification, (ii) and even for a single band and a single phonon mode, the solution of the full many body problem can be formidable. This is the main reason why realistic modeling of polarons in real materials has hereto hardly been attempted. Nevertheless, the single band and single phonon model might be very accurate for many simple polar semiconductors and insulators. In this case, all one needs to determine are the effective masses, the phonon frequencies, and the dielectric constants. Then the solution of the Fröhlich model is fairly straightforward and expected to be accurate. Furthermore, there is little reason to invoke first-principles calculations for large polarons, if all materials parameters are experimentally established.

The situation is more complicated for small polarons. As already mentioned, the Holstein model is a too simplistic approach: for small polarons, the lattice distortions are usually large around an atom, and it is not obvious how to map the distortions to a single effective mode; furthermore the implicitly assumed harmonic approximation might not be accurate if the distortions are large. However, small polarons often involve fairly large energy scales around 1 eV. In this case, one might consider to treat the ionic degrees of freedom as classical particles. In many transition metal oxides, this should be a valid approximation, since the typical vibrational frequencies are small compared to the considered energy scales. In other words, quantum fluctuations of the ions can be disregarded. The second crucial approximation is to treat the electronic degrees of freedom using a mean field theory, such as DFT. There is no need to discuss the accuracy of DFT; it often works exceedingly well. Polarons though pose a particular challenge, since they involve localization of the electrons, which many density functionals fail to describe accurately because of self-interaction errors (see Sect. 2). Compared to the model approach, the full first-principles approach is much simpler. One simply choses a large unit cell, adds an electron, and lets the system evolve into the ground state. The disadvantage compared to the model approach is about giving up on the quantum nature of the ions and restricting the treatment of the electrons to a mean field

approximation. The advantages are, however, numerous. The model Hamiltonians are obtained by expanding the full electron-phonon Hamiltonian around a reference ground state to second order in the ionic displacements (recall, one can use density functional perturbation theory to calculate the vibrational frequencies, as well as the electron-phonon coupling). Anharmonicities are therefore implicitly neglected in the model approach. In this respect the simplistic, brute force approach certainly excels the model approach.

1.2   Experimental Observations of Polarons

Several years after the first theoretical predictions, experimental observations based on different techniques started to detect the presence of polarons in real materials and to study their properties. Nowadays, it is a common practice to complement experimental data with theoretical interpretations based on electronic structure simulations. Table 2 lists significant experimental studies revealing polaron formation in oxide materials. This list also represents the variety of experimental techniques used to observe both small and large polarons, formed by both negative (electron) and positive (hole) excess charges in oxide materials.

1.2.1   Conductivity Measurements

The first observation of polarons is attributed (Stoneham et al. 2007) to an experimental study published in 1963 (Nagels et al. 1963). Experiments on oxidized uranium dioxide reported an increasing hole conductivity with raising temperature, following a behavior well described by the formula:

$$\mu \sim \frac{1}{T} \exp\left(-\frac{\Delta E}{k_B T}\right), \tag{7}$$

indicating small polaron hopping from U^{5+} to U^{4+} sites upon thermal activation, an interpretation further substantiated by the observation that the activation energy ΔE decreases with increasing oxygen concentration.

Experimental confirmations of Eq. 7 were obtained for a wide range of materials, with charge carriers injected by different types of defects. Examples are given by the hole small polarons observed in Li-doped MnO (Crevecoeur and De Wit 1970) and by the electron small polarons in oxygen-deficient cerium dioxide (Tuller and Nowick 1977). In the latter example, two excess electrons originate from every oxygen vacancy present in the system. Each excess electron localizes at one Ce^{4+} site forming a Ce^{3+} ion. Hopping is activated by temperature, and the conductivity measurements confirmed the expected trend for polarons (Eq. 7). Moreover, by measuring the Seebeck coefficient for thermo-electricity, the number of carriers was found to be temperature independent. This is in contrast with the band model for the conductivity, which predicts an increasing number of charge carriers with raising temperature. On the contrary, according to the polaron hopping model, the charge

Table 2 List of experimental observations of polarons in oxide materials using different techniques

Type	Material	Source	Exp. technique	Publication
Hole, small polaron	UO_2	Oxidation	Conductivity measurement	(Nagels et al. 1963)
Hole, small polaron	MnO	Li doping	Conductivity measurement	(Crevecoeur and De Wit 1970)
Electron, small polaron	CeO_2	O vacancies	Conductivity and Seebeck	(Tuller and Nowick 1977)
Electron, small polaron	$BaTiO_3$	Nb doping	EPR	(Possenriede et al. 1994)
Electron, small and large polaron	$a\text{-}TiO_2$ and $r\text{-}TiO_2$	Nb doping	Conductivity and optical measurements	(Zhang et al. 2007)
Electron, small polaron	$r\text{-}TiO_2$	O vacancies	Resonant photoelectron diffraction	(Krüger et al. 2008)
Electron, small polaron	$r\text{-}TiO_2$	O vacancies	EPR	(Yang et al. 2013)
Electron, large polaron	$a\text{-}TiO_2$	O vacancies	ARPES	(Moser et al. 2013)
Electron, small polaron	$r\text{-}TiO_2$	O vacancies	STM and STS	(Setvin et al. 2014a)
Electron, small polaron	$r\text{-}TiO_2$	UV irradiation or H adatom	IR spectroscopy	(Sezen et al. 2015)
Hole, small polaron	$LiNbO_3$	Visible light	IR spectroscopy	(Freytag et al. 2016)
Electron, small polaron	$r\text{-}TiO_2$	O vacancies	IR spectroscopy on adsorbates	(Cao et al. 2017)
Electron and hole, large polaron	Lead halide perovskites	Laser pulse	TR-OKE	(Miyata et al. 2017)

carriers are introduced in CeO_2 only by defects, and the mobility increases with raising temperature due to the increased phonon populations.

1.2.2 Electron Paramagnetic Resonance

Electron paramagnetic resonance (EPR) is another experimental technique able to identify lattice ions with unpaired electrons, a distinct feature of the polaron state. In EPR experiments, an external magnetic field splits the energy level of the unpaired electrons (Zeeman effect), thus determining energy levels available by emitting or absorbing a photon at a specific frequency: the sample is illuminated by light, typically at constant frequency, while the magnetic field varies, and the resonance peaks are measured as function of the magnetic field when the conditions for the level transitions are satisfied. The shape, intensity, and energy values of the

resonance peaks determine the electronic states of the atoms in the sample where charge trapping can be detected. Small polarons were identified by EPR in Nb-doped $BaTiO_3$ samples (Possenriede et al. 1994; Lenjer et al. 2002) and other materials including oxygen-deficient rutile TiO_2 samples (Yang et al. 2013).

1.2.3 Optical Measurements

The difference between small and large polarons manifests itself prominently in the optical properties of materials. An example is the response of two different polymorphs of TiO_2, rutile and anatase, to (electron) doping with Nb, reported by Zhang et al. (2007). Epitaxial thin films of the two different polymorphs were grown on appropriate substrates, $SrTiO_3(001)$ or $LaAlO_3(001)$ for anatase and $Al_2O_3(r$ and c-cut) for rutile. By substituting 4-valent Ti with 5% of 5-valent Nb, electrons were added to the system. The two types of films showed contrasting behavior in conductivity measurements: anatase films were metallic while rutile films semiconducting. Together with the similar carrier density observed in both films, this is consistent with the formation of large and small electron polarons, respectively. The optical transmittance of 80% in the visible range for epitaxial anatase films shows that these films can be characterized as transparent conductive oxides (Klein 2012).

1.2.4 Resonant Photoelectron Diffraction

The charge distribution in reduced TiO_2 rutile (110) single crystals was first determined by Krüger et al. (2008, 2012). Their experiment is based on angle-resolved x-ray photoemission (XPS). Intensity variations in a specific photoelectron peak (here Ti-$2p$) are recorded while changing the polar and azimuthal emission angle. Since forward focusing dominates the scattering of electrons with kinetic energies of a few hundred eV, the configuration of near-range atomic neighbors can be accessed in a rather direct manner and modeled in a simple cluster geometry (Chambers 1992). In their work, the authors took advantage of the fact that the Ti-$2p$ XPS peak shows a clear shoulder that is attributed to Ti^{3+}; the intensity of this feature was additionally increased by tuning the photon energy to a resonance condition (Bertel et al. 1983). The photoelectron diffraction pattern of the Ti^{4+} peak (from Ti^{4+} at regular lattice sites) and the reduced Ti^{3+} signature turned out to be drastically different, and the best fit was obtained by attributing the excess electrons to the subsurface Ti atoms. This is true independently of the way the excess electrons are introduced into the lattice, either by creating O vacancies (Krüger et al. 2008) or by adding electrons via Na deposition (Krüger et al. 2012). This points to the fact that the location of polarons in subsurface position is an intrinsic feature of rutile TiO_2.

1.2.5 Angle-Resolved Photoemission Spectroscopy

TiO_2 has been investigated by numerous different experimental techniques in addition to the methods mentioned above. Angle-resolved photoemission spectroscopy (ARPES) (Moser et al. 2013) was used to identify the presence of large polarons in the anatase polymorph of titanium dioxide. The energy dispersion close to the Fermi level measured by ARPES experiment shows satellite (shallow) states below

the conduction band, corresponding to large electron polarons, brought about by the oxygen vacancies present in the sample. By tuning the amount of oxygen vacancies (via UV irradiation Locatelli et al. 2007), the density of charge carriers can be controlled. At high vacancy concentrations, the satellite states disappear, due to the overlap of the polaronic wavefunctions, giving rise to a metallic behavior indicated by the crossing of the conduction band with the Fermi energy.

1.2.6   Scanning Tunneling Microscopy and Spectroscopy

A direct view at the polaronic states in TiO_2 can be achieved by using scanning tunneling microscopy (STM) and spectroscopy (STS) (Setvin et al. 2014a). In-gap states stand clearly from the STM and STS studies on both anatase and rutile TiO_2 polymorph, due to charge carriers (electrons) induced by oxygen vacancies or Nb doping. However, the two polymorphs present quite different polaronic states: while rutile TiO_2 is a prototypical small polaron material, anatase was found to host predominantly large polarons. STM is capable to provide information on the spatial distribution of the polaronic charge at the surface, which is helpful for the identification of small and large polarons.

1.2.7   Infrared Spectroscopy

Small polaron states have been detected in TiO_2 also by absorption infrared (IR) spectroscopy experiments (Sezen et al. 2015). The vibrational energies of the lattice bonds are typically in the infrared regime. Therefore, by exposure of the sample to IR light, resonance peaks appear in the measured transmitted and reflected beams. In the presence of small polarons, the localized charge together with the local lattice distortions contributes to form resonant peaks at characteristic vibrational frequencies. Interestingly, the polaron peaks in TiO_2 were found to be independent on the source of charge carriers, H adsorption and UV irradiation, a strong evidence of the polaron nature of the peaks. The characteristic vibrational spectra of a defect state, at a donor would, in fact, be susceptible to the type of the donor.

Infrared spectroscopy can also be used to detect polarons indirectly, by inspecting the vibrational properties of bonds in the vicinity of the trapping site. Electron small polarons were found to redshift the stretching frequency of NO molecules adsorbed on reduced TiO_2 surfaces with respect to the pristine samples (Cao et al. 2017). Analogously, hole small polarons modify the vibrational frequency of the OH impurities at Li vacancies in $LiNbO_3$ samples (Freytag et al. 2016). In the latter study, irradiation in the visible region was used to generate hole small polarons at O sites and electron polarons at Nb sites. The hole polarons localize in the vicinity of the Li vacancies, due to the a more favorable electrostatic potential, and therefore strongly contribute to the frequency shift of the OH impurities. Once the source is turned off, hole and electron polarons recombine, since their mobile character, and the original vibrational frequency is restored.

1.2.8   Time-Resolved Optical Kerr Effect

Recently, the time-resolved optical Kerr effect (TR-OKE) was used to investigate polaron formations in $CH_3NH_3PbBr_3$ and $CsPbBr_3$ perovskites (Miyata et al. 2017).

In a TR-OKE experiment, a laser pulse is sent to the material, and the polarization rotation is detected. With laser pulses of energy larger than the energy gap value, the experiment probes the TR-OKE response upon charge injection. In these lead halide perovskites, the detected signals are compatible with a description of the altered phonon dynamics in terms of the formation of hole and electron, large polarons in the $PbBr_3$ sublattice. This TR-OKE study reported also interesting insights on the polaron formation dynamics. In fact, the two materials show different rates for the polaron formation. This is attributed to the reorientations of the cations in $CH_3NH_3PbBr_3$, which determine a faster polaron formation than in $CsPbBr_3$.

2   Modeling Small Polarons by First Principles

The electronic structure community has been quite active in the study of polarons in real materials. Simulations can be helpful to understand and interpret the experimental findings or to predict the formation of polaronic states. This section provides a short overview of the first-principles methods generally used to model polarons in materials and discusses different computational procedures to acquire information on the formation and dynamics of polarons and on their mutual interaction.

2.1   Theories and Methods

Density functional theory (DFT) has been largely applied as a starting point to study polarons in materials. However, common schemes used to describe the exchange-correlation functional in DFT simulations, such as the local density and generalized gradient approximations (LDA and GGA, respectively), fail to describe properly the charge localization at atomic sites. By using LDA or GGA, the excess charges are delocalized overall in the lattice, and partial occupations of the available electronic levels are favored against integer occupations. Therefore, the modeling of polarons requires a correction in order to overcome the drawback of standard local and semilocal exchange-correlation approximations.

According to the Janak's theorem (Janak 1978), variations of the total energy E due to the electronic occupation (n_i) of the state i are given in terms of the DFT Hamiltonian eigenstates ε_i, as $dE/dn_i = \varepsilon_i$, independently of the exchange-correlation approximation. The expected behavior of the exact total energy is a piecewise linear function of the electronic occupation, with discontinuities in the first derivative for integer values ($n_i = N$) (Perdew et al. 1982; Perdew and Levy 1983), that is $d^2E/dn_i^2 = 0$ except at integer occupancies. Therefore, the energy of the state i remains constant during electron addition or removal.

At variance with the expected behavior, LDA and GGA generally result in a convex function for the total energy, i.e., $d^2E/dn_i^2 > 0$. This is schematically shown in Fig. 3. The energy change of the state i upon its own occupation reflects a spurious self-interaction effect, introduced by the type of exchange-correlation

Fig. 3 Total energy vs. occupation in DFT and HF. Total energy as a function of the electronic occupation in DFT (convex function) and HF schemes (concave function)
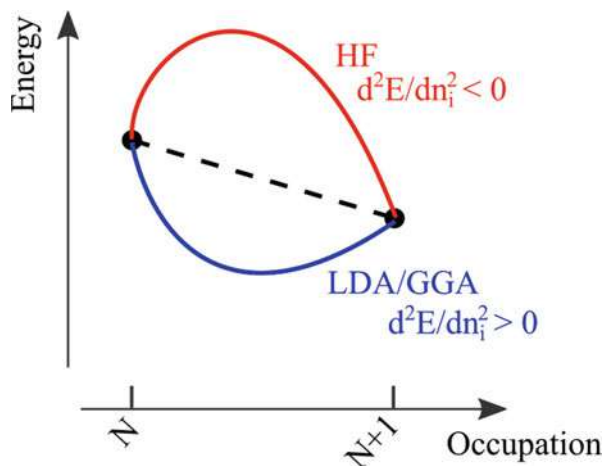
approximation. As a consequence of the convexity of the calculated energy, partial occupations are preferred over integer occupations. This leads to well-known failures of DFT simulations including the underestimation of the energy band gaps, the description of strongly correlated insulators as metals, and the difficulty to account for charge localization.

The tendency of DFT to delocalize charge can be corrected via modifications of the exchange-correlation approximations (To et al. 2005). An efficient remedy is to construct hybrid functionals by mixing LDA or GGA with the Hartree Fock (HF) exchange (Becke 1993), following the general formula

$$E_{XC}^{\mathrm{Hybrid}} = \alpha_1 E_X^{HF} + \alpha_2 E_X^{LDA/GGA} + \alpha_3 E_C^{LDA/GGA}. \tag{8}$$

The mixing ratios α's (and in some screened hybrids the screening length) can be tuned to match physical quantities determined empirically, such as the energy band gap (He and Franchini 2012). Alternatively, the hybrid functional parameters can be determined ab initio by requiring the linearity of the total energy (Kronik et al. 2012; Elmaslmane et al. 2018) or following a self-consistent procedure (He and Franchini 2017) or fitting the parameters to the dielectric function of the material (Chen et al. 2018; Cui et al. 2018). This is possible since HF theory predicts an opposite result as compared to DFT: The energy change upon electronic occupation is described by a concave function of the electronic occupation, $\mathrm{d}^2E/\mathrm{d}n_i^2 < 0$; thus, the charge localization is overestimated by HF. This overestimation is due to neglecting screening effects, which are usually not negligible in solids.

As an alternative to hybrid functionals, which are usually computationally quite demanding, corrections can be applied to the standard formulation of the density functional theory in order to restore the expected behavior for the charge localization (Maxisch et al. 2006; Nolan and Watson 2006). To this class of methods belongs the DFT+U method, where an additional term is added to the expression for the total energy. The DFT+U total energy $E_{\mathrm{DFT}+\mathrm{U}}$ is given by

$$E_{\text{DFT}+\text{U}} = E_{\text{DFT}} + E_{\text{U}}(U, J) \, , \qquad (9)$$

where E_{DFT} is the energy obtained by standard DFT, while E_{U} is an on-site correction arising from a local Hubbard-like Coulomb repulsion (U) and a Hund's parameter (J), including double-counting corrections. Various expressions for E_{U} have been proposed (Anisimov et al. 1991, 1993; Liechtenstein et al. 1995; Dudarev et al. 1998a, b), such that the integer occupation of electronic states is energetically favored. Clearly, the results depend on the choice of U and J, which is not trivial, and, as a matter of fact, these quantities are typically treated as fitting parameters, by adjusting their values such that a specific quantity (e.g., the band gap) is predicted accurately. In order to maintain the ab initio character of DFT, procedures have been defined to calculate the DFT+U parameters from first principles such as the constrained local density approximation (Gunnarsson 1990; Cococcioni and de Gironcoli 2005), where the interaction parameters are obtained by considering the total energy variation with respect to the occupation number of localized orbitals, and the constrained random-phase approximation (cRPA), which allows for an explicit calculation of the matrix element of the screened interactions U and J (Aryasetiawan et al. 2004). Alternatively, in the specific case of excess charge introduced by impurities, the $E_{\text{U}}(U, J)$ term can be substituted by an on-site angular-dependent potential that does not affect the states of the defect-free system. The potential depends on parameters which are tuned to restore the linearity of the total energy (Lany and Zunger 2009; Lany 2011).

2.2   Polaron Formation: Energetics and Structural Distortions

In the framework of DFT, regardless of the specific choice for the correction to the local/semilocal exchange-correlation approximation (DFT+U or hybrids), the stability of polarons is analyzed in terms of a set of different energies. Unfortunately, the notion of the polaronic energies is not well standardized, and it varies according to the authors' preferences. The important ingredients to consider in order to characterize a polaron system are:

1. Type of electronic state: (a) excess charge localized in a lattice site (polaron solution with integer particle occupation on an energy level localized at one lattice site) or (b) delocalized in the entire sample (no polaron, fractional occupation on several lattice sites).
2. Type of structural solution: (a) local distortions at the trapping site (polaron solution) or (b) absence of local distortions (uniform lattice, no polaron). The lattice distortions are obtained fully self-consistently through the minimization of the forces acting on the ions (To avoid any confusion, we specify that also in the non-polaron state, the forces are minimized, but since the excess charge is delocalized in the lattice, the structural changes are uniform and generally very small and become zero for very large supercells.).
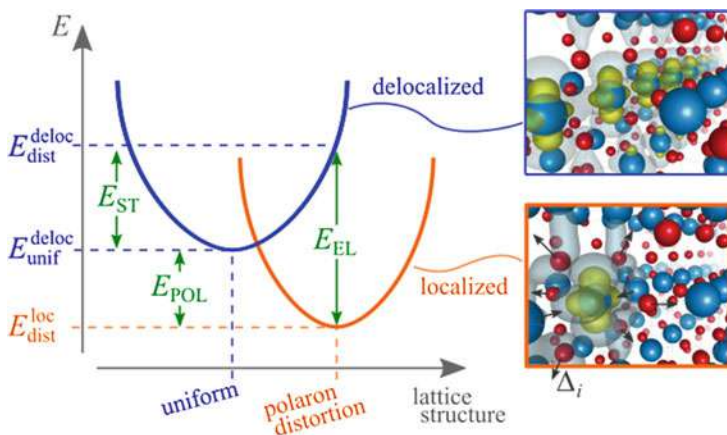
Fig. 4 Polaron energies. Sketch of the polaron formation energy (E_{POL}), the structural energy cost (E_{ST}), and the electronic energy gain (E_{EL}) obtained as combinations of the calculated total energies in the localized and delocalized solutions ($E_{\mathrm{dist}}^{\mathrm{loc}}$, $E_{\mathrm{unif}}^{\mathrm{deloc}}$, and $E_{\mathrm{dist}}^{\mathrm{deloc}}$). The delocalized and localized electronic charge densities are also shown for rutile TiO_2, together with the polaronic lattice distortions Δ_i

Considering that the polaron represents an unpaired electron, DFT-based calculations have to be performed taking spin-polarization into account. In order to force the system to relax into a delocalized solution, it is necessary and often sufficient to use a non-spin-polarized setup (this prescription is valid for both DFT+U and hybrid runs). By selectively switching on or off charge localization and/or lattice distortions, it is possible to compute the total energy of the system in different regimes (see Fig. 4):

$E_{\mathrm{dist}}^{\mathrm{loc}}$: Total energy of the polaron state: charge localization plus lattice distortions;

$E_{\mathrm{unif}}^{\mathrm{deloc}}$: Total energy of the system with delocalized charge carriers and uniform lattice;

$E_{\mathrm{dist}}^{\mathrm{deloc}}$: Total energy of the system with delocalized charge constrained into the lattice structure of the polaron state. Since that solution is only metastable, it might be difficult to realize it in practice (for instance, by enforcing a non-spin-polarized ground state).

In order to compare these energies with each other, the computational setup must be kept fix (except for the spin degree of freedom) in all type of calculations, in particular the total number of electrons (including excess electrons), the unit cell size, and the value of U and J (or the value of the mixing parameter in hybrid-DFT).

Figure 4 sketches the variations of the energy of the system in the delocalized and localized solutions as a function of lattice distortions, considering a quadratic (harmonic) energy versus structure curve. Important insights on the formation of polarons can be obtained by combining the values of $E_{\mathrm{dist}}^{\mathrm{loc}}$, $E_{\mathrm{unif}}^{\mathrm{deloc}}$, and $E_{\mathrm{dist}}^{\mathrm{deloc}}$, which

define the set of polaronic energies E_{POL} (polaron formation energy), E_{ST} (strain energy), and E_{EL} (electronic energy gain):

$$E_{\mathrm{POL}} = E_{\mathrm{dist}}^{\mathrm{loc}} - E_{\mathrm{unif}}^{\mathrm{deloc}}, \tag{10}$$

$$E_{\mathrm{ST}} = E_{\mathrm{dist}}^{\mathrm{deloc}} - E_{\mathrm{unif}}^{\mathrm{deloc}}, \tag{11}$$

$$E_{\mathrm{EL}} = E_{\mathrm{dist}}^{\mathrm{loc}} - E_{\mathrm{dist}}^{\mathrm{deloc}}. \tag{12}$$

The stability of a polaron solution can be analyzed in terms of the polaron formation energy. A negative E_{POL} stands for stable polarons, i.e., the polaronic solution is energetically more convenient than a system with delocalized charge carriers. E_{ST} quantifies the structural cost needed to distort the lattice in order to accommodate the excess charge to form a polaron, whereas E_{EL} is the electronic energy gained by localizing the charge in the distorted lattice via the electron-phonon interaction. The values of E_{ST} and E_{EL} depend on the degree of charge localization and the size of the lattice distortion (see the horizontal shift of the parabolas in Fig. 4) and on the curvature of the parabola in Fig. 4. The three polaronic energies are strongly connected. For instance, E_{POL} can be interpreted as the result of the competition between the structural cost E_{ST} and the electronic energy E_{EL} (Setvin et al. 2014a):

$$E_{\mathrm{POL}} = E_{\mathrm{EL}} + E_{\mathrm{ST}} . \tag{13}$$

In addition to the polaronic energies, purely structural properties can also provide important information on the nature and extension of the polaron state. To this aim, one can define the average bond-length distortion D around the polaron site to quantitatively measure the degree of local structural distortions (Reticcioli et al. 2018a):

$$D = \frac{1}{n} \sum_{i=1,n} |\Delta_i| \tag{14}$$

where $\Delta_i = \delta_i^{\mathrm{dist}} - \delta_i^{\mathrm{unif}}$ is the change of the bond length δ for the ion site i between the polaronic (distorted lattice) and uniform solutions and n indicates the total number of ions considered in the sum (for small polarons typically first and second nearest neighbors).

2.3   Site-Controlled Localization

Typically, charge trapping occurs at different sites of a given material. The various polaronic configurations are characterized by different energies. In general, there is no guarantee that a DFT+U or hybrid functional DFT calculation leads to the global minimum of the polaronic system. In fact, the formation of polarons could spontaneously occur at less favorable lattice sites or occur not at all. Therefore, it

is important to control the site localization of polarons, by inspecting the formation of polarons at different sites, and compare the relative formation energies. For this purpose it is essential to establish a protocol capable to selectively control charge trapping at specific sites.

Since charge localization strongly depends on the initial conditions (input) of the calculation, a selective charge trapping can be achieved by forcing initial perturbations in form of structural distortions or strong on-site Coulomb energy (Deskins et al. 2011; Shibuya et al. 2017; Hao et al. 2015). Starting from a biased setup, it is easier for the system to relax into one desired configuration at the end of the electronic *and* structural self-consistent calculation (output). Structural perturbations can be introduced manually by distorting the local structure around a given atomic site, resembling the expected polaron-induced lattice distortions. Alternatively, the initial perturbation can be achieved by chemical substitution, as explained below for the small electron-polaron case:

Step 1: Chemical substitution at the selected site where charge trapping should occur with an atom containing one more electron than the original native atom (e.g., Ti $\rightarrow$ V). Structural relaxation performed at spin-polarized DFT+U level will yield local lattice distortions around the chosen site.

Step 2: The initial element is reinserted at its original position (e.g., V $\rightarrow$ Ti). It is often necessary to use a larger value of U at the selected site, while the other atoms keep the original value of U. The manual initialization of the local magnetic moment has to take into account the presence of the localized electron at the selected site. A new relaxation is performed starting from the optimized structure obtained in Step 1. The self-consistent run should be able to maintain the polaron solution obtained in Step 1.

Step 3: A final step is necessary, performed by using the original value U for all atoms. In this case it is recommended to initialize the orbitals with the one obtained in Step 2. In fact, by using a random initialization, it can happen that the self-consistent loop will end up in a different polaron solution (different polaron site) or in a delocalized solution (Papageorgiou et al. 2010; Deskins et al. 2009; Meredig et al. 2010; Morgan and Watson 2009).

The effective final localization of the electron at the end of each step can be verified by analyzing the local magnetic moment at the selected site. Step 2 (i.e., using a larger U) usually helps to localize the electron at the selected site. However, it is possible, in simple problems, to skip this step and obtain charge localization at the selected site by simply performing Steps 1 and 3. This strategy can be extended to the hybrid-DFT level, by using the orbitals and the optimized structure obtained in Step 3 as an input for the hybrid functional calculations. In case more than one electron needs to be localized at selected sites, the Steps 1 to 3 can be performed for every selected site separately, one after the other or, alternatively, at the same time.

A systematic use of this strategy allows the identification of the polaronic ground state (global minimum) of the system (Reticcioli et al. 2018a).

2.4   Polaron Dynamics

As mentioned in the introduction, polarons are mobile quasiparticles, whose mobility can be activated or increased by temperature. Polaron hopping can be modeled in DFT, by merging first-principles total energies to the Marcus (Marcus and Sutin 1985) and Emin-Holstein-Austin-Mott (EHAM) formalisms (Deskins and Dupuis 2007; Brunschwig et al. 1980; Spreafico and VandeVondele 2014) or, more effectively, by performing first-principles molecular dynamics (FPMD) (Kresse and Hafner 1993; Kowalski et al. 2010; Setvin et al. 2014a) or nudged elastic band (NEB) calculations (Henkelman et al. 2000; Bondarenko et al. 2015; Janotti et al. 2013).

FPMD data can reveal important insights into polaron dynamics, including site population analysis (sites visited by the polarons during their hopping motion), temperature- and/or concentration-dependent polaron-related transitions, and polaron diffusion trajectories. By means of a detailed statistical analysis, one can acquire information on the interaction with specific lattice sites, with lattice or chemical defects, or on the mutual polaron-polaron interactions (Kowalski et al. 2010; Setvin et al. 2014a; Hao et al. 2015; Reticcioli et al. 2018a). An example of polaron dynamics in oxygen-deficient $SrTiO_3$ is given in Fig. 5, showing the temperature-dependent interaction between two polarons and the (immobile) oxygen vacancy (V_O): at T = 700 K one polaron remains anchored to V_O and the second one has a reduced hopping activity; at higher T (T = 1000 K), both polarons become very mobile and start to explore sites far away from V_O, but the strong Coulomb attraction between the negatively charged polaron and the positively charged V_O regularly brings polarons back into the vicinity of V_O, forming polaron-V_O complexes (Hao et al. 2015).

The statistical analysis is usually performed by investigating the occurrences of charge trapping at each lattice site in the simulation cell (site-decomposed population analysis) and by computing different types of correlation functions aiming to describe the interaction of a polaron with other polarons or lattice defects:

(i) Polaron-polaron site correlation function. The polaron-polaron site correlation function $S_{pol-pol}$ can be defined as the distribution of the site distance i along a particular direction in the crystal, between two polarons at a given time-step t, averaged over all the FPMD time steps τ:

$$S_{pol-pol}(i) = \frac{1}{N}\frac{1}{\tau}\sum_{t=0}^{\tau}\sum_{j}\rho_j(t)\rho_{j+i}(t) \,, \qquad (15)$$

where N is the number of atomic sites and $\rho_j(t)$ indicates the polaronic site density at time t, and it is equal to 1 for the j-th site hosting a polaron and 0 otherwise.
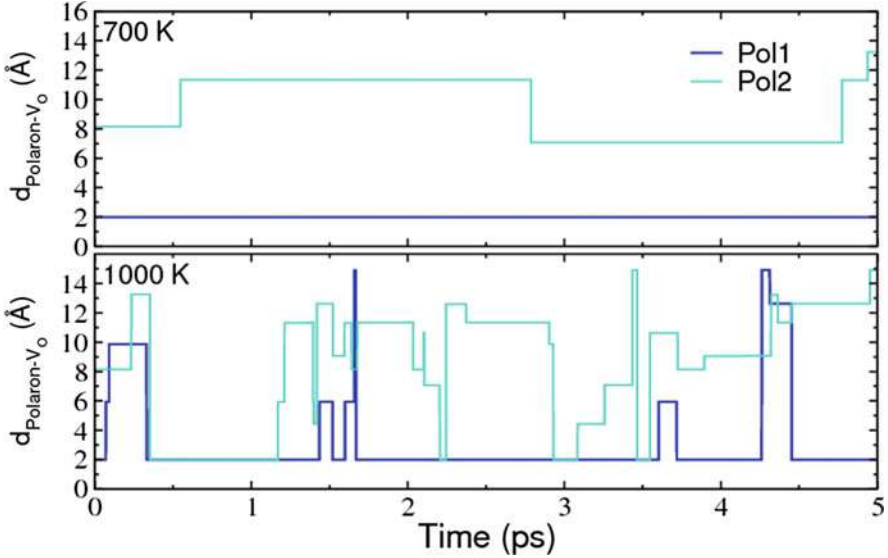
Fig. 5 Polaron dynamics. Example of small polaron dynamics in reduced bulk SrTiO₃. Two excess electrons donated by one oxygen vacancy form two polarons whose mobility is temperature dependent: almost immobile at 700 K and polaron-hopping at 1000 K. The position of the polarons is given in terms of their distance to the oxygen vacancy, which remains at rest at any temperature. (Figure taken from Hao et al. 2015)

(ii) Radial correlation function. Analogously, the polaron-polaron $R_{pol-pol}$ and defect-polaron $R_{def-pol}$ radial correlation functions can be defined as a function of the distance r:

$$R_{pol-pol}(r) = \frac{1}{\tau} \sum_{t=0}^{\tau} \sum_{(q,p)} \delta(|\mathbf{r}_q - \mathbf{r}_p|, r, t) ,$$

$$R_{def-pol}(r) = \frac{1}{\tau} \sum_{t=0}^{\tau} \sum_{(def,p)} \delta(|\mathbf{r}_{def} - \mathbf{r}_p|, r, t) ,$$

(16)

where the variables $\delta(|\mathbf{r}_{def} - \mathbf{r}_q|, r, t)$ and $\delta(|\mathbf{r}_q - \mathbf{r}_p|, r, t)$ assume the value 1 if, at time t, the polaron p is at distance r from the defect at position $\mathbf{r}_{def}$ or from the polaron q at position $\mathbf{r}_q$, respectively, and are 0 otherwise.

These statistical quantities ($S_{pol-pol}$ and $R_{pol-pol}$) furnish information on the spatial distribution of polarons in the lattice, by indicating the statistically more favorable polaron configurations.

The FPMD calculation is also useful to investigate the stability of the polarons under different conditions, as an alternative to the site-controlled localization

strategy discussed above. In fact, a FPMD run leads to a set of high-temperature structures, hosting polarons in various configurations, which might be difficult to access by simple structural relaxations. A useful strategy is to use the high-T structures obtained by FPMD as starting structures for relaxations and check their stability by inspecting the corresponding polaron formation energies E_{POL} and ultimately determine the global minimum (polaronic ground state) (Setvin et al. 2014a; Reticcioli et al. 2017). This strategy is obviously similar to simulated annealing.

As mentioned above, in addition to the FPMD approach, polaron dynamics can be investigated by a sequence of static calculations along pre-defined trajectories, using a linear interpolation scheme (LIS) (Deskins and Dupuis 2007; Janotti et al. 2013) or the NEB approach. In the LIS scheme, the energy barrier for the transition of a polaron between different polaronic configurations can be estimated by static DFT calculations on intermediate distorted structures (Deskins and Dupuis 2007). Such intermediate structures $q(x)$ are obtained by a linear interpolation of the ionic positions assumed at the initial (A) and final (B) polaronic states, i.e., $q(x) = (1 - x)q_A + xq_B$, which define the polaron pathway from A to B (Deskins and Dupuis 2007), see Fig. 6. At each considered configuration, an electronic self-consistent calculation is performed in order to compute the total energy at each step and to construct an energy diagram of the polaron transition as a function of the reaction coordinate x, as those shown in Fig. 6 (Sun et al. 2017). The form of the energy profile determines whether the polaron transfer is an adiabatic or diabatic (i.e., nonadiabatic) process (Fig. 6a) (Deskins and Dupuis 2007). If the coupling between the initial and final state V_{AB} is strong (large V_{AB}, see Fig. 6a), the activation barrier for polaron transfer decreases, and the energy curve exhibits a smooth transition: this corresponds to an adiabatic mechanism which occurs via thermal hopping. The alternative case is when the interaction between the initial and final states is low and the transfer follows a nonadiabatic process. In this case, the energy curve exhibits a cusp at the intermediate state, and the electron transfer occurs via quantum tunneling.

It is important to note that, unfortunately, different functionals can lead to a different description of the polaron transfer process. For instance, DFT+U and hybrid functional approaches can lead to qualitative different results, as shown in Fig. 6 (Sun et al. 2017). In DFT+U, the partial localization is unfavored, the charge carrier is strongly localized in one site only, the interaction between the initial and final configuration is weak, and the hopping occurs typically diabatically. In this case the energy as a function of the lattice distortions resembles the intersection of two parabola, as shown in Fig. 6c. The cusp is most likely an artifact of the one-center terms in the DFT+U approach, which does not allow for localization of the electron between two lattice sites. Conversely, hybrid functionals predict usually an adiabatic hopping, with a more realistic gradual transfer of the electronic charge from one trapping site to the other one, as shown in Fig. 6b. As a consequence, activation barriers for polaronic hopping are usually smaller in hybrid-DFT compared to DFT+U.
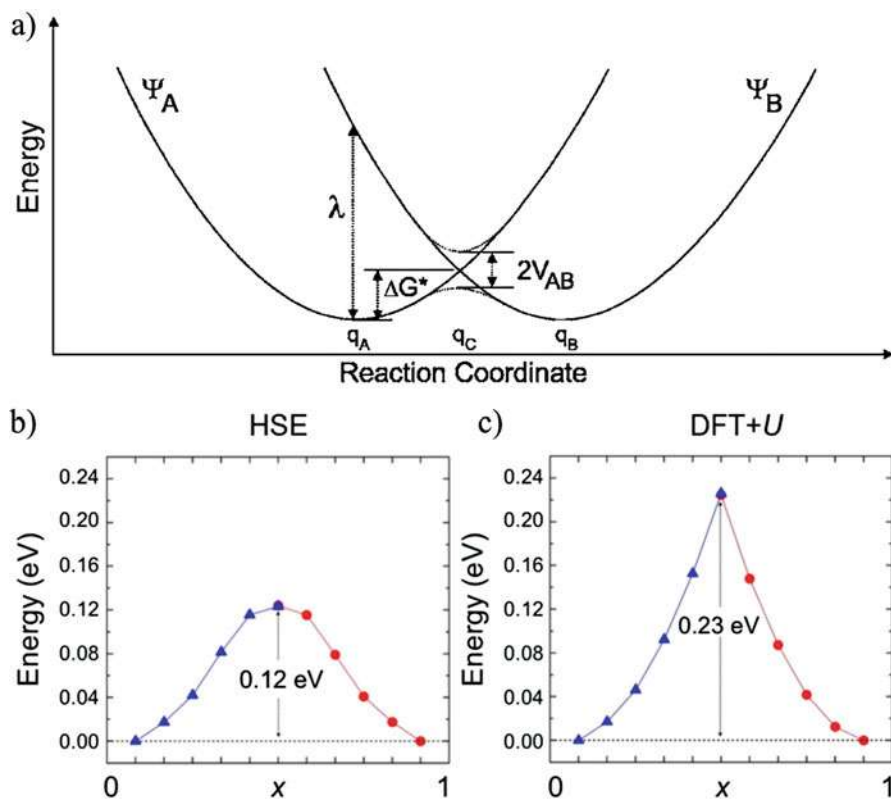
Fig. 6 Polaron transfer. (**a**) General scheme of the Marcus-Emin-Holstein-Austin-Mott theory showing the adiabatic and the nonadiabatic mechanism. (Taken from Deskins and Dupuis 2007). First-principles results of polaron transfer in CeO_2 based on (**b**) hybrid-DFT (HSE parametrization) and (**c**) DFT+U. Depending on the computational scheme, the polaron transfer is described as an adiabatic (HSE) or nonadiabatic (DFT+U) process, resulting in different activation barriers. (This Figure is adapted from Sun et al. 2017)

3   Small Polarons in TiO_2

Charge trapping and the formation of polarons is a pervasive phenomenon in transition metal oxide compounds, in particular, at the surface. As already outlined, polaron formation influences the fundamental physical and chemical properties of materials, manifested by a local alteration of the bond lengths, a change of the formal valence at the specific polaronic site, and the emergence of a characteristic peak localized in the gap region (Nagels et al. 1963; Verdi et al. 2017; Sezen et al. 2015; Freytag et al. 2016). These changes affect virtually all functionalities of the material in practical applications. This section is devoted to the presentation of selected results on polaron effects in TiO_2, one of the most studied transition metal oxides and a prototypical polaronic material. The results are obtained using the array

of first-principles schemes described in the previous section. Considering that small polarons show similar properties among transition metal oxides, the methodologies adopted for TiO_2 should be representative for this class of compounds.

TiO_2 has been studied by several experimental techniques, making the basic properties of this material well characterized (Diebold et al. 2010). TiO_2 is used in many applications, primarily as white pigment in paints and cosmetic products and as a physical blocker for ultraviolet UVA in almost every sunscreen, it is added to cement and tiles (to give the material sterilizing, deodorizing, and antifouling properties), it is utilized in self-cleaning glasses as optical and corrosion-protector (this type of glass is coated in a thin layer of transparent anatase), it is a component of Grätzel cells (in nanostructured form), and it can be used as varistor in electric devices.

The first-principles modeling of polarons is performed by introducing excess charge carriers in a variety of ways:

1. Introducing defects, such as oxygen vacancies, interstitial Ti atoms, or Nb impurities.
2. By suitably changing (by hand) the number of electrons of the system. In this way, one mimics excess charges injected by UV irradiation in perfect (i.e., defect-free) samples (Sezen et al. 2015) or cases where the donor and excess charge are spatially separated.
3. In surface environments, by means of adsorbates (e.g. hydrogen).

As already mentioned, the stability of the polaron state depends on many factors, in particular, the location of the trapping site (a particularly delicate issue at surfaces where the broken symmetry results in various inequivalent Ti sites), the interaction with defects, and the concentration of charge carriers (polaron-polaron interaction). The system is generally characterized by an energy profile with several local polaronic minima (even the delocalized solution might trap in a local minimum), and there is no guarantee that single-shot runs will lead to the identification of the most favorable configuration; this issue is discussed in Sect. 2.3.

3.1 Rutile and Anatase

TiO_2 exists in various crystal structures (polymorphs). Rutile and anatase, the most common and stable TiO_2 polymorphs, bear the formation of polarons (Yin et al. 2018), with some important distinctions. The two polymorphs show different properties upon injection of an excess charge into the system, e.g., anatase becomes metallic, while rutile remains semiconducting upon Nb doping (Zhang et al. 2007) (as mentioned in Sect. 1), consistent with the formation of large and small polarons, respectively. Large polarons have been reported in anatase also by ARPES (Moser et al. 2013) and STM (Setvin et al. 2014a) measurements and interpreted by first-principles calculations (Setvin et al. 2014a; Verdi et al. 2017). Conversely, small polarons have been found to localize easily in pristine or defective rutile, while their

formation in anatase is more seldom. Small polarons in anatase have been observed only in samples containing chemical or structural defects (e.g., oxygen vacancies and step edges) (Setvin et al. 2014a, b).

The theoretical analysis based on first-principles calculations has shed some light on this distinct behavior of rutile and anatase (Setvin et al. 2014a; Spreafico and VandeVondele 2014). In particular it was found that the trapping energy in rutile is significantly larger than in anatase, where a larger U (probably unrealistically large) is required to form a small polaron (see Fig. 7b). By considering a realistic U of ≈ 4 eV, obtained fully ab initio by cRPA, small polarons are formed in rutile but not in anatase (Setvin et al. 2014a). This behavior is explained by inspecting the density of states, shown Fig. 7a. The formation of a polaron involves the perturbation of the conduction band minimum (CBM), which has a different character in the two TiO_2 polymorphs: the CB minimum in anatase lies at a lower energy than in rutile, and it exhibits a wider bandwidth, suggesting that a larger U is required to change this favorable configuration (Setvin et al. 2014a).

On the other hand, at reducing conditions, i.e., under conditions where O vacancies (V_O) are present, small polarons can be formed on the surface of both polymorphs, but their degree of mobility is rather different, as demonstrated by the FPMD analysis shown in Fig. 7c. In rutile, polarons (orange lines) are very mobile and hop preferentially among subsurface and surface sites, while in anatase the polaron is immobile, and it stays attached to the V_O, forming a polaron-V_O defect-like complex (Setvin et al. 2014a) (the following sections are devoted to a more detailed discussion of the small polaron properties of rutile TiO_2).

The representative examples discussed in this chapter are mostly focused on the rutile phase, in particular on the $TiO_2(110)$ surface.

3.2   Small Polarons on the Surface of Rutile TiO$_2$

Rutile TiO_2 crystallizes in a tetragonal symmetry with each Ti atom surrounded by six O atoms in a distorted octahedral configuration (Grant 1959). Neighboring octahedra share one corner and are rotated by 90°. The titanium atoms are in a nominal 4+ oxidation state (Ti^{4+}), while the oxygens are in a 2− state (O^{2-}). Rutile can be easily reduced by intrinsic defects, such as interstitial Ti atoms and double-charged oxygen vacancies, and by doping and molecular adsorption. Regardless of the source used to obtain charge carriers in TiO_2, excess electrons show similar behaviors (Janotti et al. 2013; Deák et al. 2015; Moses et al. 2016): they are trapped at a titanium site, which becomes a Ti^{3+} site, a hallmark revealed and confirmed by EPR experiments (Yang et al. 2009, 2013).

DFT-based simulations are capable to describe the formation of small polarons in rutile, but the stability of the polaronic states, defined in terms of the polaron formation energy, E_{POL}, does depend on the computational approach and on the input parameters (Yin et al. 2018), such as the on-site Coulomb interaction in DFT+U (Calzado et al. 2008; Setvin et al. 2014a) and the amount of mixing in hybrid functional DFT (Spreafico and VandeVondele 2014). The calculated E_{POL}
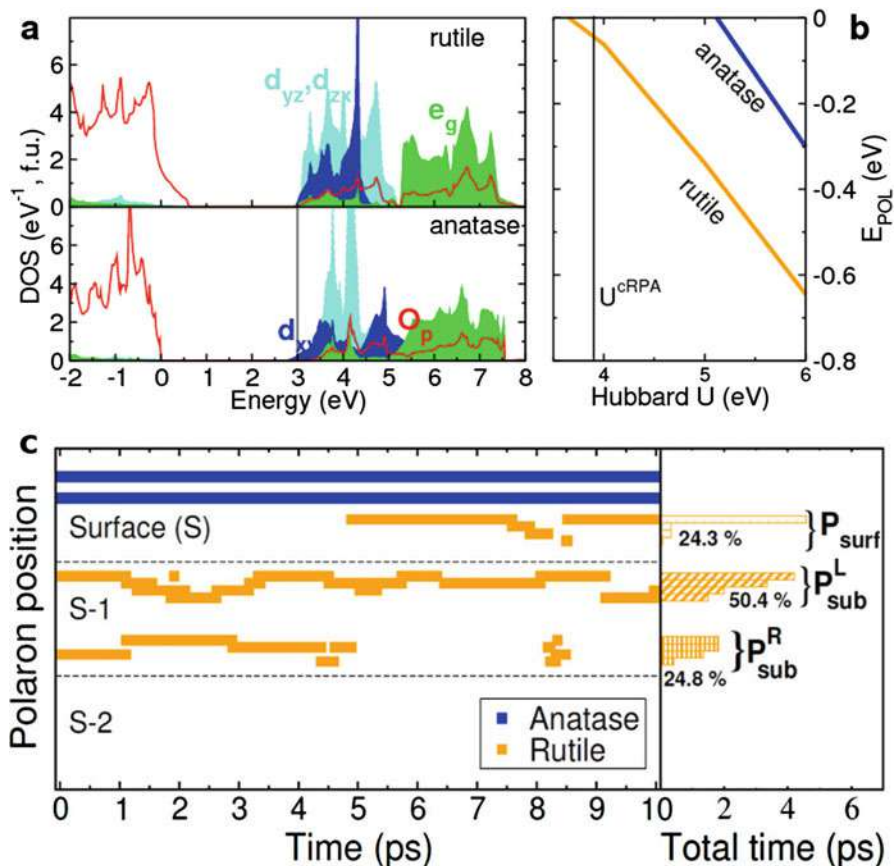
Fig. 7 Small polaron formation in TiO$_2$. Basic polaron properties of rutile and anatase TiO$_2$: (**a**) Density of states and (**b**) polaron formation energies in the bulk phases and (**c**) FPMD trajectories for oxygen-deficient surfaces. (Adapted from Setvin et al. 2014a)

values are spread in a broad range between 0.1 and 0.8 eV (Yin et al. 2018). Despite these quantitative differences, the predicted properties are qualitatively in good agreement among the various methods and also with the experiments. It has been argued that beyond DFT methods, such as RPA, could result in a more reliable quantitative determination of E_{POL} (Spreafico and VandeVondele 2014), but the high computational cost of the RPA method and the difficulty to implement forces at the RPA level (Ramberger et al. 2017) (in existing RPA calculations the lattice has been always relaxed at DFT+U or hybrid-DFT level) have limited its applicability. In fact, the large size of the cell, needed to accommodate the polaron together with the induced lattice distortions, prohibits simulations with more advanced quantum chemistry methods, and DFT+U and to a lesser extent hybrid-DFT remain the most popular choice.

Fig. 8 The rutile TiO₂(110). Side view of the (1 × 1) phase of rutile TiO₂(110). The atoms marked by *A* are the Ti$_{5c}$ atoms on the *S*0 layer and the atoms below; analogously, the atoms marked by *B* are the Ti$_{56}$ atoms on the *S*0 layer and the atoms below. (This Figure is adapted from Reticcioli et al. 2017)
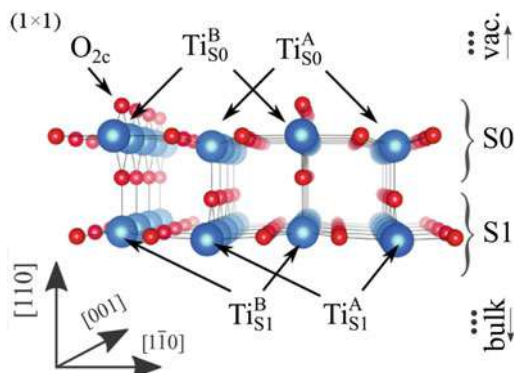
The rutile surface deserves particular attention in the study of polarons, since charge trapping is more favored in the proximity of the surface than in the bulk. The most stable surface of rutile, the TiO₂(110) surface, consists of a bulk-terminated (1 × 1) surface (see Fig. 8), with large relaxations of the atoms at the surface layers, predominantly along the [110] direction. The surface slab is formed by a sequence of tri-layer building blocks, constituted by a central layer comprising an equal number of Ti^{4+} and O^{2-} atoms sandwiched between two layers of oxygen atoms. The topmost (110) layer (the surface layer) is very corrugated since it terminates with undercoordinated (twofold) oxygen atoms (O$_{2c}$) arranged in rows along [001]. The layer below the O$_{2c}$ rows is constituted by trifold coordinated oxygen [001] rows alternated to sixfold coordinated (Ti$_{6c}$) and fivefold coordinated (Ti$_{5c}$) titanium [001] rows (see Fig. 8). Obviously, the surface breaks the symmetry of the Ti and O atoms and exhibits a higher degree of flexibility and different type of structural distortions compared to the bulk-like layers.

The O$_{2c}$ atoms are relatively easy to remove by thermal annealing, irradiation, electron bombardment, or sputtering, because of their coordinative undersaturation, and this is the main source of excess charge for rutile surfaces. In fact, the propensity to O$_{2c}$ removal reduces the system, and each V_O results in the injection of two excess electrons, which are eligible to form small polarons. The ability to control the amount of defects and the concentration of polarons is extremely important, as this can be used to tune the chemical and physical properties of the surface. For instance, under extreme reducing conditions, TiO₂(110) undergoes a structural reconstruction, from the (1 × 1) to a (1 × 2) phase, doubling the periodicity along [1̄10] (Onishi and Iwasawa 1994, 1996; Onishi et al. 1995; Reticcioli et al. 2017). Different models have been proposed to describe the atomic structure of this reconstruction. The debate is still open, but the Ti₂O₃ model proposed by Onishi appears to be the most probable one (Wang et al. 2014b; Reticcioli et al. 2017).

3.2.1   Polaron Configurations and Properties

Preferable trapping site. The first question to address while modeling the polaron formation is the determination of the optimal polaron sites. According to FPMD

calculations, Ti sites in the subsurface layer ($S1$) are the most favorable trapping centers for excess electrons. Specifically, the site-dependent analysis of E_{POL} shows that the most favorable sites are the Ti atoms (Ti_{S1}^A) at the $S1$ layer below the Ti_{5c} atoms, as indicated by the histogram analysis shown in Fig. 7c. Formation of polarons at the $S1$ (Ti_{S1}^B) sites below the Ti_{6c} atoms is less probable, due to an unfavorable local structural relaxation. Polaron formation at deeper layers is hindered by the stiffness of the lattice, which rises the energy cost to locally distort the lattice. The FPMD analysis in Fig. 7c also shows that polaron formation at the surface layer ($S0$) is not particularly favorable, mostly due to the reduced electron screening at the surface and large local distortions.

Polaron orbital symmetry. A second important aspect is the orbital symmetry of the polaron state. Depending on the hosting site, the excess electron occupies different d orbitals. Which d orbital becomes populated is predominantly determined by the local chemical environment, the crystal field splitting, and the flexibility of the neighboring atoms. Polarons in the Ti_{5c} sites on the surface layer (Ti_{S0}^A) are characterized by a $d_{xz} - d_{yz}$ orbital symmetry as evidenced by the polaron charge density plot displayed in Fig. 9a, b. On the other hand, below the surface, in the subsurface layer ($S1$) and in deeper layers, small polarons assume $d_{z^2} - d_{x^2-y^2}$ and $d_{x^2-y^2}$ orbital symmetries, alternately along [110] and [1$\bar{1}$0] (see Fig. 9c–f). This stacking is the result of the orientation of the local environment (i.e., the orientation of the coordination octahedron around the trapping Ti site). The lattice distortions are larger for the Ti_{5c} polarons (the Ti site hosting the polaron relaxes outward along [110], and the surrounding O atoms and the nearest-neighbor Ti atoms are pushed away by the excess electron), while the relaxations in deeper layers are smaller (the Ti-O distance in the TiO_6 octahedron increases, and the nearest Ti^{4+} sites relax toward the polaronic site).

Statistical analysis: polaron-polaron and polaron-vacancy interaction. Another relevant aspect to understand the physics of polarons is the Coulomb-like polaron-polaron and polaron-defect (in this case V_O) interaction. Being positively charged, a V_O acts as an attractive center for the small polarons and, from the structural point of view, renders the local structure more flexible and reduces the structural cost E_{ST} to distort the lattice. Conversely, the polaron-polaron interaction is repulsive and particularly effective at small distances. The relative distance among polarons and between polarons and V_O considerably affects the energy and the degree of localization of the characteristic in-gap polaronic states. For instance, at short distance the interaction between polarons can be strong enough to split the polaron levels and lead to the onset of double in-gap peaks well separated in energy (Reticcioli et al. 2018a). As mentioned in the methodology section, a useful tool to decipher these intricate interactions is the statistical analysis of the polaron energies in the explored configuration space. The results for polarons in rutile TiO_2 are summarized in Fig. 10.

Figure 10a, b clearly show that the polaron energy decreases with increasing polaron-polaron distance and with decreasing polaron-V_O distance. Polaron pairs at nearest-neighbor Ti sites along [001] (i.e., a polaron-polaron distance of about 3 Å) is very unstable, due to the strong polaron-polaron repulsion. This Coulomb
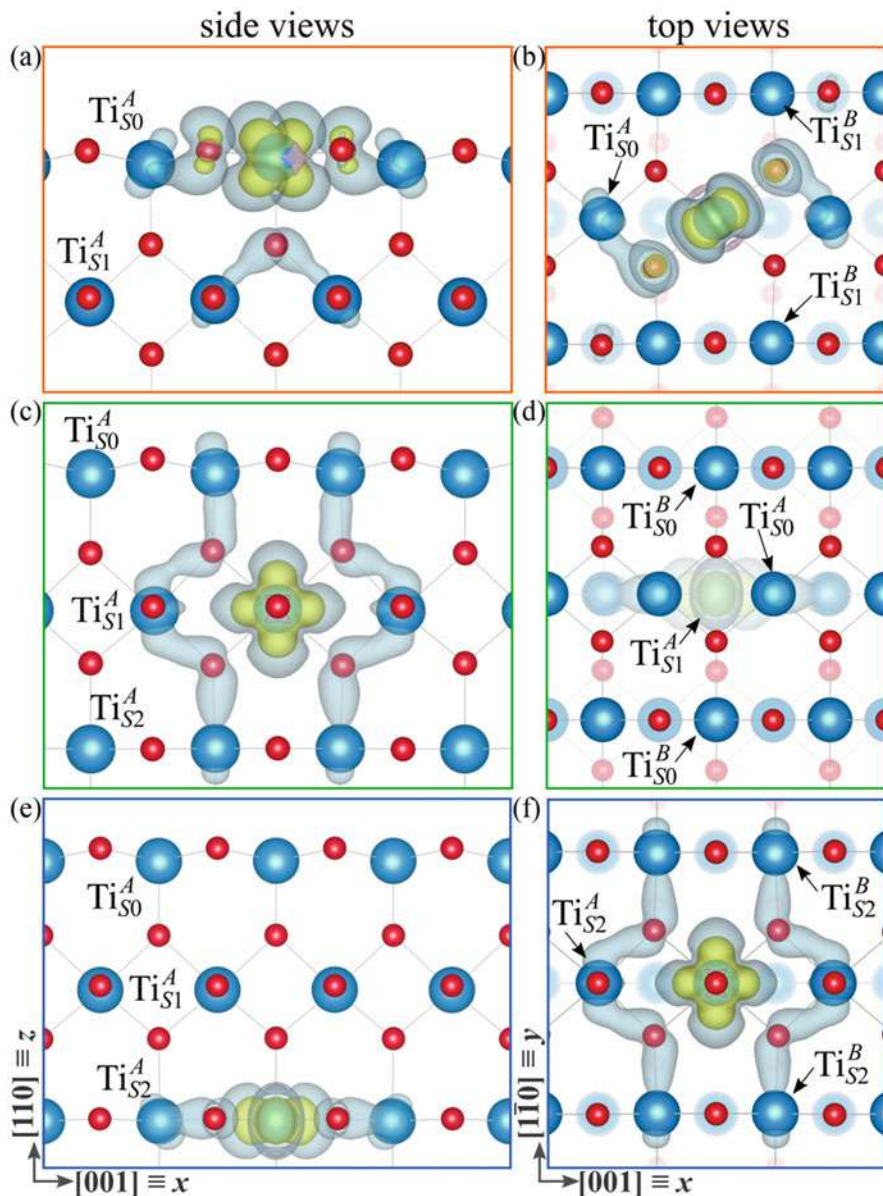
Fig. 9 Small polarons for the rutile TiO$_2$(110) surface. Side (**a, c, e**) and top (**b, d, f**) views of the electronic charge density of polarons on the surface $S0$ (**a, b**), sub-surface $S1$ (**c, d**) and sub-sub-surface $S2$ (**e, f**) layers. The atoms marked by A are the Ti$_{5c}$ atoms on the $S0$ layer and the Ti atoms below; analogously, the atoms marked by B are the Ti$_{6c}$ atoms on the $S0$ layer and the Ti atoms below. (This Figure is adapted from Reticcioli et al. 2018a)
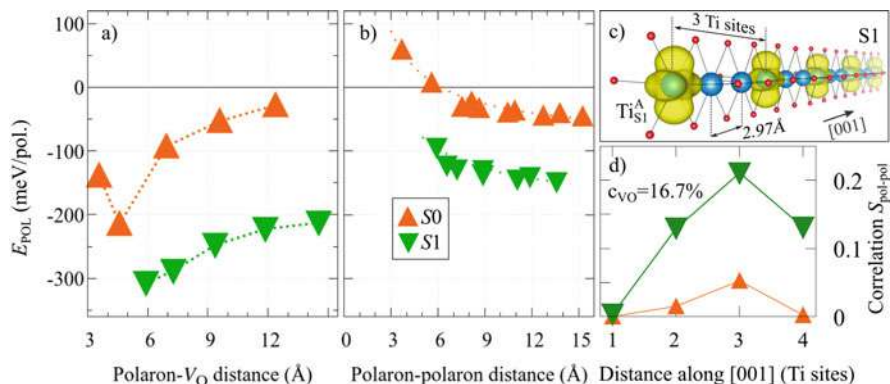
Fig. 10 **Statistics of polarons on rutile TiO$_2$(110):** (**a**) polaron-V_O and (**b**) polaron-polaron interactions, (**c**) sketch of the optimal 1×3 polaron pattern in the sub-surface of rutile TiO$_2$ at the critical V_O concentration (16.7%), and (**d**) the spatial correlation function. (Adapted from Reticcioli et al. 2017)

repulsion is enhanced by the overlap of the polaron charges. In fact, only 70% of the excess charge is confined at the trapping site, while the remaining charge is spread around the surrounding atoms and hinders nearest-neighbor polaronic configurations. For the sake of clarity, it should be noted that the polaron-polaron curve shown in Fig. 10b is obtained for a perfect crystal without any V_O, in order to decouple the polaron-polaron interaction from the polaron-V_O interaction. In defective samples, the optimal polaronic configuration is the result of the balance between these two opposite effects: within a simplified picture, one can conclude that polarons try to maximize their mutual distance and to minimize the polaron-V_O distance.

Clearly the most favorable polaron arrangement and its energetics is concentration dependent. In rutile TiO$_2$, the highest V_O concentration experimentally achievable is $\approx$16.7% (Reticcioli et al. 2017). At low concentrations, polarons occupy $S1$ sites. By progressively increasing the reduction level, the excess electrons form increasingly more polarons at the most convenient sites in $S1$, finally forming a characteristic 1×3 pattern at the optimal concentration of 16.7% (Reticcioli et al. 2017; Shibuya et al. 2017). This is well reproduced by theory, as exemplified by the peak at a three-site distance in the correlation function shown in Fig. 10d. A pictorial view of the 1×3 pattern is shown in Fig. 10c. Larger V_O concentrations are energetically problematic, since additional polarons in $S1$ would result in strong polaron-polaron repulsion. To solve this instability, the system undergoes a structural reconstruction, as discussed in the following.

Surface phase diagram (polaron-induced surface reconstruction). The existence of a critical concentration of polarons determines a limit for the oxygen vacancy formation via O$_{2c}$ removal. In fact, rutile TiO$_2$ is able to host up to 16.7% V_O on the surface, which corresponds to a concentration of 33% polarons in the $S1$ Ti
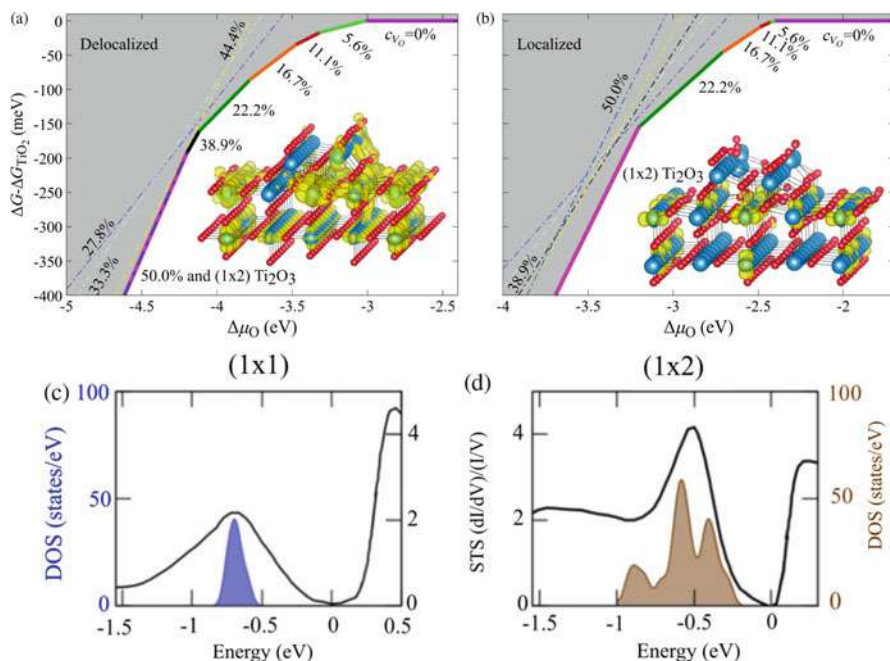
Fig. 11 Rutile TiO$_2$(110) phase diagram. The surface energy of the (1×1) phase with various concentrations of oxygen vacancies (c_{V_O}) and of the (1×2) phase is reported as a function of the oxygen chemical potential μ_O, as obtained by delocalized (**a**) and localized (**b**) calculations. Panels (**c**) and (**d**) display a comparison between the experimental (STS) and simulated (DFT+U) polaronic peak in the (1×1) and reconstructed (1×2) surfaces. (This Figure is adapted from Reticcioli et al. 2017)

[001] rows (each V_O introduces two excess electrons). However, the system can host larger quantities of excess electrons via a structural reconstruction. The (1×1) surface undergoes a (1×2) reconstruction, specifically the Ti$_2$O$_3$ model (Onishi and Iwasawa 1994; Reticcioli et al. 2017), which is able to conveniently handle a larger reduction of the system (50%), as discussed in Reticcioli et al. (2017). This surface reconstruction problem can be described by DFT+U by construction of an appropriate phase stability diagram (see Fig. 11b), by calculating the surface free energy using standard ab initio atomistic thermodynamics (Reuter and Scheffler 2001). The resulting phase diagram is shown in Fig. 11. Close to the critical concentration, the free energy of the reconstructed surface is lower than the corresponding unreconstructed one, thereby marking the structural phase transition. Remarkably, by neglecting polaron effects (panel (a), delocalized setup), DFT+U calculations find the (1×2) phase unstable, in disagreement with observations (Reticcioli et al. 2017).

It is important to remark that the structural reconstruction is associated with changes of the in-gap states, as shown in Fig. 11c, d: The localized polaron peak,
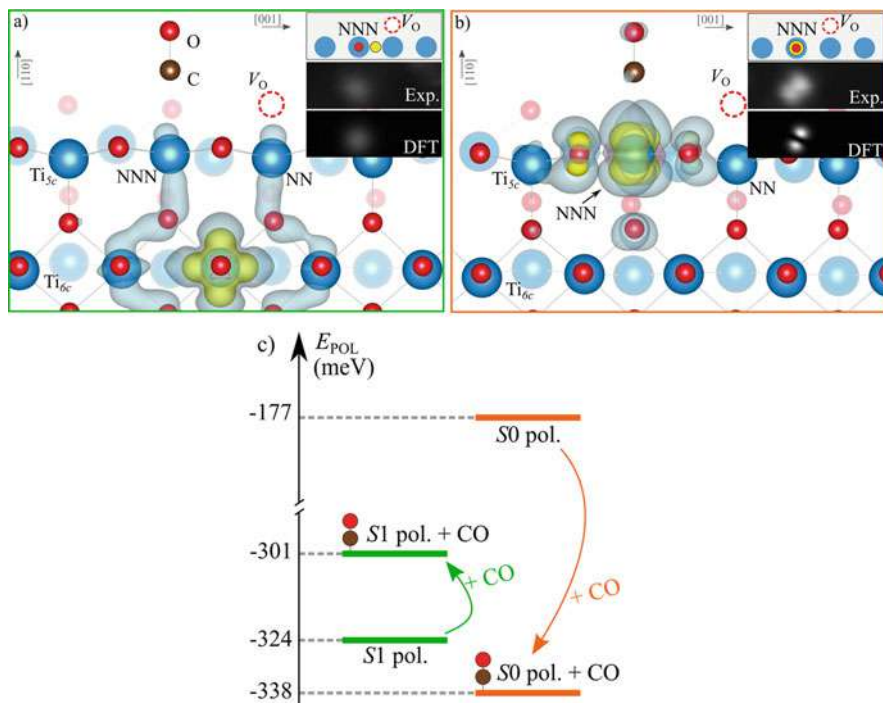
Fig. 12 CO adsorption on rutile TiO$_2$(110). The CO adsorption at Ti$_{5c}$ sites in presence of subsurface (**a**) and surface (**b**) polarons. The insets sketch the geometric configuration and report the simulated and experimental STM signatures. The polaron formation energy perturbed by the CO is compared to the unperturbed case (**c**). (This figure is adapted from Reticcioli et al. 2019)

typical of the (1×1) surface (see the corresponding band dispersion in Fig. 2), becomes broader in the reconstructed (1×2) phase as a consequence of the higher amount of excess charge localized on the surface.

Interaction between polarons and adsorbates. A final paradigmatic case of the importance on polarons on surfaces is the interaction of polarons with adsorbates. This interaction is essential to decode the initial stage of catalytic reactions at defective surfaces. It has been shown that even when adsorbates do not transfer any charge carrier to the substrate, polarons can affect the energetics and configuration of the adsorption process. This is the case for CO adsorption on TiO$_2$(110) (Reticcioli et al. 2019), where the polaron stability is altered by the presence of this adsorbate, as shown in Fig. 12.

Adsorbed CO shows attractive coupling with polarons on the surface layer and repulsive interaction with polarons in the subsurface. As a result, upon CO adsorption, the most convenient polaron site is not in $S1$ anymore: polarons are attracted onto the surface layer ($S0$) and bind with the CO (Fig. 12c). The effect of polarons on $S0$ stands out clearly from the STM analysis of the TiO$_2$ sample. A polaron at $S0$, coupled with the CO, is characterized by a bright filled-state STM

signal with two lobes tilted with respect [001] (Fig. 12b), while the $S1$ polaron, right beneath the CO, generates a feeble single spot, due to a lack of polaronic charge transfer toward the adsorbate (Fig. 12a).

4   Summary

Small polarons can easily form in oxide materials, in particular near the surface where the increased structural flexibility, molecular adsorption, and the easy formation of surface defects facilitate the transfer of charge and its coupling with the lattice. Small polarons can be effectively studied using first-principles techniques in the framework of DFT. The main issue to be aware of is that small polarons involve the localization of charge on a particular lattice site. Usually this goes in hand with a change of the transition metal oxidation state, for instance, in TiO_2, from Ti^{4+} to Ti^{3+}. Such a localization is often not well described by standard semilocal LDA and GGA functionals because of self-interaction errors. Therefore it is compulsory to use functionals beyond the semilocal approximation, for instance, GGA+U or hybrid functionals. It goes without saying that such functionals always involve some empiricism, since increasing U or the amount of exact exchange will favor charge localization and, thus, the formation of polarons. Currently no entirely satisfactory approach exists giving an unambiguous answer whether polaron formation is favorable or not. But there are many principles – such as determination of U using some first-principles approaches or inspection that the energy versus electron number shows a straight line behavior – that can guide our choices. Ultimately, comparison with the experiment, however, remains necessary and mandatory.

While modeling small polarons, the most important aspects to address are the trapping process, polaron mobility, and polaron electronic properties. Using the computational prescriptions presented in this chapter, it is possible to explore the configuration space by either guiding the excess charge to localize in selected lattice sites or by performing FPMD calculations. The analysis of the polaron energies (polaron formation energy, structural energy cost, and electronic energy gain) allows to identify the global polaronic minimum and its fundamental energy profile. In order to fully characterize the polaronic state, it is useful to inspect the degree and extent of the local structural distortions, the mutual interaction among polarons, as well as their coupling with defects. This can be done by performing a statistical analysis in terms of the polaron-polaron site correlation function and radial correlation function. A statistical scrutiny of the FPMD data furnishes also important insight on the mobility pattern. An alternative way to study polaron mobility is Markus theory and the Emin-Holstein-Austin-Mott formalism, which model polaron hopping between two sites and permit to discern between adiabatic or nonadiabatic processes. The electronic characteristics of small polarons are defined by the typical band-gap state formed below the conduction band. For an isolated polaron, this state is well localized and typically resides 1 eV below the conduction band. Its specific orbital character (symmetry of the d level where the excess charge is trapped) depends on the local structural and chemical environment. For interacting

polarons, the situation is more complicated as spin-exchange effects; Coulomb repulsion and hybridization can significantly modify the energy location and the width of the polaron state.

Apart from their intrinsic significance, polarons can induce dramatic changes in the hosting materials and can be used as a mean to drive and control different types of transitions, including metal-to-insulator transitions (not discussed here, an example is given in Franchini et al. 2009), structural transitions, or adsorption processes. Two examples were discussed in this chapter: (i) polaron-induced surface reconstruction (by increasing the polaron concentration the rutile $TiO_2(110)$ undergoes a transition from a (1×1) to a (1×2) surface termination) and (ii) substantial changes of the adsorption energy (the interaction between CO and polarons changes the polaron configuration and affects the CO adsorption scheme).

Finally, it is important to underline that the quality of the theoretical modeling can be assessed by a direct comparison with experimental data, including EPR (by recognizing the change in the formal valence state of the trapping site; for instance in TiO_2 polaron formation changes the Ti valence state from Ti^{4+} to Ti^{3+}), spectroscopy (energy level and band dispersion), or STM (by a direct comparison between measured and simulated STM images).

References

Alexandrov AS, Devreese JT (2010) Advances in polaron physics. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-01896-1

Anisimov VI, Zaanen J, Andersen OK (1991) Band theory and Mott insulators: Hubbard U instead of Stoner I. Phys Rev B 44(3):943–954. https://doi.org/10.1103/PhysRevB.44.943

Anisimov VI, Solovyev IV, Korotin MA, Czyzyk MT, Sawatzky GA (1993) Density-functional theory and NiO photoemission spectra. Phys Rev B 48(23):16929–16934. https://doi.org/10.1103/PhysRevB.48.16929

Aryasetiawan F, Imada M, Georges A, Kotliar G, Biermann S, Lichtenstein AI (2004) Frequency-dependent local interactions and low-energy effective models from electronic structure calculations. Phys Rev B 70(19):195104. https://doi.org/10.1103/PhysRevB.70.195104

Ashcroft NW, Mermin ND (1976) Solid state physics. Cambridge University Press, Cambridge

Becke AD (1993) A new mixing of Hartree–Fock and local density-functional theories. J Chem Phys 98(2):1372–1377

Bertel E, Stockbauer R, Madey TE (1983) Resonant electron emission in Ti and TiO_2. Phys Rev B 27(3):1939–1942. https://doi.org/10.1103/PhysRevB.27.1939

Bondarenko N, Eriksson O, Skorodumova NV (2015) Polaron mobility in oxygen-deficient and lithium-doped tungsten trioxide. Phys Rev B 92:165119. https://doi.org/10.1103/PhysRevB.92.165119

Brunschwig BS, Logan J, Newton MD, Sutin N (1980) A semiclassical treatment of electron-exchange reactions. Application to the hexaaquoiron(ii)-hexaaquoiron(iii) system. J Am Chem Soc 102(18):5798–5809. https://doi.org/10.1021/ja00538a017

Calzado CJ, Hernández NC, Sanz JF (2008) Effect of on-site Coulomb repulsion term U on the band-gap states of the reduced rutile (110) TiO$_2$ surface. Phys Rev B 77(4):045118. https://doi.org/10.1103/PhysRevB.77.045118

Cao Y, Yu M, Qi S, Huang S, Wang T, Xu M, Hu S, Yan S (2017) Scenarios of polaron-involved molecular adsorption on reduced TiO$_2$(110) surfaces. Sci Rep 7(1):6148. https://doi.org/10.1038/s41598-017-06557-6, http://www.nature.com/articles/s41598-017-06557-6

Chambers S (1992) Elastic scattering and interference of backscattered primary, Auger and X-ray photoelectrons at high kinetic energy: principles and applications. Surf Sci Rep 16(6):261–331. https://doi.org/10.1016/0167-5729(92)90016-5, http://linkinghub.elsevier.com/retrieve/pii/0167572992900165

Chen W, Miceli G, Rignanese GM, Pasquarello A (2018) Nonempirical dielectric-dependent hybrid functional with range separation for semiconductors and insulators. Phys Rev Mater 2:073803. https://doi.org/10.1103/PhysRevMaterials.2.073803

Cococcioni M, de Gironcoli S (2005) Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. Phys Rev B 71(3):035105. https://doi.org/10.1103/PhysRevB.71.035105

Cortecchia D, Yin J, Bruno A, Lo SZA, Gurzadyan GG, Mhaisalkar S, Brédas JL, Soci C (2017) Polaron self-localization in white-light emitting hybrid perovskites. J Mater Chem C 5(11):2771–2780. https://doi.org/10.1039/c7tc00366h, http://xlink.rsc.org/?DOI=C7TC00366H

Crawford JH, Slifkin LM Jr (1972) Point defects in solids. Vol. 1: General and Ionic Crystals. Plenum Press, New York

Crevecoeur C, De Wit H (1970) Electrical conductivity of Li doped MnO. J Phys Chem Solids 31(4):783–791. https://doi.org/10.1016/0022-3697(70)90212-X, http://linkinghub.elsevier.com/retrieve/pii/002236977090212X

Cui ZH, Wang YC, Zhang MY, Xu X, Jiang H (2018) Doubly screened hybrid functional: an accurate first-principles approach for both narrow- and wide-gap semiconductors. J Phys Chem Lett 9(9):2338–2345. https://doi.org/10.1021/acs.jpclett.8b00919, pMID: 29669414

Deák P, Aradi B, Frauenheim T (2015) Oxygen deficiency in TiO$_2$: similarities and differences between the Ti self-interstitial and the O vacancy in bulk rutile and anatase. Phys Rev B 92(4):045204. https://doi.org/10.1103/PhysRevB.92.045204

Deskins NA, Dupuis M (2007) Electron transport via polaron hopping in bulk TiO$_2$: a density functional theory characterization. Phys Rev B 75(19):195212. https://doi.org/10.1103/PhysRevB.75.195212

Deskins NA, Rousseau R, Dupuis M (2009) Localized electronic states from surface hydroxyls and polarons in TiO$_2$(110). J Phys Chem C 113(33):14583–14586. https://doi.org/10.1021/jp9037655

Deskins NA, Rousseau R, Dupuis M (2011) Distribution of Ti^{3+} surface sites in reduced TiO$_2$. J Phys Chem C 115(15):7562–7572. https://doi.org/10.1021/jp2001139

Devreese JT (1996) Polarons. In: Encyclopedia of applied physics. Wiley-VCH Publishers, Inc., vol. 14, pp 383–409. http://arxiv.org/abs/cond-mat/0004497

Diebold U, Li SC, Schmid M (2010) Oxide surface science. Ann Rev Phys Chem 61(1):129–148. https://doi.org/10.1146/annurev.physchem.012809.103254

Dudarev SL, Botton GA, Savrasov SY, Humphreys CJ, Sutton AP (1998a) Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA+U study. Phys Rev B 57(3):1505–1509. https://doi.org/10.1103/PhysRevB.57.1505

Dudarev SL, Botton GA, Savrasov SY, Szotek Z, Temmerman WM, Sutton AP (1998b) Electronic structure and elastic properties of strongly correlated metal oxides from first principles: LSDA+U, SIC-LSDA and EELS study of UO$_2$ and NiO. Phys Status Solidi (a) 166(1):429. https://doi.org/10.1002/(sici)1521-396x(199803)166:1<429::aid-pssa429>3.0.co;2-f, https://doi.org/10.1002/(SICI)1521-396X(199803)166:1%3C429::AID-PSSA429%3E3.0.CO;2-F

Elmaslmane AR, Wetherell J, Hodgson MJP, McKenna KP, Godby RW (2018) Accuracy of electron densities obtained via Koopmans-compliant hybrid functionals. Phys Rev Mater 2(4):040801. https://doi.org/10.1103/PhysRevMaterials.2.040801

Emin D (2013) Polarons. Cambridge University Press, New York. http://www.cambridge.org/9780521519069

Feynman RP (1955) Slow electrons in a polar crystal. Phys Rev 97(3):660–665. https://doi.org/10.1103/PhysRev.97.660

Franchini C, Kresse G, Podloucky R (2009) Polaronic hole trapping in doped babio₃. Phys Rev Lett 102:256402. https://doi.org/10.1103/PhysRevLett.102.256402

Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86:253–305. https://doi.org/10.1103/RevModPhys.86.253

Freytag F, Corradi G, Imlau M (2016) Atomic insight to lattice distortions caused by carrier self-trapping in oxide materials. Sci Rep 6(1):36929. https://doi.org/10.1038/srep36929, http://www.nature.com/articles/srep36929

Fröhlich H (1954) Electrons in lattice fields. Adv Phys 3(11):325–361. https://doi.org/10.1080/00018735400101213

Fröhlich H, Pelzer H, Zienau S (1950) Properties of slow electrons in polar materials. Philos Mag J Sci 41(314):221–242. https://doi.org/10.1080/14786445008521794

Gerlach B, Smondyrev MA (2008) Upper and lower bounds for the large polaron dispersion in 1, 2, or 3 dimensions. Phys Rev B 77(17):174303. https://doi.org/10.1103/PhysRevB.77.174303

Gerlach B, Kalina F, Smondyrev M (2003) On the LO-polaron dispersion in D dimensions. Phys Status Solidi (b) 237(1):204–214. https://doi.org/10.1002/pssb.200301770

Giustino F (2017) Electron-phonon interactions from first principles. Rev Mod Phys 89:015003. https://doi.org/10.1103/RevModPhys.89.015003

Grant FA (1959) Properties of Rutile (Titanium Dioxide). Rev Mod Phys 31(3):646–674. https://doi.org/10.1103/RevModPhys.31.646

Gunnarsson O (1990) Calculation of parameters in model Hamiltonians. Phys Rev B 41(1):514–518. https://doi.org/10.1103/PhysRevB.41.514

Hahn T, Klimin S, Tempere J, Devreese JT, Franchini C (2018) Diagrammatic Monte Carlo study of Fröhlich polaron dispersion in two and three dimensions. Phys Rev B 97(13):134305. http://arxiv.org/abs/1803.09608, https://doi.org/10.1103/PhysRevB.97.134305

Hao X, Wang Z, Schmid M, Diebold U, Franchini C (2015) Coexistence of trapped and free excess electrons in SrTiO₃. Phys Rev B 91(8):085204. https://doi.org/10.1103/PhysRevB.91.085204

He J, Franchini C (2012) Screened hybrid functional applied to $3d^0 \rightarrow 3d^8$ transition-metal perovskites LaMO₃ (M = Sc-Cu): influence of the exchange mixing parameter on the structural, electronic and magnetic properties. Phys Rev B 86(23):235117. https://doi.org/10.1103/PhysRevB.86.235117, 1209.0486

He J, Franchini C (2017) Assessing the performance of self-consistent hybrid functional for band gap calculation in oxide semiconductors. J Phys Condens Matter 29(45):454004. https://doi.org/10.1088/1361-648X/aa867e, https://iopscience.iop.org/article/10.1088/1361-648X/aa867e

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113(22):9901–9904. https://doi.org/10.1063/1.1329672

Holstein T (1959) Studies of polaron motion – part I. Ann Phys 8(3):325–342. https://doi.org/10.1016/0003-4916(59)90002-8, http://linkinghub.elsevier.com/retrieve/pii/0003491659900028

Holstein T (2000) Studies of polaron motion – part II – the small polaron. Ann Phys 281(1–2):725–773. https://doi.org/10.1006/aphy.2000.6021, http://www.sciencedirect.com/science/article/pii/S0003491600960214

Janak J (1978) Proof that $\partial E/\partial n_i = \varepsilon$ in density-functional theory. Phys Rev B 18(2):7165–7168. https://doi.org/10.1103/PhysRevB.18.7165

Janotti A, Franchini C, Varley JB, Kresse G, Van de Walle CG (2013) Dual behavior of excess electrons in rutile TiO₂. Phys Status Solidi (RRL) Rapid Res Lett 7(3):199–203. https://doi.org/10.1002/pssr.201206464

Klein A (2012) Transparent conducting oxides: electronic structure-property relationship from photoelectron spectroscopy with in situ sample preparation. J Am Ceram Soc. https://doi.org/10.1111/jace.12143

Kowalski PM, Camellone MF, Nair NN, Meyer B, Marx D (2010) Charge localization dynamics induced by oxygen vacancies on the $TiO_2(110)$ surface. Phys Rev Lett 105(14):5–8. https://doi.org/10.1103/PhysRevLett.105.146405

Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. Phys Rev B 47(1):558–561. https://doi.org/10.1103/PhysRevB.47.558

Kronik L, Stein T, Refaely-Abramson S, Baer R (2012) Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals. J Chem Theory Comput 8(5):1515–1531. https://doi.org/10.1021/ct2009363

Krüger P, Bourgeois S, Domenichini B, Magnan H, Chandesris D, Le Fèvre P, Flank AM, Jupille J, Floreano L, Cossaro A, Verdini A, Morgante A (2008) Defect states at the $TiO_2(110)$ surface probed by resonant photoelectron diffraction. Phys Rev Lett 100(5):055501. https://doi.org/10.1103/PhysRevLett.100.055501

Krüger P, Jupille J, Bourgeois S, Domenichini B, Verdini A, Floreano L, Morgante A (2012) Intrinsic nature of the excess electron distribution at the $TiO_2(110)$ surface. Phys Rev Lett 108(12):126803. https://doi.org/10.1103/PhysRevLett.108.126803

Landau LD (1933) Über die Bewegung der Elektronen im Kristallgitter. Phys Z Sowjet 664(3):644–645

Landau LD, Pekar SI (1965) The effective mass of the polaron. In: Collected papers of L.D. Landau, vol 53. Pergamon Press Ltd. and Gordon and Breach, Science Publishers, Inc. pp 478–483. https://doi.org/10.1016/B978-0-08-010586-4.50072-9, ujp.bitp.kiev.ua, http://linkinghub.elsevier.com/retrieve/pii/B9780080105864500729

Lany S (2011) Predicting polaronic defect states by means of generalized Koopmans density functional calculations. Phys Status Solidi (B) 248(5):1052–1060. https://doi.org/10.1002/pssb.201046274

Lany S (2015) Semiconducting transition metal oxides. J Phys Condens Matter 27(28):283203. https://doi.org/10.1088/0953-8984/27/28/283203, http://stacks.iop.org/0953-8984/27/i=28/a=283203?key=crossref.983b68f3b7276a611510ef5879e02e75

Lany S, Zunger A (2009) Polaronic hole localization and multiple hole binding of acceptors in oxide wide-gap semiconductors. Phys Rev B 80(8):085202. https://doi.org/10.1103/PhysRevB.80.085202

Lenjer S, Schirmer OF, Hesse H, Kool TW (2002) Conduction states in oxide perovskites: three manifestations of Ti^{3+} Jahn-Teller polarons in barium titanate. Phys Rev B 66(16):165106. https://doi.org/10.1103/PhysRevB.66.165106

Liechtenstein AI, Anisimov VI, Zaanen J (1995) Density-functional theory and strong interactions: orbital ordering in Mott-Hubbard insulators. Phys Rev B 52(8):5467–5471. https://doi.org/10.1103/PhysRevB.52.R5467

Linsebigler AL, Lu G, Yates JT (1995) Photocatalysis on TiO_2 surfaces: principles, mechanisms, and selected results. Chem Rev 95(3):735–758. https://doi.org/10.1021/cr00035a013

Locatelli A, Pabisiak T, Pavlovska A, Mentes TO, Aballe L, Kiejna A, Bauer E (2007) One-dimensional Au on TiO_2. J Phys Condens Matter 19(8):082202. https://doi.org/10.1088/0953-8984/19/8/082202, http://stacks.iop.org/0953-8984/19/i=8/a=082202?key=crossref.dabd3b55239f988fa949f60f9e906b9a

Marcus R, Sutin N (1985) Electron transfers in chemistry and biology. Biochim Biophys Acta (BBA) Rev Bioenergetics 811(3):265–322. https://doi.org/10.1016/0304-4173(85)90014-X, http://www.sciencedirect.com/science/article/pii/030441738590%014X

Maxisch T, Zhou F, Ceder G (2006) Ab initio study of the migration of small polarons in olivine Li_xFePO_4 and their association with lithium ions and vacancies. Phys Rev B Condens Matter Mater Phys 73(10):1–6. https://doi.org/10.1103/PhysRevB.73.104301

Meredig B, Thompson A, Hansen HA, Wolverton C, van de Walle A (2010) Method for locating low-energy solutions within DFT+U. Phys Rev B 82(19):195128. https://doi.org/10.1103/PhysRevB.82.195128

Mishchenko A, Prokof'ev N, Sakamoto A, Svistunov B (2000) Diagrammatic quantum Monte Carlo study of the Fröhlich polaron. Phys Rev B 62(10):6317–6336. https://doi.org/10.1103/PhysRevB.62.6317%5Cn, https://doi.org/10.1103/PhysRevB.62.6317

Miyata K, Meggiolaro D, Trinh MT, Joshi PP, Mosconi E, Jones SC, De Angelis F, Zhu XY (2017) Large polarons in lead halide perovskites. Sci Adv 3(8):e1701217. https://doi.org/10.1126/sciadv.1701217

Morgan BJ, Watson GW (2009) Polaronic trapping of electrons and holes by native defects in anatase TiO_2. Phys Rev B Condens Matter Mater Phys 80(23):2–5. https://doi.org/10.1103/PhysRevB.80.233102

Moser S, Moreschini L, Jaćimović J, Barišić OS, Berger H, Magrez A, Chang YJ, Kim KS, Bostwick A, Rotenberg E, Forró L, Grioni M (2013) Tunable polaronic conduction in anatase TiO_2. Phys Rev Lett 110(19):1–5. https://doi.org/10.1103/PhysRevLett.110.196403

Moses PG, Janotti A, Franchini C, Kresse G, Van de Walle CG (2016) Donor defects and small polarons on the TiO_2(110) surface. J Appl Phys 119(18):181503. https://doi.org/10.1063/1.4948239

Mott NF, Stoneham AM (1977) The lifetime of electrons, holes and excitons before self-trapping. J Phys C Solid State Phys 10(17):3391–3398. https://doi.org/10.1088/0022-3719/10/17/022, http://stacks.iop.org/0022-3719/10/i=17/a=022?key=crossref.62bab474a6fe30e880f6b92499b5d42b

Nagels P, Denayer M, Devreese J (1963) Electrical properties of single crystals of uranium dioxide. Solid State Commun 1(2):35–40. https://doi.org/10.1016/0038-1098(63)90388-0, http://linkinghub.elsevier.com/retrieve/pii/0038109863903880

Nolan M, Watson GW (2006) Hole localization in Al doped silica: a DFT+U description. J Chem Phys 125(14):144701. https://doi.org/10.1063/1.2354468

Onishi H, Iwasawa Y (1994) Reconstruction of TiO_2(110) surface: STM study with atomic-scale resolution. Surf Sci 313(1–2):L783–L789. https://doi.org/10.1016/0039-6028(94)91146-0, http://linkinghub.elsevier.com/retrieve/pii/0039602894911460

Onishi H, Iwasawa Y (1996) Dynamic visualization of a metal-oxide-surface/gas-phase reaction: time-resolved observation by scanning tunneling microscopy at 800 K. Phys Rev Lett 76(5):791–794. https://doi.org/10.1103/PhysRevLett.76.791

Onishi H, Fukui Ki, Iwasawa Y (1995) Atomic-scale surface structures of TiO_2(110) determined by scanning tunneling microscopy: a new surface-limited phase of titanium oxide. Bull Chem Soc Japan 68(9):2447–2458. https://doi.org/10.1246/bcsj.68.2447

Pacchioni G (2015) Numerical simulations of defective structures: the nature of oxygen vacancy in non-reducible (MgO, SiO_2, ZrO_2) and reducible (TiO_2, NiO, WO_3) oxides. In: Jupille J, Thornton G (eds) Defects at oxide surfaces, chap 1. Springer, pp 1–28. https://doi.org/10.1007/978-3-319-14367-5

Papageorgiou AC, Beglitis NS, Pang CL, Teobaldi G, Cabailh G, Chen Q, Fisher AJ, Hofer WA, Thornton G (2010) Electron traps and their effect on the surface chemistry of TiO_2(110). Proc Nat Acad Sci 107(6):2391–2396. http://www.pnas.org/content/107/6/2391, https://doi.org/10.1073/pnas.0911349107

Pekar SI (1948) Local quantum states of electrons in an ideal ion crystal. Zh Eksp Teor Fiz 18:481–486

Perdew JP, Levy M (1983) Physical content of the exact Kohn-Sham orbital energies: band gaps and derivative discontinuities. Phys Rev Lett 51(20):1884–1887. http://www.ncbi.nlm.nih.gov/pubmed/622, https://doi.org/10.1103/PhysRevLett.51.1884

Perdew JP, Parr RG, Levy M, Balduz JL (1982) Density-functional theory for fractional particle number: derivative discontinuities of the energy. Phys Rev Lett 49(23):1691–1694. https://doi.org/10.1103/PhysRevLett.49.1691

Possenriede E, Kröse H, Varnhorst T, Scharfschwerdt R, Schirmer OF (1994) Shallow acceptor and electron conduction states in $BaTiO_3$. Ferroelectrics 151(1):199–204. https://doi.org/10.1080/00150199408244743

Prokof'ev NV, Svistunov BV (1998) Polaron problem by diagrammatic quantum Monte Carlo. Phys Rev Lett 81(12):2514–2517. https://doi.org/10.1103/PhysRevLett.81.2514

Ramakumar R, Das AN (2004) Polaron cross-overs and d-wave superconductivity in Hubbard-Holstein model. Eur Phys J B 41(2):197–200. https://doi.org/10.1140/epjb/e2004-00309-4, 0611355

Ramberger B, Schäfer T, Kresse G (2017) Analytic interatomic forces in the random phase approximation. Phys Rev Lett 118:106403. https://doi.org/10.1103/PhysRevLett.118.106403

Rashba EI (2005) Polarons. In: Bassani F, Liedl GL, Wyder P (eds) Encyclopedia of condensed matter physics, chap 4. Elsevier, Oxford, pp 347–355

Reticcioli M, Setvin M, Hao X, Flauger P, Kresse G, Schmid M, Diebold U, Franchini C (2017) Polaron-driven surface reconstructions. Phys Rev X 7(3):031053. https://doi.org/10.1103/PhysRevX.7.031053

Reticcioli M, Setvin M, Schmid M, Diebold U, Franchini C (2018a) Formation and dynamics of small polarons on the rutile TiO_2 (110) surface. Phys Rev B 98(4):045306. http://arxiv.org/abs/1805.01849, https://doi.org/10.1103/PhysRevB.98.045306

Reticcioli M, Sokolović I, Schmid M, Diebold U, Setvin M, Franchini C (2019) Interplay between adsorbates and polarons: CO on rutile TiO_2(110). Phys Rev Lett 122: 016805. https://doi.org/10.1103/PhysRevLett.122.016805

Reuter K, Scheffler M (2001) Composition, structure, and stability of ruo_2(110) as a function of oxygen pressure. Phys Rev B 65:035406. https://doi.org/10.1103/PhysRevB.65.035406

Rønnow HM, Renner C, Aeppli G, Kimura T, Tokura Y (2006) Polarons and confinement of electronic motion to two dimensions in a layered manganite. Nature 440(7087):1025–1028. https://doi.org/10.1038/nature04650, http://www.nature.com/articles/nature04650

Rosenfelder R, Schreiber AW (2001) On the best quadratic approximation in Feynman's path integral treatment of the polaron. Phys Lett A 284(2):63–71. https://doi.org/10.1016/S0375-9601(01)00287-0, http://www.sciencedirect.com/science/article/pii/S0375960101002870

Salje E, Alexandrov AS, Liang WY (2005) Polarons and bipolarons in high-T_c superconductors and related materials. Cambridge University Press, Cambridge

Setvin M, Franchini C, Hao X, Schmid M, Janotti A, Kaltak M, Van De Walle CG, Kresse G, Diebold U (2014a) Direct view at excess electrons in TiO_2 rutile and anatase. Phys Rev Lett 113(8):1–5. https://doi.org/10.1103/PhysRevLett.113.086402

Setvin M, Hao X, Daniel B, Pavelec J, Novotny Z, Parkinson GS, Schmid M, Kresse G, Franchini C, Diebold U (2014b) Charge trapping at the step edges of TiO_2 anatase (101). Angew Chem Int Ed 53(18):4714–4716. https://doi.org/10.1002/anie.201309796

Sezen H, Buchholz M, Nefedov A, Natzeck C, Heissler S, Di Valentin C, Wöll C (2015) Probing electrons in TiO_2 polaronic trap states by IR-absorption: evidence for the existence of hydrogenic states. Sci Rep 4(1):3808. https://doi.org/10.1038/srep03808, http://www.nature.com/articles/srep03808

Shibuya T, Yasuoka K, Mirbt S, Sanyal B (2017) Subsurface polaron concentration as a factor in the chemistry of reduced TiO_2 (110) surfaces. J Phys Chem C 121(21):11325–11334. https://doi.org/10.1021/acs.jpcc.7b00935

Spreafico C, VandeVondele J (2014) The nature of excess electrons in anatase and rutile from hybrid DFT and RPA. Phys Chem Chem Phys 16(47):26144–26152. http://pubs.rsc.org/en/content/articlehtml/2014/cp/c4cp03981e, http://xlink.rsc.org/?DOI=C4CP03981E, https://doi.org/10.1039/C4CP03981E

Stoneham AM, Gavartin J, Shluger AL, Kimmel AV, Muñoz Ramo D, Rønnow HM, Aeppli G, Renner C (2007) Trapping, self-trapping and the polaron family. J Phys Condens Matter 19(25):255208. https://doi.org/10.1088/0953-8984/19/25/255208, http://stacks.iop.org/0953-8984/19/i=25/a=255208

Sun L, Huang X, Wang L, Janotti A (2017) Disentangling the role of small polarons and oxygen vacancies in CeO_2. Phys Rev B 95(24):245101. https://doi.org/10.1103/PhysRevB.95.245101

Teresa JMD, Ibarra MR, Algarabel PA, Ritter C, Marquina C, Blasco J, García J, del Moral A, Arnold Z (1997) Evidence for magnetic polarons in the magnetoresistive perovskites. Nature 386(6622):256–259. https://doi.org/10.1038/386256a0

To J, Sokol AA, French SA, Kaltsoyannis N, Catlow CRA (2005) Hole localization in $[AlO_4]^0$ defects in silica materials. J Chem Phys 122(14):144704. https://doi.org/10.1063/1.1880972

Tuller H, Nowick A (1977) Small polaron electron transport in reduced CeO_2 single crystals. J Phys Chem Solids 38(8):859–867. https://doi.org/10.1016/0022-3697(77)90124-X, http://linkinghub.elsevier.com/retrieve/pii/002236977790124X

Verdi C, Caruso F, Giustino F (2017) Origin of the crossover from polarons to Fermi liquids in transition metal oxides. Nat Commun 8:1–7. https://doi.org/10.1038/ncomms15769

Wang M, Bi C, Li L, Long S, Liu Q, Lv H, Lu N, Sun P, Liu M (2014a) Thermoelectric seebeck effect in oxide-based resistive switching memory. Nat Commun 5(1):4598. https://doi.org/10.1038/ncomms5598, http://www.nature.com/articles/ncomms5598

Wang Q, Oganov AR, Zhu Q, Zhou XF (2014b) New reconstructions of the (110) surface of rutile TiO_2 predicted by an evolutionary method. Phys Rev Lett 113(26):1–5. https://doi.org/10.1103/PhysRevLett.113.266101

Yang S, Halliburton LE, Manivannan A, Bunton PH, Baker DB, Klemm M, Horn S, Fujishima A (2009) Photoinduced electron paramagnetic resonance study of electron traps in TiO_2 crystals: oxygen vacancies and Ti^{3+} ions. Appl Phys Lett 94(16):162114. https://doi.org/10.1063/1.3124656

Yang S, Brant AT, Giles NC, Halliburton LE (2013) Intrinsic small polarons in rutile TiO_2. Phys Rev B 87(12):125201. https://doi.org/10.1103/PhysRevB.87.125201

Yin Wj, Wen B, Zhou C, Selloni A, Liu Lm (2018) Excess electrons in reduced rutile and anatase TiO_2. Surf Sci Rep 73(2):58–82. https://doi.org/10.1016/j.surfrep.2018.02.003, http://linkinghub.elsevier.com/retrieve/pii/S0167572918300128

Zhang SX, Kundaliya DC, Yu W, Dhar S, Young SY, Salamanca-Riba LG, Ogale SB, Vispute RD, Venkatesan T (2007) Niobium doped TiO_2: intrinsic transparent metallic anatase versus highly resistive rutile phase. J Appl Phys 102(1):013701. https://doi.org/10.1063/1.2750407

Defect Equilibria and Kinetics in Crystalline Insulating Oxides: Bulk and Hetero-interfaces

45

Mostafa Youssef, Jing Yang, and Bilge Yildiz

Contents

Abstract

Metal oxides fuel our modern technology. In order to sustain a continuous technological advancement, we strive to understand, predict, and control the

M. Youssef (✉)
Department of Mechanical Engineering, The American University in Cairo, New Cairo, Egypt

Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: mostafa.youssef@aucegypt.edu; myoussef@mit.edu

J. Yang
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

B. Yildiz (✉)
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: byildiz@mit.edu

behavior of metal oxides under different thermodynamic conditions. Since defects are responsible for the major part of the properties of metal oxides, it is highly desirable to have powerful predictive models for defect equilibria and kinetics in oxides and their interfaces. In this chapter we show that a fruitful coupling between electronic structure methods, thermodynamics, electrostatics, and transport theory provides a coherent framework for the study of defect equilibria and kinetics of metal oxides, in their bulk as well as near their interfaces. We present demonstrations of this framework on the ZrO_2 model system and discuss the remaining fronts that need further research and method development efforts.

1 Introduction

Our continuous search for a deeper understanding of metal oxides is motivated by their prevailing presence in modern technologies and the fact that we live in a planet furnished by metal oxides. Energy conversion and storage devices, modern computer memories, modern electronics and spintronics, and many other technologies rely on metal oxides. The earth crust and mantle, the reservoir of our materials, and the origin of earthquakes consist predominantly of metal oxides. Oxides protect our metallic structures from continuous environmental degradation. Even water, the liquid of life, is an oxide.

Upon attaining equilibrium, metal oxides crystallize in ordered crystalline structures that are dictated by the chemical composition of the oxide and the surrounding thermodynamic boundary conditions. Most metal oxides around us and in technological applications are crystalline oxides. But it is also possible to trap metal oxides in nonequilibrium disordered states, amorphous, but these fall beyond the scope of this chapter. Crystalline oxides, however, are not truly perfect. They do contain imperfections which are termed "defects." Defects can be three-dimensional such as cracks and voids, two-dimensional such as grain boundaries and hetero-interfaces, one-dimensional such as dislocations, or zero-dimensional such as point and electronic defects. The presence of zero-dimensional defects minimizes the overall free energy of the crystal mainly because of their configurational entropy. This chapter explains the equilibria and kinetics of zero-dimensional defects in the bulk of metal oxides and also in the presence of an interface between two metal oxides. Whenever we refer to "defects" in the rest of this chapter, it implicitly implies zero-dimensional defects.

One might wonder about the necessity of studying defects in oxides given that we anticipate them to occupy a tiny fraction of the host lattice sites. This anticipation is justified since we are studying crystalline oxides, and if the concentration of defects becomes excessive, then we would rather classify this oxide as amorphous or disordered. The truth is even the smallest fraction of defects can significantly alter the electric, optical, magnetic, mechanical, and thermal properties of the oxide (Kalinin et al. 2012; Kalinin and Spaldin 2013; Baiutti et al. 2015; Cao et al. 2015). In fact there are many useful properties of oxides that are solely

attributed to the presence of defects. A canonical example is ZrO_2. This oxide is a poor electric conductor intrinsically. However, doping and introducing oxygen vacancies in the host crystal can significantly increase its electrical conductivity to the extent of becoming a useful component in solid oxide fuel cells (Minh 1993; Fergus 2006; Wachsman and Lee 2011). A more dramatic effect is phase transitions induced by defects. The whole crystal structure of a metal oxide can change upon the introduction of defects in the lattice. An example for these is the phase transition among the two phases $SrCoO_{2.5}$ and $SrCoO_3$ which is induced by controlling the oxygen defects (Lu and Yildiz 2016). These two phases have completely different electronic, magnetic, and mechanical properties. Thus, defects provide a large engineering landscape to tune the properties of metal oxides to fit into applications, and as such it is of great importance to quantitatively understand the equilibria and kinetics of defects in oxides. "Defect equilibria" is the study of the defect concentrations and their variations under the action of thermodynamic forces in the bulk of the crystal and at higher dimensional defects, whereas "defect kinetics" in this chapter implies the analysis of the mobility of these defects. The thermodynamic forces include temperature, chemical potential, stress, and electromagnetic fields. First, we discuss the equilibria in the bulk of an oxide far away from any interfaces; next, we extend the analysis to hetero-interfaces between two different oxides since this setup appears frequently in applications related to property control at nanoscale; then we discuss defect kinetics; and finally we present a summary and outlook.

2 Defect Equilibria in the Bulk of Oxides

2.1 Theory

We limit our presentation for the bulk equilibria to non-interacting defects in binary insulating metal oxides. The notion of non-interacting defects implies that each defect type is sufficiently isolated from other defects in the lattice such that its abundance can be described by a formation free-energy function that has no dependence on neighboring defects. Interacting defects in oxides is a nontrivial problem which is an active research field (Žguns et al. 2017). The restriction to binary oxides is to avoid unnecessary complexity of representing the defect equilibria at this introductory level and also to avoid a persistent ambiguity in comparing theoretical calculations to experiments (Smyth 2000). If the oxide is not binary, then number of independent chemical potentials of the elements of the oxide is greater than one. Experimentally it is common to control only the chemical potential of oxygen, and the chemical potentials of the rest of the elements equilibrate in an uncontrolled manner. Thus, even if we have a detailed theoretical calculation for the equilibria as a function of all independent chemical potentials, it is still unclear how to compare with experiments. Because of this we limit the presentation to binary oxides for which specifying the chemical potential of oxygen is sufficient to determine the chemical potential of the metal. Finally, we chose to

restrict the presentation to insulating oxides by which we mean oxides that have a nonzero electronic band gap, and as such we are excluding metallic oxides. Metallic oxides are dynamically strongly correlated systems from an electronic point of view (Goodenough 1971; Lankhorst et al. 1997), and understanding their defect physics is still an ongoing research. For example, it is unclear if notions such as charged defects, holes, and electrons which are legitimate descriptions of insulating oxides would apply to metallic ones.

To setup a defect equilibria model of a binary oxide whose chemical formula is M_xO_y, we need first to determine all zero-dimensional defects that we have to take into account. There are two main categories of zero-dimensional defects, electronic and ionic (or point). Ionic defect includes oxygen vacancies V_O, metal vacancies V_M, oxygen interstitials O_i, metal interstitials M_i, oxygen antisites (oxygen ions setting on metal lattice sites) O_M, and metal antisites M_O. These are the so-called native or intrinsic defects. It is worth noting that there might be more than one interstitial site in the unit cell of the oxide (Youssef and Yildiz 2012), and in this case, one has to enlist in the model all possible interstitial defects. In addition to these six types of native ionic defects, it is also possible to have defect complexes due to the association of two or more of these defects. For example, an oxygen vacancy and a metal vacancy can associate together to form a Schottky defect. Another example is the concurrent presence of a metal vacancy and a metal interstitial leading to the formation of the so-called Frenkel defect or the concurrent presence of an oxygen vacancy and an oxygen interstitial leading to the formation of an anti-Frenkel defect. In fact, the antisite defect is a defect complex formed by the association of an interstitial of the metal (oxygen) and an oxygen (a metal) vacancy. The number of possible complex defects is large and grows significantly as the oxide admits extrinsic defects (explained below). Typically one relies on prior literature or experiments to have hints about the possible defect complexes that need to be considered in the equilibria model; otherwise a systematic search can be very time-consuming. Each one of the ionic defects can have multiple charge states. The charge meant here is a relative charge with respect to the original lattice site. So if a lattice site was occupied by an oxide ion $(2-)$ and then this ion is not present anymore, then the resulting defect is $2+$ oxygen vacancy where the $2+$ is the charge of null with respect to $(2-)$. Such a vacancy is denoted by $V_O^{\bullet\bullet}$ using Kröger-Vink notation. If an electron localizes on the vacant site of the $2+$ vacancy, its charge state becomes $1+$, $V_O^{\bullet}$, and if a second electron localizes on that site, then the vacancy becomes neutral, $V_O^{\times}$. It is recommended to study the literature of the particular oxide of interest and perform some tests using the available computational techniques to determine the possible charge states for each defect. For example, one is advised to study references (Kohan et al. 2000; Zhang et al. 2001) for ZnO defects and references (Foster et al. 2002a; Gavartin et al. 2006) for HfO_2 defects. In addition to intrinsic defects, the oxide also can admit extrinsic defects of some element X. These defects can be substitutional on the oxygen sublattice X_O, substitutional on the metal sublattice X_M, or interstitial X_i. Electronic defects, on the other hand, include holes and electrons. If an electron (a hole) lives in the extended conduction (valence) band

states and its charge density is spread over many ions, then it is considered free or a large polaron. Computing the concentration of such a defect requires an integration of the extended band states they occupy as discussed below. On the other hand, if the electron (hole) is localized on a single ion and occupies an in-gap state, then it is considered a small polaron. In computing the concentration of a small polaron, it is treated as an ionic defect since it is localized on a single ionic site.

Once we decided which defects to include in the equilibria model, we have to choose a thermodynamic ensemble that is a best representation for the real life or experimental situation we are trying to model and simulate. To clarify this, we examine the first differential of the internal energy dU of an oxide:

$$dU = TS - PdV + \mu_O dN_O + \mu_M dN_M + \mu_e dN_e + \varphi dq + \overrightarrow{\mathbf{E}} \cdot d\overrightarrow{\mathbf{p}} + \overrightarrow{\mathbf{B}} \cdot d\overrightarrow{\mathbf{I}},\tag{1}$$

where T, S, P, and V are the temperature, entropy, pressure, and volume, respectively. The chemical potentials μ_O, μ_M, and μ_e are those of oxygen, cation M, and electrons, respectively; and N_O, N_M, and n_e are the number of particles of oxygen, cation M, and electrons, respectively. φ is the electrostatic potential and q is the charge transferred. $\overrightarrow{\mathbf{E}}$ is an external homogenous electric field, and $\overrightarrow{\mathbf{p}}$ is the macroscopic electric dipole moment. $\overrightarrow{\mathbf{B}}$ is an external homogenous magnetic field and $\overrightarrow{\mathbf{I}}$ is the macroscopic magnetic dipole moment. Other conjugate pairs (force-displacement) can appear in the above equation if there is a need to include them (e.g., a surface-tension-like term). From each conjugate pair, we have to choose the one that is controlled in the experiment or real-life application we are trying to model. So out of the temperature and entropy pair, certainly we would choose the temperature. Out of the pressure and volume pair, we can choose either, and if we decide to choose the volume, we end up with a Helmholtz-like thermodynamic potential, whereas if we choose the pressure, we end up with a Gibbs-like thermodynamic potential. Typically in experiments we control the chemical potential of oxygen by controlling the oxygen partial pressure, and this in fact determines also the chemical potential of the metal μ_M since both of them are constrained by:

$$\mu_{M_x O_y} = x\mu_M + y\mu_O,\tag{2}$$

where $\mu_{M_x O_y}$ is the chemical potential of the oxide. Out of the chemical potential of electrons and number of electrons, we choose the former as a natural variable not because it can be controlled in experiments, but for theoretical convenience (Freysoldt et al. 2014). Once the list of all natural variables has been selected, a partial Legendre transformation of U can be performed to obtain a convenient thermodynamic potential that describes the oxide crystal (Callen 1985). Minimization of this potential decides the equilibrium state of the crystal and determines an expression for the concentration of the ionic defects in terms of what is known

as the free energy of formation of the defect. Equation 3 gives the most common thermodynamic potential used in modeling defects in the absence of electromagnetic fields which is commonly termed Gibbs free energy (G) in the defects community although it might be more appropriate to call it a grand potential:

$$G = U - TS + PV - \mu_O N_O - \mu_e n_e, \tag{3}$$

where T, P, μ_O, and μ_e were chosen as natural variables. For a given defect d whose charge is q, it is possible to define the Gibbs free energy of formation of this defect to be:

$$G_{d,q}^{form} = \left(U^{def} - U^{perf} \right) + P \left(V^{def} - V^{perf} \right) - T \left(S^{*def} - S^{*perf} \right)$$
$$- \mu_O \left(N_O^{def} - N_O^{perf} \right) + q\mu_e. \tag{4}$$

The definition is a straightforward difference in the terms of the thermodynamic potential between the defective crystal and the perfect crystal at fixed natural variables. There are some important observations to note. *First*, the entropy in the above expression includes all non-configurational entropies (mainly vibrational and band electronic). The contribution of configurational entropies to the abundance of the defects appears in the final expression of the defect concentration. *Second*, it looks unclear how this formation expression represents cation defects when it does not contain any information about the cation chemical potential or number of particles. Here, one appeals to the natural variables we choose for our thermodynamic potential and to the experimental setup we are simulating. Since we only control μ_O, then cation defects need to be created by exchanging oxygen between the crystal and the surroundings of the crystal. Thus, a cation vacancy in M_xO_y is formed by transferring y/x oxygen atoms from the surroundings into the crystal. Similarly creating cation interstitial is formed by destroying an $MO_{y/x}$ unit, such that y/x oxygen atoms are transferred into the surroundings and a metal atom remains in the crystal as interstitial. To summarize, when it comes to cation defects, the formation expression given above is still usable by careful consideration of the number of oxygen particles in the defective and perfect crystals. *Third*, if one is interested in analyzing dopants and their solubility in the oxide, then we need to add an extra term to the thermodynamic potential (Eq. (3) above). If the dopant is X and we select its chemical potential μ_X as a natural variable, then we add the term $(-\mu_X N_X)$ to G. Of course in the perfect crystal of the oxide, the number of particles of the dopant N_X is null. Consequently, Eq. 4 for the formation free energy needs to be updated by a term $(-\mu_X N_X^{def})$. Finally, all the terms of $G_{d,q}^{form}$ are usually computed from first-principles electronic structure methods. Density functional theory (DFT) (Kohn 1965) and its advanced derivatives (Perdew 2013) are becoming the electronic structure method of choice to study the defects of oxides. DFT is a reformulation of the quantum mechanical description of matter from a 3N-dimensional wave function to a three-dimensional electron density, where N is the number of electrons. The

literature of computing $G_{d,q}^{form}$ is mature (Freysoldt et al. 2014) and is evolving rapidly to account for new thermodynamic forces that has not been considered previously for metal oxides such as electric fields (Youssef et al. 2017a) and the simultaneous action of temperature and pressure (Youssef et al. 2017b).

Creating an ionic or electronic defect requires paying a free-energy penalty, but that increases the free energy of the defective crystal with the respect to that of the perfect one. Simultaneously the creation of the defect (ionic or electronic) results in an increase in the configurational entropy which lowers the free energy of the defective crystal with respect to the free energy of the perfect crystal. At a given set of thermodynamic conditions (natural variables discussed above), there is a unique minimal free energy of the defective crystal (less than that of the prefect crystal) obtained by a unique set of concentrations of the defects. The minimization has been carried out analytically (Schmalzried 1995; Kasamatsu et al. 2011) in the limit of non-interacting defects, and the resulting expression for the concentration of an ionic defect d with charge q is:

$$[d^q] = n_d \frac{\exp\left(\frac{-G_{d,q}^{form}}{k_B T}\right)}{1 + \sum\limits_{q^*} \sum\limits_{d^*} \exp\left(\frac{-G_{d^*,q^*}^{form}}{k_B T}\right)} \tag{5}$$

Here n_d represents the number of possible and degenerate ways of including the defect d in the lattice space occupied by the chemical formula M_xO_y. k_B is Boltzmann constant as expected, and the summation in the denominator is taken over all defects d^* with charge q^* that can occupy the same lattice site of the defect d. For example, the defect site of a 2+ oxygen vacancy can also be occupied by a neutral oxygen vacancy, a fluorine substitutional of charge 1+, and any other defect that resides on the oxygen sublattice. If $G_{d,q}^{form} \gg k_B T$ for all defects, then the summation in the denominator can be ignored, and we recover the familiar Boltzmann-like expression. The minimization of the free energy of the defective crystal also produces expressions for the concentration of free electrons $[e]$ and free holes $[h]$ which are:

$$[e] = \int\limits_{E_{CBM}}^{\infty} g_c(E) \frac{dE}{1 + \exp\left(\frac{E - \mu_e}{k_B T}\right)}, \quad [h] = \int\limits_{-\infty}^{E_{VBM}} g_v(E) \frac{dE}{1 + \exp\left(\frac{\mu_e - E}{k_B T}\right)}, \tag{6}$$

where E is the electron energy; E_{CBM} and E_{VBM} are the edges of the conduction and valence bands, respectively; and $g_c(E)$ and $gv(E)$ are the electronic density of states in the conduction and valence bands, respectively. The density of states can be computed using electronic structure methods under the assumption that they do not change with temperature or defect concentration.

Once we specify the control variables T, P, μ_O, and μ_X, then all defect concentrations (ionic and electronic) turn into a single-variable function of the

chemical potential of electrons μ_e. The latter cannot be specified; instead it is determined by the condition of charge neutrality inside the bulk of the oxide, which can be formulated as:

$$\sum_{d,q} q \left[d^q \right] + [h] - [e] = 0. \tag{7}$$

Solving the charge neutrality equation which can be done by a root-finding numerical method such as the bisection method, we can obtain a unique value μ_e that achieves charge neutrality. The latter can be used to determine the concentrations of all defects at the specified thermodynamic conditions. It is important to check that the formation free energies of all defects at the resulting μ_e are greater than zero. A negative formation free energy for a defect indicates crystal instability (Todorova and Neugebauer 2014) and an inaccessible thermodynamic regime for the oxide phase under consideration (e.g., under these thermodynamic conditions that oxide decomposes into a metal and oxygen gas).

After computing the defect concentrations as a function of the natural variables (T, P, μ_O), it remains to choose a graphical representation for these concentrations. For that purpose, it is common to fix T and P and plot the concentrations as a function of the chemical potential of oxygen (or oxygen partial pressure). The resulting plot is usually termed Kröger-Vink diagram. Notice that number of particles (defect concentrations) and chemical potential are conjugate thermodynamic variables, and as such the Kröger-Vink diagram comes in the spirit of T-S and P-V diagrams that are very common in thermodynamics. The slopes of the curves on T-S and P-V diagrams are related to very fundamental thermodynamic properties, such as the specific heat and the compressibility, respectively. Similarly, the slopes of the defect concentrations on Kröger-Vink diagrams (although they do not have names) are of great importance since they can be measured experimentally via electric conductivity measurements. By measuring these slopes, it is possible to guess the underlying dominant charged defects utilizing the charge neutrality condition and the law of mass action. This experimental approach is analogous and synergistic to the modeling and simulation approach which we explained in this section.

2.2 Defect Equilibria in Intrinsic and Doped ZrO$_2$

An example for computed defect equilibria (Kröger-Vink diagram) is shown in the left panel of Fig. 1 for tetragonal ZrO$_2$ (Youssef and Yildiz 2012). It is clear that oxygen chemical potential variation has a dramatic effect on the dominant defects in the oxide. At very high oxygen partial pressure, 2+ oxygen vacancies $V_O^{\bullet\bullet}$ and 1+ holes p_v are charge compensated with 4- zirconium vacancies V_{Zr}''''. Then at around oxygen partial pressure of 10^{-3} atm, the charge compensation mechanism changes to be among $V_O^{\bullet\bullet}$ and -1 electrons. Another key observation is the characteristic slopes of the concentrations of the dominant charged defects, 0 slope at high partial pressure and $-1/6$ slope at low partial pressure, which are obtained from the
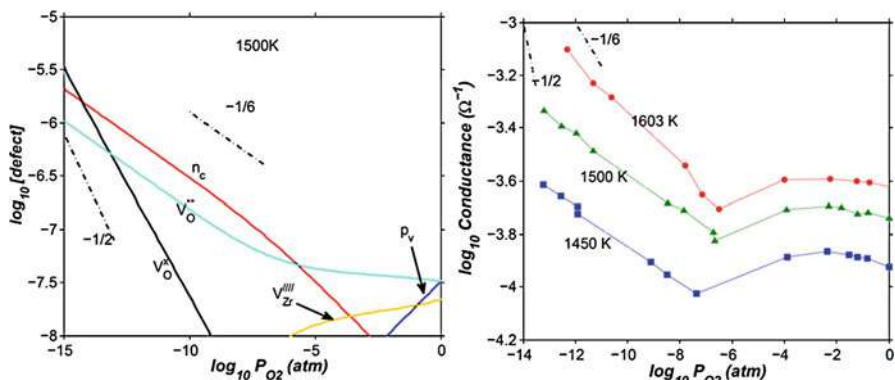
Fig. 1 (left) Calculated Kröger-Vink diagram for tetragonal ZrO_2 at 1500 K. (right) Electric conductivity measurements on tetragonal ZrO_2 by Kofstad and Ruzicka (Kofstad and Ruzicka 1963). (The figure is adapted from Youssef and Yildiz (2012)). In the left panel, n_c and p_v denote conduction band electrons and valence band holes, respectively.

defect equilibria model and also consistent with electric conductivity measurements (Kofstad and Ruzicka 1963) shown on the right panel of the figure. This consistency gives mutual credibility to both theoretical and experimental approaches. Note, however, that at very low oxygen partial pressure, theory predicts the predomination of the neutral oxygen vacancies $V_O^\times$. Being neutral, this defect is not amenable to discovery via electric conductivity experiments. The electronic structure approach that provided the raw formation free energies of the defects presented in Fig. 1 is standard DFT equipped with some corrections (Youssef and Yildiz 2012). It is possible to resort to more accurate electronic structure methods (Perdew 2013) to confirm and possibly correct the predictions of standard DFT especially in situations where experimental validation becomes difficult as in the case of neutral defects.

It is also possible to use defect equilibria models to predict the solubility of an element inside the oxide. This is essentially the sum of the concentrations of all the defects of this element in this oxide at the highest possible chemical potential of the element. A further utility of the defect equilibria models is computing the solubility of an undesirable element X when the oxide is doped intentionally by an element Y. An example for this calculation is shown in Fig. 2 which shows the solubility of hydrogen in monoclinic ZrO_2 as a function of 3D transition metals dissolved in the same oxide (Youssef et al. 2016). Hydrogen is the undesirable dopant here because hydrogen pickup and embrittlement is a main design-limiting factor for zirconium alloys in water-induced corrosion environments (Cox 2005). Three-dimensional transition metals are added intentionally to ZrO_2 to engineer its hydrogen resistance. The plot shows that Co when dissolved in ZrO_2 can increase the solubility of hydrogen by about 5 orders of magnitude compared to undoped ZrO_2.

It is possible to recast the results presented in Fig. 2 into a design paradigm as shown in Fig. 3 which presents the solubility of hydrogen as a function of

temperature and electron chemical potential μ_e (Youssef et al. 2016). The figure
shows that hydrogen solubility has a "valley" shape due to the presence of a
minimum solubility at any given temperature. The task now is to search for
dopants which either minimize the solubility (triangle markers) or do something
else desirable such as maximizing μ_e (circle markers). Note that at maximum μ_e,
the dominant hydrogen defect is substitutional on Zr site which is a less mobile
form of hydrogen compared to interstitial protons that dominate the minimum μ_e
part of the valley. In reference (Youssef et al. 2016), a search was performed to
found dopants that maximize μ_e which would be beneficial for protecting Zr alloys
against the ingress of hydrogen, and a list of not previously explored elements were
identified (Ta, P, Mo, W).

3   Defect Redistribution at Oxide Hetero-Interfaces

3.1   Theory

Modeling point defect equilibria from first-principle calculations have proven itself
a powerful tool for understanding defect chemistry of oxide materials in their bulk
as discussed in the previous section. In this section, we introduce how to extend
this predictive scheme to interface systems. Suppose we have obtained bulk band
structure and defect equilibria of two oxide materials; there are three factors that
drive charged defect redistribution if the two materials form a hetero-interface: first,
the discontinuities of valence band ΔE_v and conduction band ΔE_c immediately at
the interface without considering any defect. Theoretically, this band discontinuity
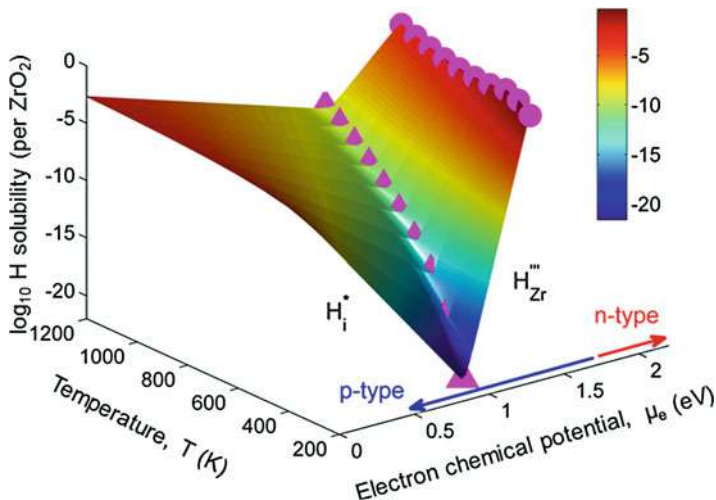
Fig. 3 Hydrogen solubility in monoclinic ZrO_2 as a function of temperature and electron chemical potential at oxygen- and water-rich conditions. The triangles indicate the minimum solubility "valley" of hydrogen, whereas the circles represent the solubility at the maximum electron chemical potential. To the right of the "valley," the dominant hydrogen defect is substitutional on Zr site with a charge of 3-, and to the left of the "valley," the dominant hydrogen defect is interstitial proton. (The figure is adapted from reference Youssef et al. (2016))

corresponds to the large electrostatic potential gradient at ideally abrupt defect-free interfaces. This discontinuity is localized in the few atomic layers in the immediate vicinity of the interface and is independent of external perturbations such as dilute doping and applied electric bias (Van de Walle and Martin 1987). Second, charged species have different electrochemical potentials in equilibrium bulk materials. When two materials are put into contact, electronic and ionic defects start to diffuse to reequilibrate, driven by this chemical potential difference. This leads to buildup of local electric field and in turn drives drift flux in the opposite direction. When the system equilibrates, drift and diffusion fluxes cancel out, leading to a reduced form of drift-diffusion equation with zero flux (Markowich 1990), which is equivalent to the Poisson-Boltzmann formalism used by the solid state ionics field (Maier 2004). Third, atoms near an interface are in a distinct bonding environment compared to atoms in bulk. Defect formation energy will change due to the interfacial lattice relaxation and chemical composition variation. This adds the segregation energy, $E_{D^q}^{seg}$, to defect formation energy of ionic defects in the vicinity of the interface. This region where the segregation energy is non-negligible is termed as the core zone. Noting that now we have an electrostatic distribution profile across the interface, electrochemical potential should be used to describe the energetics and kinetics of charged defects:

$$\tilde{\mu}_{D^q} = \mu_0 + \mu_{cfg} + q\phi. \tag{8}$$

Here μ_0 is the concentration-independent contribution. In the context of our model and because of the complexity of studying interfaces with electronic structure methods, we ignore the formation entropies, $\mu_0 = E_{D^q}^{f,\text{bulk}} + \Delta E_{D^q}^f$. $\Delta E_{D^q}^f$ represents the energetic preference of defects segregating to or depleting from the interfacial core region due to the bonding environment variation. μ_{cfg} represents the configurational entropy generated by defects. ϕ is the electrostatic potential in the space-charge layer. The drift-diffusion flux is defined by:

$$J_{D^q} = -M_{D^q} \nabla \tilde{\mu}_{D^q} \tag{9}$$

where M_{D^q} is the mobility. Assuming that there is no generation or annihilation of defect during the transport process, the transport equation is:

$$\frac{\partial [D^q]}{\partial t} = -\nabla \cdot J_{D^q} \tag{10}$$

Under equilibrium, the fluxes of all species should be zero, which means the electrochemical potential is constant. Equation (10) reduces to:

$$J_{D^q} = 0 \tag{11}$$

The defect electrochemical potential in bulk is given by:

$$\overset{\sim\text{bulk}}{\mu}_{D^q} = E_{D^q}^{f,\text{bulk}} + k_B T \ln\left([D^q]_{\text{bulk}}\right), \tag{12}$$

assuming dilute limit is satisfied. Since equilibrium is reached across the interface, $\tilde{\mu}_{D^q}(z)$ is equal to $\overset{\sim\text{bulk}}{\mu}_{D^q}$ throughout the oxide material, we arrive at the equation for defect concentration:

$$[D^q](z) = [D^q]_{\text{bulk}} \exp\left(-\frac{q\Delta\phi(z) + \Delta E_{D^q}^f}{k_B T}\right) \tag{13}$$

With this expression we can now self-consistently solve Poisson's equation:

$$\frac{d^2\phi_{\text{scl}}(z)}{dz^2} = -\frac{\rho(z)}{\varepsilon\varepsilon_0}, \tag{14}$$

where ε is the relative permittivity of the oxide and ε_0 is the permittivity of vacuum.

Under the constraint of global charge neutrality:

$$\int_0^L \rho(z)dz = 0 \tag{15}$$

where $\rho(z)$ is the total charge concentration including electrons, holes, and all types of charged point defects:

$$\rho(z) = \sum_{D,q} q \left[D^q \right](z) + p_v(z) - n_c(z) \tag{16}$$

Under equilibrium, bulk defect chemistry is retrieved far away from the interface. The corresponding boundary conditions are:

$$\frac{d\phi_{scl}}{dz}\bigg|_{z=0} = 0, \ \ \frac{d\phi_{scl}}{dz}\bigg|_{z=L} = 0 \tag{17}$$

z equals 0 and L represents the starting and end points of the modeled hetero-junction. This framework for describing defect equilibria at hetero-interface is summarized in Fig. 4. The three factors listed above can be obtained from different simulation and experimental methods; here we will focus on first-principles electronic structure approach. Density functional theory (DFT) calculations can provide bulk defect formation energies, segregation energies, and band offsets. These DFT results can be fed into the model described above to generate the defect equilibria profiles across the hetero-interface. Though the discussion here is focused on oxide hetero-interface, similar scheme is also applicable on grain boundaries (Souza 2009), surface segregation (Polfus et al. 2016), and metal/oxide interface (Kasamatsu et al. 2011).

3.2   Defect Equilibria Near ZrO₂(001)/Cr₂O₃ Interface

As an example, we demonstrate the capability of the above described framework on the model system monoclinic-ZrO_2(001)/Cr_2O_3(01-$\bar{1}$2) interface (Yang et al. 2017). This interface system is of interest in the context of high-temperature water-induced corrosion of Zr alloys (Pêcheur et al. 1992; Hatano and Sugisaki 1997). Due to the low solubility of alloying elements (such as Cr) in zirconium oxide, alloying elements precipitate and further get oxidized in the passive film. As a result, Cr alloying of Zr forms Cr_2O_3 precipitates within the ZrO_2 matrix. The space-charge effect of such hetero-interfaces can greatly alter oxygen and hydrogen incorporation kinetics and therefore is important to examine quantitatively. First, we take a close look at the electrostatic profiles close to the interface as plotted in Fig. 5. Here we see a positive built-in potential and downward band bending on ZrO_2 side and upward band bending on Cr_2O_3 side. The positively charged defect species are depleted, and negatively charged defects accumulate in both the space-charge and core zones of ZrO_2, and the reverse happens for Cr_2O_3. It is immediately observable that the electrostatic potential profile is not differentiable at the interface, which corresponds to an abrupt change in electric field. This is because the electric displacement field $D = \varepsilon\varepsilon_0 E$ is continuous throughout, as required by Poisson's equation, but ε takes different values on the two sides. D is not smooth both at the interface and at the
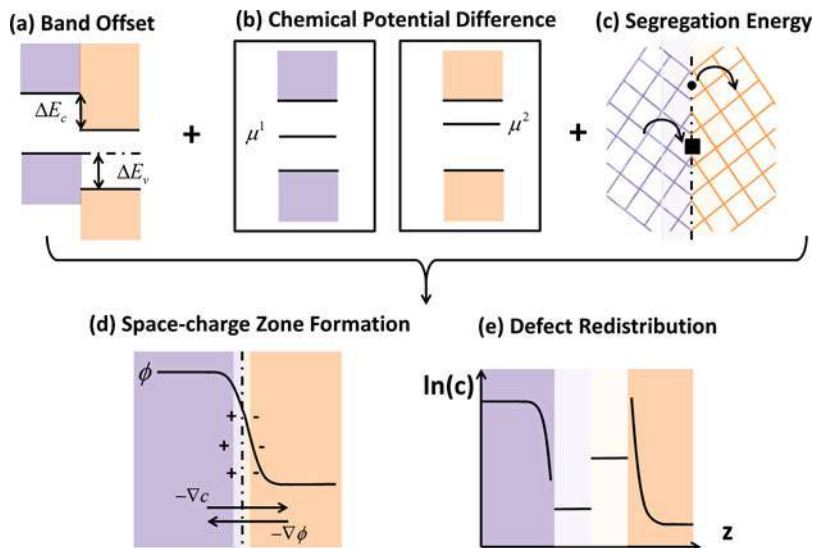
Fig. 4 Schematics for a comprehensive framework for predicting point defect equilibria across hetero-interfaces. Three important factors can be assessed at the atomistic level by DFT calculation: (**a**) band offset of the ideal interface, (**b**) bulk defect equilibria on each side of the interface, and (**c**) defect segregation or depletion due to interfacial structural and compositional change. At a hetero-interface, these three factors determine the difference in chemical potential of charged defects and thus drive their redistribution (Maier 2004; Markowich et al. 1990; Franciosi and Van de Walle 1996). When equilibrium is reached, the electrochemical potential should be constant throughout for electronic defects and constant for ionic defects in the oxide side in which they exist. The steady-state drift-diffusion model can be solved in a self-consistent way which produces (**d**) the electrostatic profiles in the space-charge and core zones, and (**e**) the defect redistribution profiles across the interface with explicit consideration for the interfacial core zone. Upon equilibrium, diffusion flux driven by concentration gradient ∇c should cancel out with drift flux driven by electrostatic potential $\nabla \phi$. In (**d**) and (**e**), light-colored regions in the immediate vicinity of the interface represent core regions, which extend for only several atomic layers. (Figure reproduced from Yang et al. (2017))

boundaries of core layer due to the abrupt change in charge concentration caused by segregated species.

To examine the effect of defect redistribution on transport properties of ZrO_2 with embedded Cr_2O_3 particles, in Fig. 6 we show the spatial distribution of free electrons, electrons and holes, and doubly charged oxygen vacancy across such interface with varying oxygen partial pressure. By looking at the concentration distribution of oxygen-related defects and electronic defects, we can analyze how electron transport and oxygen diffusion are changed in the space-charge layer of ZrO_2 compared to its bulk due to the change in the concentration of these defects in the space-charge layer. In Fig. 6a because the bands bend down on the ZrO_2 side, the concentration of electrons is increased in the space-charge layer compared to bulk, and subsequently holes are decreased as well in this layer. In addition, the
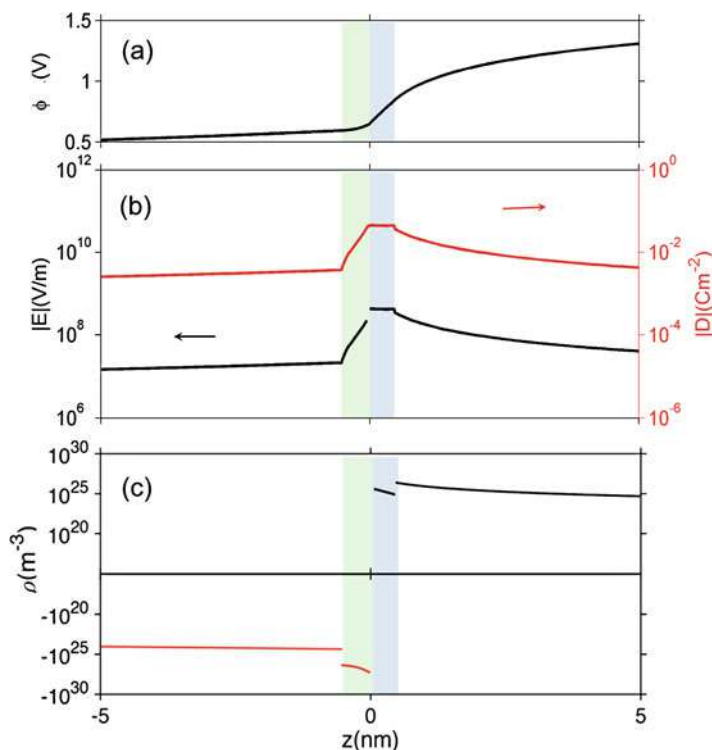
Fig. 5 Electrostatic profiles calculated for m-$ZrO_2(001)$/$Cr_2O_3(01\bar{1}2)$ interface. (**a**) Electrostatic potential ϕ, (**b**) The magnitude of the electric field E and electric displacement field D and (**c**) charge concentration ρ close to the interface at 1200 K and $P_{O_2} = 10^{-5}$ atm. Regions with green and blue background color represent core layer on ZrO_2 and Cr_2O_3 side, respectively. (Figure reproduced from Yang et al. (2017))

sum of electron and hole concentrations is plotted in Fig. 6b. If we assume that the mobilities of electrons and holes are on the same order of magnitude, this quantity should give us an estimation of the change in total electronic conductivity in the space-charge layer. Figure 6b shows that when $P_{O_2} < 10^{-12}$ atm, the space-charge layer of ZrO_2 is in its n-type regime. In this case, the total electronic conductivity in the space-charge zone is increased. On the other hand, when ZrO_2 becomes p-type at higher P_{O_2}, the total electronic conductivity is decreased in the space-charge layer in comparison to bulk due to the downward band bending.

A major advantage of this first-principles-based method is that it decouples 0 K effect from finite temperature effect, allowing for a quantitative comparison of the two effects on the electrostatic properties. In Fig. 7 we plot electrostatic potential V, electric field E, and charge density ρ from both DFT-calculated profile for defect-free interface and space-charge model considering charged defect redistribution. These two effects coexist in every interface between two semiconductors (or
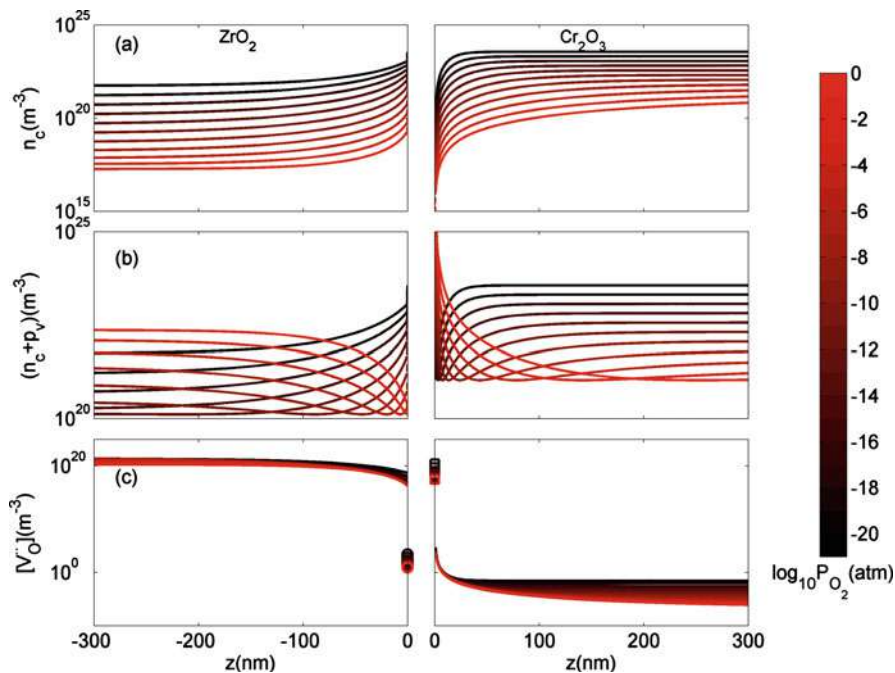
Fig. 6 Spatially resolved concentrations of (**a**) electrons, (**b**) sum of electrons and holes, and (**c**) $V_O^{\bullet\bullet}$ as a function of oxygen partial pressure at T = 1200 K. Circles and squares in panel (**c**) represent the concentrations in the core layer on the ZrO_2 and Cr_2O_3 sides, respectively. (Figure reproduced from Yang et al. (2017))

insulators) and can both influence the interface properties. However, because of the difference in spatial scales of these two effects, either could be dominating at different spatial positions around the interface. The highly localized nature of the field and the dipole arising from the defect-free interface gives rise to the catalytic activity of exposed or barely hidden interfaces (Lee and Selloni 2014). It can also influence the transport processes right at the interface (parallel or across). On the other hand, the field and dipole generated by defects at finite temperatures influence the transport processes over an extended region around the interface.

4   Defect Kinetics

Upon establishing the defect equilibria for a binary metal oxide at a given thermodynamic condition, it is possible to determine the self-diffusivity of the oxide anion or the metal cation under these conditions. For example, the total self-diffusion coefficient D_{tot} of the oxide anion can be written as:
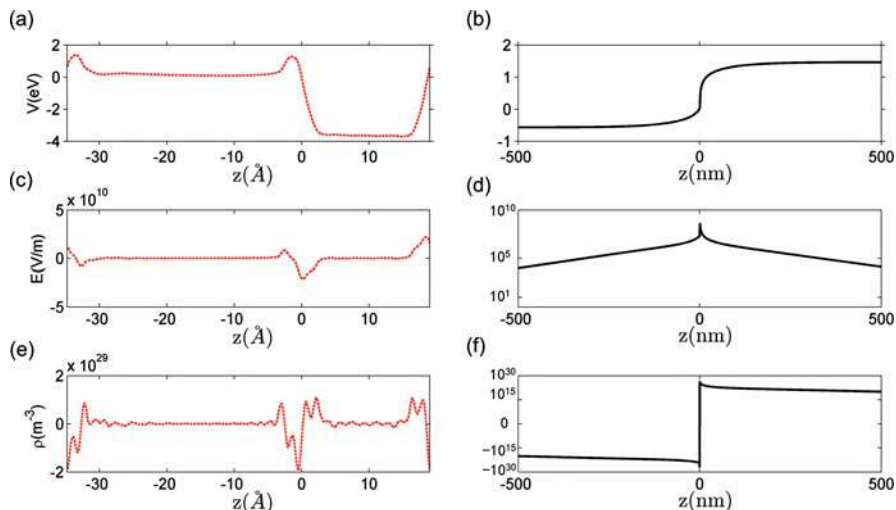
Fig. 7 Macroscopic average of (**a**) electrostatic potential V, (**c**) electric field E, and (**e**) charge density ρ from the perfect interface calculated by DFT at 0 K, compared with the same quantities (**b**), (**d**), and (**f**) contributed by the redistribution of defects at T = 1200 K and $P_{O_2} = 10^{-5}$ atm. (Figure reproduced from (Yang et al. 2017))

$$D_{tot} = \sum_{d,q} \left[d^q \right] D_d^q \tag{18}$$

where the summation is taken over all oxide anion defects d and the possible charge states q for these defects and each term is simply the concentration of the defect $[d^q]$ multiplied by its diffusivity D_d^q. A similar expression can be written for the cation self-diffusivity and also for an extrinsic dopant. We described computing the concentrations of the defects, and it remains to clarify evaluating the defect diffusivities. The most direct way to calculate the diffusivities is through molecular dynamics simulation whose input structure contains defect concentrations corresponding to the defect concentrations evaluated by the equilibria model. Although this approach requires no a priori knowledge of the defect diffusion mechanisms, it has a major drawback. The temperature of the simulation needs to be very high (on the order of 1000 K) so that the defect diffusion events are observable during the course of the simulation. The reason for this is that solid state diffusion is a temperature-activated process and as such at low temperatures the diffusion events become very rare. To overcome this challenge, one can rely on rare-event simulation methods to compute the activation barrier E_b of each diffusion mechanism. The most relevant method for the defect diffusion in the dilute limit is the so-called nudged elastic band (NEB) method and its descendants (Henkelman et al. 2000). For this method, all what is needed is the initial state of the defect, the final state, and of course a total energy calculation method such as DFT. Then, the NEB algorithm is capable of finding the minimum energy pathway and the corresponding activation barrier. Clearly this

approach requires a prior knowledge of the defect diffusion mechanism to prepare the initial and final states accordingly. However, for binary oxides the most relevant diffusion pathways are four which we define here for the case of the oxide anion diffusion, although the same rationale applies to cation diffusion as well. *First*, we have the vacancy mechanism in which an oxide ion hops from a lattice site to fill in a neighboring oxygen vacancy. Equivalently, we can say that the vacancy hops from one lattice site to a neighboring one. *Second* is the antisite mechanism in which an oxygen ion sitting on the cation sublattice hops to a neighboring cation vacancy. *Third* is the direct interstitial mechanism in which an interstitial oxide ion hops from an interstitial site to a neighboring interstitial site. *Fourth* is the interstitialcy mechanism in which an interstitial ion knocks on a neighboring lattice oxygen ion and eventually the original lattice oxygen ion becomes interstitial and the original interstitial becomes a lattice oxygen ion. The net effect is a net displacement for an interstitial oxide ion. The above discussion focused on the diffusion of point defects. Localized electronic defects (small polarons) hop in a similar fashion to point defects, although their hopping involves some complications related to the adiabaticity of the hopping process (Emin and Holstein 1969; Holstein 2000; Austin and Mott 2001; Wang and Bevan 2016). The very same complication arises as well in the theory of electron transfer in chemical reactions (Marcus and Sutin 1985; Marcus 1993). On the other hand, the transport of delocalized electronic defects (large polarons) requires a different approach such as solving Boltzmann transport equation (Madsen and Singh 2006). In this chapter, we focus on ionic defect diffusion.

The complexity of the crystal structure of the metal oxide dictates the next step after sampling the activation barriers of the relevant diffusion mechanisms. Recall that the main goal at this step is calculating D_d^q so that we can obtain the overall self-diffusivity as in Eq. (18). We discuss two approaches to calculate D_d^q. For simple structures which do not contain nonequivalent sites for the ions (e.g., cubic fluorite oxides such as ZrO_2, CeO_2, ThO_2, and UO_2), there is a unique activation barrier E_b along each crystallographic direction as shown schematically in Fig. 8a. In this case, it is straightforward to compute the defect diffusivity from random walk diffusion theory (Tilley 2008).

The premise of random walk diffusion theory is that a defect migrates (jumps) between neighboring equivalent sites corresponding to local energy minima. The migration takes place between the two equivalent sites and across an energy activation barrier E_b which is readily computed using the NEB method. Each migration step is thermally activated and is independent of the previous ones, that is, there is no memory. Under these assumptions, the diffusivity of a defect d whose charge is q a long a crystallographic direction k is given by:

$$D_{d,k}^q = \frac{1}{2}\zeta_k \lambda_k^2 v_{d,k} \exp\left(\frac{-E_b}{k_B T}\right), \qquad (19)$$

Fig. 8 Two possible energy landscapes for the diffusion of a defect along a crystallographic direction in a metal oxide crystal. (**a**) All energy minima are equivalent, and there is a unique activation barrier E_b characterizing diffusion along this direction. (**b**) There are two (or more) nonequivalent energy minima along this path and hence two different activation barriers E_{b1} and E_{b2}
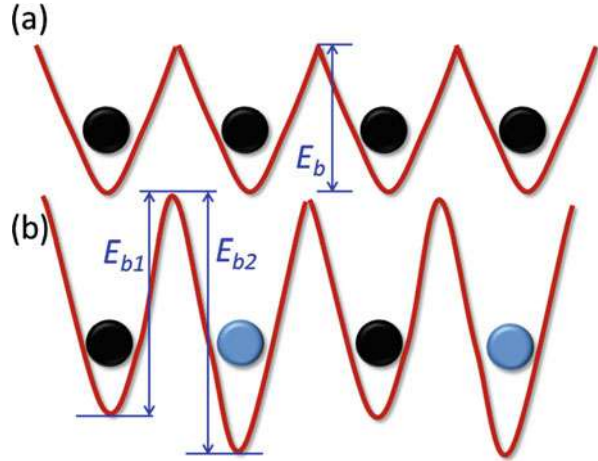
where ζ_k is the multiplicity of the crystallographic direction k. For example, in the cubic fluorite structure, the multiplicity of the [001] direction is 3 because [001], [010], and [100] are crystallographically equivalent. λ_k is the jump length along the k direction. $v_{d,k}$ is the attempt frequency for the defect d to migrate along the crystallographic direction k. For solids this frequency is between 1 and 10 THz. Once $D_{d,k}^q$ is computed for the most relevant crystallographic directions (the ones that lead to short diffusive jumps on the order of the lattice parameter), it is possible to calculate the "isotropic" diffusivity D_d^q of each defect by some averaging procedure. The diffusion of oxygen in tetragonal ZrO_2 (Youssef and Yildiz 2014), and oxygen in ZnO (Erhart and Albe 2006) were analyzed using this approach.

On the other hand, if the crystal structure is complicated due to the presence of many nonequivalent sites for the same ions as in Fig. 8b, then simple random diffusion theory with its analytical form (Eq. 19) cannot account for this complexity. A viable alternative is on-lattice kinetic Monte Carlo (kMC) simulation (Voter 2007). A kMC simulation models in a stochastic way the time evolution of a system under the assumption that all relevant processes and their relative probabilities are known a priori. In the context of modeling the diffusion process, kMC generates a series of atomistic diffusion events according to the probability distribution of these as dictated by the activation barrier of these diffusion jumps. In each kMC step, a diffusion event for the defect is chosen using a random number from a list of pre-calculated diffusion jumps (using NEB). Then, the clock is advanced according the activation barrier of the selected event. A high-activation barrier event corresponds to a long time span and vice versa. After a converged number of kMC steps, the defect diffusivity can be calculated from the mean square displacement $\langle \vec{r}(t)^2 \rangle$ of the defect using Einstein's expression:

$$\langle \vec{r}(t)^2 \rangle = 6D_d^q t \tag{20}$$

Fig. 9 Total oxygen
self-diffusivity in monoclinic
ZrO_2 as a function of
temperature and oxygen
partial pressure. On an
isotherm, the diffusivity is
dominated by 2− interstitial
for oxygen-rich conditions
and 2+ vacancies for oxygen
poor conditions. (The figure
is adapted from Yang et al.
(2018))
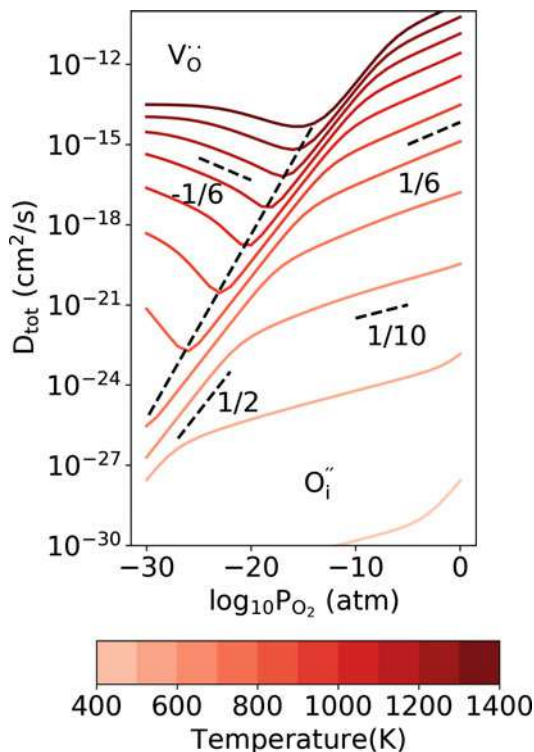
where t is the time. The kMC simulation has to be repeated at each temperature of interest and of course for each defect type. An excellent introduction for the kMC method can be found in reference (Voter 2007). Whether the defect diffusivity was computed using random walk or kMC, the final self-diffusivity of the ion is obtained by summing over all the defects of this ion using Eq. (18). An example of computing oxygen self-diffusivity is shown in Fig. 9 for the case of monoclinic ZrO_2 (Yang et al. 2018) as a function of temperature and oxygen partial pressure. Monoclinic phase is of interest because it is the stable phase for ZrO_2 below 1440 K (Abriata et al. 1986), and so it forms during oxidation of Zr and because of other functional oxides bearing the same structure (Foster et al. 2002b). Generating this plot required computing the overall defect equilibria (Youssef et al. 2016), sampling the activation barriers for all oxygen defects, and performing kMC simulations for those defects that have multi nonequivalent lattice sites. This effort leads to a complete picture for oxygen defect kinetics in monoclinic ZrO_2; in particular we see that each diffusivity isotherm has a minimum. Toward oxygen-rich conditions, the diffusivity increases thanks to interstitial defects, and toward oxygen poor conditions, the diffusivity increases too, but this time due to oxygen vacancies. Such comprehensive picture facilitates defect engineering and design where we can adjust the thermodynamic conditions to obtain a desired diffusivity.

5   Summary and Outlook

The basic question we tried to answer in this chapter is the following: Given certain thermodynamic conditions, and given a certain arrangement for a metal oxide such as uniform bulk or with extended defects such as a hetero-interface, what is the equilibrium concentration and diffusivities of the zero-dimensional defects? The answer to this question requires a coupling between thermodynamics, electrostatics, transport theory, and electronic structure methods. Such coupling provides a fruitful field to study defect equilibria in metal oxides. We gave examples based on ZrO_2 and its interface with Cr_2O_3. However, the approaches and ideas have broad applicability. From a practical point of view, having a model and simulator for defect equilibria with accurate prediction capability opens the door for material design and defect engineering. We can dub this field as defect genomics. For a given application or technology, we can select optimal dopants, optimal interfaces, or optimal thermodynamic conditions that can maximize or minimize certain defects.

In spite of being perceived as a mature field, the physical chemistry of defects still needs major research efforts. A cross-validation between theory and experiments for the defect equilibria models for some of the most technologically important oxides (e.g., iron oxides) is still missing. Metallic oxides which are typically dynamically strongly correlated systems pose severe challenges to electronic structure theory and hence to predicting their defect equilibria. These oxides are becoming increasingly important for various applications, and engineering their defects on physical grounds calls for advances in understanding their electronic structure. This, in turn, requires more advanced methods beyond standard density functional theory. Ternary, quaternary, and higher multielement oxides are rising in many applications, but the number of possible defects that can arise in them is intractable. Automated defect study algorithms might be needed to systematically analyze the possible defects of these systems and to pinpoint the most important ones. Again cross-validation with experiments is crucial here as well. Moreover, studying defect redistribution at interfaces, surfaces, and dislocations of metallic oxides and multielement oxides requires more elaborate effort on top of what we presented for binary insulting oxides. Finally, although we focused here on studying the effect of temperature and chemical potential on the equilibria of defects, studying the effect of other thermodynamic forces such as electromagnetic fields and non-hydrostatic stresses (Youssef et al. 2017a, b; Chi et al. 2018) is equally important and deserves more attention from the defect physical chemistry community.

References

Abriata JP, Garcés J, Versaci R (1986) The O–Zr (Oxygen-Zirconium) system. Bull Alloy Phase Diagr 7(2):116–124

Austin IG, Mott NF (2001) Polarons in crystalline and non-crystalline materials. Adv Phys 50(7):757–812

Baiutti F, Logvenov G, Gregori G, Cristiani G, Wang Y, Sigle W, van Aken PA, Maier J (2015) High-temperature superconductivity in space-charge regions of lanthanum cuprate induced by two-dimensional doping. Nat Commun 6:8586

Callen HB (1985) Thermodynamics and an introduction to thermostatistics, 2nd edn. Wiley, New York

Cao Q, Cheng Y-F, Bi H, Zhao X, Yuan K, Liu Q, Li Q, Wang M, Che R (2015) Crystal defect-mediated band-gap engineering: a new strategy for tuning the optical properties of Ag 2 Se quantum dots toward enhanced hydrogen evolution performance. J Mater Chem A 3(40):20051–20055

Chi Y-T, Youssef M, Sun L, Van Vliet KJ, Yildiz B (2018) Accessible switching of electronic defect type in $SrTiO_3$ via biaxial strain. Phys Rev Mater 2(5):055801

Cox B (2005) Some thoughts on the mechanisms of in-reactor corrosion of zirconium alloys. J Nucl Mater 336(2):331–368

Emin D, Holstein T (1969) Studies of small-polaron motion IV. Adiabatic theory of the Hall effect. Ann Phys 53(3):439–520

Erhart P, Albe K (2006) First-principles study of migration mechanisms and diffusion of oxygen in zinc oxide. Phys Rev B 73(11):115207

Fergus JW (2006) Electrolytes for solid oxide fuel cells. J Power Sources 162(1):30–40

Foster AS, Lopez Gejo F, Shluger AL, Nieminen RM (2002a) Vacancy and interstitial defects in hafnia. Phys Rev B 65(17):174117

Foster AS, Shluger AL, Nieminen RM (2002b) Mechanism of Interstitial Oxygen Diffusion in Hafnia. Phys Rev Lett 89(22):225901

Franciosi A, Van de Walle CG (1996) Heterojunction band offset engineering. Surf Sci Rep 25(1):1–140

Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86(1):253–305

Gavartin JL, Muñoz Ramo D, Shluger AL, Bersuker G, Lee BH (2006) Negative oxygen vacancies in HfO_2 as charge traps in high-k stacks. Appl Phys Lett 89(8):082908

Goodenough JB (1971) Metallic oxides. Prog Solid State Chem 5:145–399

Hatano Y, Sugisaki M (1997) Auger Electron Spectroscopy Study of Oxidation Behavior of Iron and Chromium in $Zr(Fe,Cr)_2$ Precipitate in Zircaloy-4. J Nucl Sci Technol 34(3):264–268

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113(22):9901–9904

Holstein T (2000) Studies of Polaron Motion: Part II. The 'Small' Polaron. Ann Phys 281(1):725–773

Kalinin SV, Spaldin NA (2013) Functional Ion Defects in Transition Metal Oxides. Science 341(6148):858–859

Kalinin SV, Borisevich A, Fong D (2012) Beyond Condensed Matter Physics on the Nanoscale: The Role of Ionic and Electrochemical Phenomena in the Physical Functionalities of Oxide Materials. ACS Nano 6(12):10423–10437

Kasamatsu S, Tada T, Watanabe S (2011) Theoretical analysis of space charge layer formation at metal/ionic conductor interfaces. Solid State Ion 183(1):20–25

Kofstad P, Ruzicka DJ (1963) On the Defect Structure of ZrO_2 and HfO_2. J Electrochem Soc 110(3):181–184

Kohan AF, Ceder G, Morgan D, Van de Walle CG (2000) First-principles study of native point defects in ZnO. Phys Rev B 61(22):15019–15027

Kohn W (1965) Self-Consistent Equations Including Exchange and Correlation Effects. Phys Rev 140(4A):A1133–A1138

Lankhorst MHR, Bouwmeester HJM, Verweij H (1997) Thermodynamics and Transport of Ionic and Electronic Defects in Crystalline Oxides. J Am Ceram Soc 80(9):2175–2198

Lee JH, Selloni A (2014) TiO_2/Ferroelectric Heterostructures as Dynamic Polarization-Promoted Catalysts for Photochemical and Electrochemical Oxidation of Water. Phys Rev Lett 112(19):196102

Lu Q, Yildiz B (2016) Voltage-Controlled Topotactic Phase Transition in Thin-Film $SrCoO_x$ Monitored by In Situ X-ray Diffraction. Nano Lett 16(2):1186–1193

Madsen GKH, Singh DJ (2006) BoltzTraP. A code for calculating band-structure dependent quantities. Comput Phys Commun 175(1):67–71

Maier J (2004) Physical chemistry of ionic materials: ions and electrons in solids. John Wiley & Sons, Chichester

Marcus RA (1993) Electron transfer reactions in chemistry. Theory and experiment. Rev Mod Phys 65(3):599–610

Marcus RA, Sutin N (1985) Electron transfers in chemistry and biology. Biochim Biophys Acta BBA – Rev Bioenerg 811(3):265–322

Markowich PA, Ringhofer CA, Schmeiser C (1990) Semiconductor equations. Springer-Verlag, Wien

Minh NQ (1993) Ceramic Fuel Cells. J Am Ceram Soc 76(3):563–588

Pêcheur D, Lefebvre F, Motta AT, Lemaignan C, Wadier JF (1992) Precipitate evolution in the Zircaloy-4 oxide layer. J Nucl Mater 189(3):318–332

Perdew JP (2013) Climbing the ladder of density functional approximations. MRS Bull 38(9): 743–750

Polfus JM, Bjørheim TS, Norby T, Bredesen R (2016) Surface defect chemistry of Y-substituted and hydrated $BaZrO_3$ with subsurface space-charge regions. J Mater Chem A 4(19): 7437–7444

Schmalzried H (1995) Chemical kinetics of solids. VCH Verlagsgesellschaft mbH, Weinheim

Smyth DM (2000) The defect chemistry of metal oxides. Oxford University Press

Souza RAD (2009) The formation of equilibrium space-charge zones at grain boundaries in the perovskite oxide $SrTiO_3$. Phys Chem Chem Phys 11(43):9939–9969

Tilley RJD (2008) Defects in solids, Wiley-Interscience

Todorova M, Neugebauer J (2014) Extending the Concept of Defect Chemistry from Semiconductor Physics to Electrochemistry. Phys Rev Appl 1(1):014001

Van de Walle CG, Martin RM (1987) Theoretical study of band offsets at semiconductor interfaces. Phys Rev B 35(15):8154–8165

Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Sickafus KE, Kotomin EA, Uberuaga BP (eds) Radiation effects in solids. Springer, Netherlands, pp 1–23

Wachsman ED, Lee KT (2011) Lowering the Temperature of Solid Oxide Fuel Cells. Science 334(6058):935–939

Wang Z, Bevan KH (2016) Exploring the impact of semicore level electronic relaxation on polaron dynamics: An adiabatic *ab initio* study of $FePO_4$. Phys Rev B 93(2):024303

Yang J, Youssef M, Yildiz B (2017) Predicting point defect equilibria across oxide hetero-interfaces: model system of ZrO2/Cr2O3. Phys Chem Chem Phys 19(5):3869–3883

Yang J, Youssef M, Yildiz B (2018) Oxygen self-diffusion mechanisms in monoclinic Zr_2 revealed and quantified by density functional theory, random walk analysis, and kinetic Monte Carlo calculations. Phys Rev B 97(2):024114

Youssef M, Yildiz B (2012) Intrinsic point-defect equilibria in tetragonal ZrO_2: Density functional theory analysis with finite-temperature effects. Phys Rev B 86(14):144109

Youssef M, Yildiz B (2014) Predicting self-diffusion in metal oxides from first principles: The case of oxygen in tetragonal ZrO_2. Phys Rev B 89(2):024105

Youssef M, Yang M, Yildiz B (2016) Doping in the Valley of Hydrogen Solubility: A Route to Designing Hydrogen-Resistant Zirconium Alloys. Phys Rev Appl 5(1):014008

Youssef M, Van Vliet KJ, Yildiz B (2017a) Polarizing Oxygen Vacancies in Insulating Metal Oxides under a High Electric Field. Phys Rev Lett 119(12):126002

Youssef M, Yildiz B, Vliet KJV (2017b) Thermomechanical stabilization of electron small polarons in $SrTiO_3$ assessed by the quasiharmonic approximation. Phys Rev B 95(16): 161110

Žguns PA, Ruban AV, Skorodumova NV (2017) Ordering and phase separation in Gd-doped ceria: a combined DFT, cluster expansion and Monte Carlo study. Phys Chem Chem Phys 19(39):26606–26620

Zhang SB, Wei S-H, Zunger A (2001) Intrinsic n-type versus p-type doping asymmetry and the defect physics of ZnO. Phys Rev B 63(7):075205

Oxide Heterostructures from a Realistic Many-Body Perspective

46

Frank Lechermann

Contents

Abstract

Oxide heterostructures are a new class of materials by design that open the possibility for engineering challenging electronic properties, in particular, correlation effects beyond an effective single-particle description. This short review tries to highlight some of the demanding aspects and questions, motivated by the goal to describe the encountered physics from first principles. The state-of-the-art methodology to approach realistic many-body effects in strongly correlated oxides, the combination of density functional theory with dynamical mean-field theory, will be briefly introduced. Discussed examples deal with

F. Lechermann (✉)
Institut für Theoretische Physik, Universität Hamburg, Hamburg, Germany
e-mail: Frank.Lechermann@physnet.uni-hamburg.de

prominent Mott-band and band-band type of oxide heterostructures, where different electronic characteristics may be stabilized within a single architectured oxide material.

1   Introduction

Since the early days of quantum solid-state research, transition-metal oxides are known to pose very challenging problems in condensed matter physics and materials chemistry (Mott and Peierls 1937). The delicate balance between localization and itinerancy of electrons stemming from partially filled d-shells in these systems is at the origin of many intriguing phenomena. Spanned by the basic electronic phases, namely, band insulator, Mott insulator, and conventional metal, the rich transition-metal-oxide phase space encloses, e.g., delicate forms of transport and magnetism (Imada et al. 1998). The identification of high-temperature superconductivity (coined high-T_c) in CuO_2-based systems (Bednorz and Müller 1986), eventually well above the liquid-nitrogen temperature, still marks the hallmark finding in this group of compounds.

A new chapter in the research of such materials systems opened in the early 2000s, when systematic studies of oxide heterostructures appeared (e.g., Ohtomo et al. 2002). Ever since, that topical field belongs to a key focus in condensed matter and materials science (see, e.g., Zubko et al. (2011), Hwang et al. (2012), and Chakhalian et al. (2014) for reviews). Important advancements in experimental preparation techniques allow researchers to design oxide materials beyond known appearance in nature. Materials developments from this area may hence be relevant for future technological applications. However, due to the unique combination of the well-known demanding physics of bulk transition-metal oxides with the modern architecturing possibilities, oxide heterostructures furthermore challenge known paradigms in condensed matter physics. The coexistence and mutual affection of intriguing bulk-like electronic phases, each up to now associated with a given oxide compound, within a single oxide heterostructure may be imagined. Moreover, due to the seminal role of the interface in these novel systems, new exotic phases, unknown in present bulk compounds, may emerge.

As of today, already a faithful theoretical investigation of bulk transition-metal oxides with a weakly screened local Coulomb interaction remains a difficult task. Restriction to (effective) single-particle schemes often misses key oxide phenomenologies, such as, e.g., the many-body concept of Hubbard bands or local-moment formation. Therefore, the standard first-principles tool of materials science, density functional theory (DFT) in the Kohn-Sham representation, is in many cases not sufficient. Realistic methods beyond conventional DFT that highlight the importance of strong local Coulomb interactions in connection with true many-body electron states are on their way of becoming rather routinely applicable to the correlated electronic structure of bulk oxides. But the sophisticated structural aspects of oxide heterostructures still promote the task to another level of complexity.

The goal of this review is to highlight the many-body character of some oxide heterostructure problems and to discuss recent first-principles approaches to deal with the physics. We focus on works based on the combination of DFT with the dynamical mean-field theory (DMFT). This is motivated by the fact that the DFT+DMFT framework has evolved to the state-of-the-art realistic many-body approach to cope with the strong-coupling problem of challenging oxide electron systems. In order to set the stage, the starting Sect. 2 provides a brief reminder on the phenomenology of strongly correlated electrons and some thoughts on the relevance of heterostructure physics. Section 3 delivers an account of the DFT+DMFT technique, while Sect. 4 presents a selected overview of existing applications to oxide heterostructure problems.

2   Preliminary Considerations

In the given context, it is relevant to appreciate the many-body phenomenology of correlated electrons and its difference to sole effective single-particle picturings. Therefore, a quick summary of the key characteristics of strongly correlated electron systems renders the text more self-contained. Second, some general thoughts on the research motivation in the field of oxide heterostructures seem in order.

2.1   Brief Reminder on Strong Electronic Correlations in a Solid

In various transition-metal oxides, the key many-body aspects reside within the electronic subsystem. The Pauli principle and the mutual Coulomb interaction among the electrons give rise to an intricate many-particle wave function. While the former exchange physics, in its pure appearance, may be cast into a Slater determinant, the explicit interaction leads to more intriguing modifications of the wave function. Electronic correlations, as defined from quantum chemistry, are routed in the latter. Though the Coulomb interactions is long-ranged, screening processes in the solid restrict in most cases its actual relevant range to rather short distances. To a good approximation for many systems, a sole *local* interaction appears sufficient to describe the dominant physics. Given that viewpoint, the Hubbard model stands out as the seminal picturing of the competition between localization and itinerancy among interacting condensed matter electrons, i.e.:

$$H_{\text{Hubbard}} = -t \sum_{\langle i,j \rangle \sigma} \left(\hat{c}_{i\sigma}^{\dagger} \hat{c}_{j\sigma} + \text{h.c.} \right) + U \sum_{i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \tag{1}$$

whereby i, j are site indices, $\sigma = \uparrow, \downarrow$ is the spin projection, and $\hat{n}_{\sigma} = \hat{c}_{\sigma}^{\dagger} \hat{c}_{\sigma}$. The nearest-neighbor hopping t describes the electron tendency to delocalize on the lattice, giving rise to the noninteracting band structure, whereas the Hubbard U marks the local Coulomb interaction. The given basic Hamiltonian is appropriate for a (effective) one-band problem. Description of multiband systems asks in addition
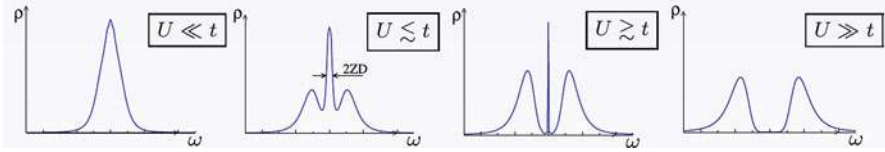
Fig. 1 Local spectral function $\rho(\omega)$ of the Hubbard model on a three-dimensional lattice for different U/t ratios. The quantity D marks the half bandwidth. The nomenclature $A(\omega) = \rho(\omega)$ is often understood, if the explicit k-dependence in the complete spectral function is not discussed

for an implementation of the local Hund's physics, as enforced by the Hund's exchange J_{H}.

Depending on the ratio U/t, Fig. 1 displays the principle behavior of the k-integrated spectral function $\rho(\omega) = \sum_{\mathbf{k}} A(\mathbf{k}, \omega)$ on a three-dimensional lattice for the so-called half-filled case $n = 1$, i.e., nominally one electron per lattice site. For small U/t, the system is a very good metal, close to a Fermi gas, with maximum spectral weight at the Fermi level ε_{F}. Note that in that regime, the spectral function reduces to the common density of states (DOS). On the contrary for large U/t, the lattice is in an insulating state, since the electrons localize in real space because of the strong Coulomb repulsion. In spectral terms, removing/adding an electron is associated with states in the lower/upper Hubbard band. Importantly, this *Mott-insulating* state is strikingly different from a conventional band insulator which arises due to complete band filling in reciprocal space.

For U/t in between the named limits, the system is in the most interesting regime of a correlated metal. Subtle spectral-weight balancing between the emerging Hubbard bands and the coherent quasiparticle (QP) peak close to ε_{F} takes place. The QP peak is the reminder of the original band states, i.e., encloses states with a well-defined wave vector k, but strongly renormalized and with a weight $Z \leq 1$. The apparent *band-narrowing* included in Z deviating from unity is readily understood: the effective hopping t is reduced due to the electron's hesitation to delocalize when having to pay the Coulomb penalty U while meeting another electron on the nearby lattice site. An existing QP weight Z defines a *Fermi liquid*, since more formally, the quantity Z is derived from the electronic self-energy $\Sigma(\mathbf{k}, \omega)$ via $Z = (1 - \partial \Sigma / \partial \omega|_{\omega \to 0})^{-1}$ (see, e.g., Imada et al. (1998) for more details). Notably, the electronic self-energy integrates all many-body effects and is the key quantity of the interacting system. Most common metals are Fermi liquids, but it is obvious that there may exist other forms of correlated metals, where a linear self-energy term at low energy is missing. Robust QPs are then absent and the metal is, somewhat vaguely, termed *non-Fermi liquid* (NFL). Various unusual electronic phases of transition-metal oxides are associated or in proximity to NFL characteristics, such as, e.g., in the phase diagram of the high-T_{c} cuprates.

2.2 Why Oxide Heterostructures?

Deviations from the standard band picture of lattice electrons are often encountered in bulk oxides. Therefore by heterostructuring these materials, a unique playground

for investigating, tailoring, and designing correlation effects is opened. The challenges are twofold in this respect. First, *known* bulk correlation phenomena are transferred into the heterostructure environment and may be tuned by various means. Second, by interfacing different electronic bulk phases, *new* interface phases are generated, eventually even without a distinct bulk analogon. A very strict separation between these both research directions is albeit delicate.

In the first challenge, common correlation features, e.g., strong spectral-weight transfer, formation of Hubbard satellites, enhanced susceptibilities, local-moment formation, Kondo physics, magnetic ordering, metal-insulator transition, and charge-density-wave or superconducting instabilities, are studied within heterostructure architectures. Proximity to an interface, different structural/geometrical relaxations/constraints, symmetry breakings due to layering (cf. Fig. 2), or polarization effects because of a heterostructure-adapted electric field are only a few impacts that are able to influence and modify the original bulk electron system. Already a lot of work has been performed in this context, and much is cited in the review articles given in the introduction.

For a concrete example, the physics of doped Mott insulators is a key research field in strongly correlated condensed matter. Bulk doping however poses many difficulties in view of a well-defined theoretical description. Most importantly, the intertwining with disorder mechanisms in the electronic and structural sector often hinders a straightforward modeling. Those problems may be overcome in Mott-oxide heterostructures, since electron, hole, or structural (i.e., effects only due to the different ion size of the valence-identical dopant) doping can nowadays be realized by respective doping layers, e.g., via molecular beam epitaxy (cf. Eckstein and Bozovic (1995) and Stemmer and Millis (2013) for reviews). Therewith, the correlated doping physics becomes accessible in a well-defined manner by model-Hamiltonian and/or first-principles techniques. In general, accompanied by this progress, the examination of the influence of defects in strongly correlated materials has gained renewed interest, being explored by considering the detailed defect chemistry together with a state-of-the-art treatment of electronic correlations.

The second challenge is even more demanding and might be associated with the physics of conducting quasi-two-dimensional electron interface systems emerging in oxide heterostructures consisting of bulk band insulators. In principle, selected surface systems may also be counted in this regard. Yet the clarification of the

Fig. 2 Principle issue of symmetry breaking induced by an interface in an oxide heterostructure, leading to layer-dependent electronic self-energies Σ_{ij}
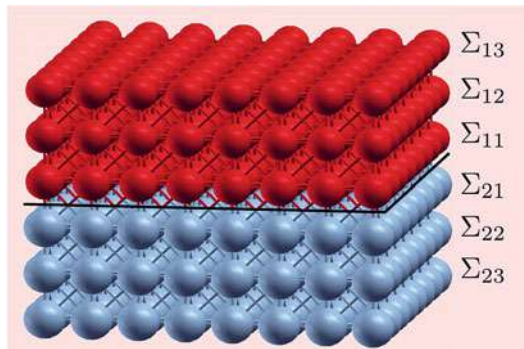
uniqueness of such engineered electron phases beyond respective bulk counterparts is still ongoing research. Topological nontrivial electronic states are surely to be considered, but there the explicit interplay between the topological aspect and electronic correlations is still largely uncharted territory. In the following, these latter material possibilities will not be covered, but the interested reader finds some ideas on this in a recent review (Keimer and Moore 2017).

3 Theoretical Approach: Realistic Many-Body Theory

3.1 Electronic Density Functional Theory (DFT)

Electronic density functional theory marks an everlasting milestone in condensed matter research, it is and will remain a key step in the first-principles investigation of matter. There are excellent reviews and books on DFT, so there is no need here to iterate thereon. In essence, the framework maps the problem of interacting electrons onto the problem of noninteracting particles within a complicated effective potential. In the present context, it is vital to note that although DFT represents *in principle* a complete many-body account of interacting electrons, the most common Kohn-Sham representation based on the conventional local-density or generalized-gradient approximations (i.e., LDA or GGA) marks this method as an effective single-particle approach. In condensed matter, it therefore describes band electrons, whereby the original many-body effects are cast into the named effective potential.

3.2 Dynamical Mean-Field Theory (DMFT)

When it comes to strongly correlated lattice electrons, the dynamical mean-field theory (Metzner and Vollhardt 1989; Georges and Kotliar 1992) has the reputation of being the many-body scheme with the best compromise between generality, accuracy, and performance. Also DMFT describes a mapping, and it is here from the problem of interacting lattice electrons onto the problem of a quantum impurity within a self-consistent energy-dependent bath, as sketched for the Hubbard model in Fig. 3.

The one-particle Green's function provides seminal access to the spectral properties and the total energy of an interacting electron system on a lattice. For chemical potential μ, and Hamiltonian $H(\mathbf{k})$ at wave vector $\mathbf{k}$, it reads:

$$G(\mathbf{k}, i\omega_n) = [i\omega_n + \mu - H(\mathbf{k}) - \Sigma(\mathbf{k}, i\omega_n)]^{-1} \quad . \tag{2}$$

Note that here and in the following part of the text, fermionic Matsubara frequencies $\omega_n := (2n + 1)\pi T$ are employed to emphasize the treatment at *finite* temperature. The analytical continuation to real frequencies ω in actual calculations may, e.g., be performed via the maximum entropy method (see, e.g., Imada et al. (1998) and Georges et al. (1996) for more general details).
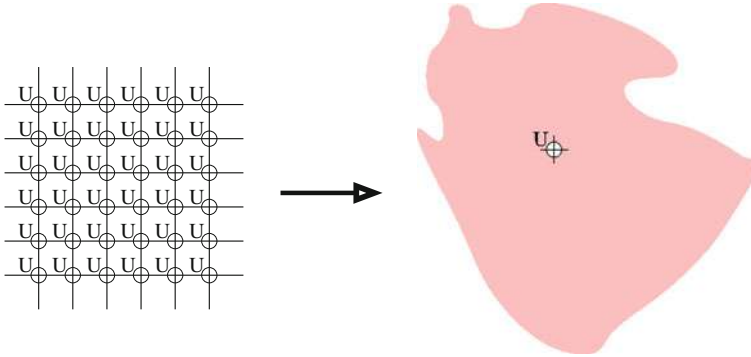
Fig. 3 Sketch of the DMFT mapping of an interacting lattice problem (left) onto a problem of an impurity within an energy-dependent bath (right), assuming a sole on-site Coulomb interaction U

In DMFT, the local Green's function is approximated with the help of a k-independent impurity self-energy $\Sigma_{\mathrm{imp}}(i\omega_{\mathrm{n}})$, i.e.:

$$G_{\mathrm{loc}}^{\mathrm{DMFT}}(i\omega_{\mathrm{n}}) = \sum_{\mathbf{k}}[i\omega_{\mathrm{n}} + \mu - H(\mathbf{k}) - \Sigma_{\mathrm{imp}}(i\omega_{\mathrm{n}})]^{-1}, \tag{3}$$

whereby the corresponding impurity problem reads:

$$\Sigma_{\mathrm{imp}}(i\omega_{\mathrm{n}}) = \mathscr{G}_0(i\omega_{\mathrm{n}})^{-1} - G_{\mathrm{imp}}(i\omega_{\mathrm{n}})^{-1} \quad . \tag{4}$$

The Weiss field $\mathscr{G}_0(i\omega_{\mathrm{n}})$ is a unique function of the local Hamiltonian (expressed within a localized basis), and, importantly, the DMFT self-consistency condition implies $G_{\mathrm{imp}} = G_{\mathrm{loc}}^{\mathrm{DMFT}}$. The calculational loop is depicted in the DMFT-loop box of Fig. 4, and quantum impurity solvers based, e.g., on quantum Monte Carlo, exact diagonalization, etc. yield the solution. For more details, we refer to Georges et al. (1996) for a review. Note that local-interaction diagrams are included to all orders in this non-perturbative theory. The vital energy dependence of the Weiss field ensures the qualitatively correct description of low-energy QP features as well as high-energy incoherent (Hubbard) excitations. Extensions to overcome the restriction to a k-independent self-energy, e.g., via cluster schemes, are available. But those will not be further pursued in the present text.

3.3   Combining DFT and DMFT

The first explicit promotions of DMFT to the realistic level through a merging with Kohn-Sham DFT has been realized end of the 1990s (Anisimov et al. 1997; Lichtenstein and Katsnelson 1998). Importantly, since the many-body part incorporates a Hubbard-model(-like) picturing of a suitably chosen *correlated*
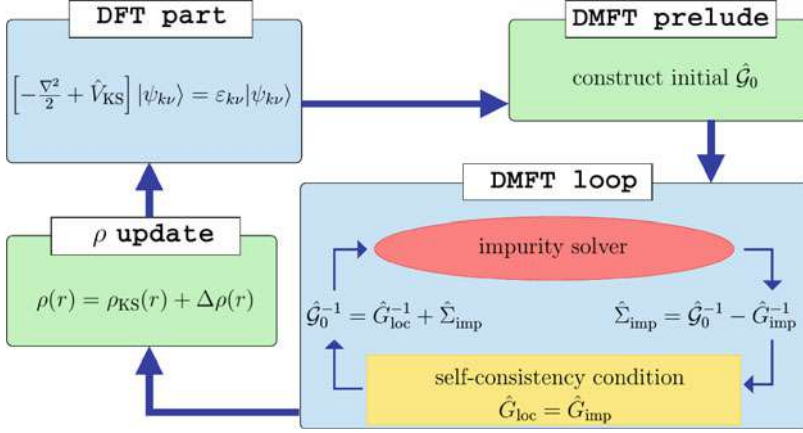
Fig. 4 State-of-the-art charge self-consistent DFT+DMFT loop. (After Lechermann et al. 2006). The calculation usually starts from a self-consistent Kohn-Sham solution. The correlated subspace is defined and the initial Weiss field $\mathscr{G}_0$ constructed. Afterward, a single (or more) DMFT step is performed. The obtained self-energies are upfolded and an updated charge density $\rho(\mathbf{r})$ is computed. A new charge density implies a new Kohn-Sham potential, and a single new Kohn-Sham step is performed; therefrom a new Weiss field is generated, etc

subspace, a (partly) local-orbital representation is an essential building block of the DFT+DMFT framework. Linear-muffin-tin orbitals (Andersen 1975); Wannier(like) functions, e.g., of maximally localized kind (Marzari et al. 2012); or projected-local orbitals (Amadon et al. 2008; Anisimov et al. 2005) are usually in charge of that representation. The correlated subspace is understood as a quantum-numbered real-space region where correlated electrons hide. The key interfacing blocks of the complete DFT+DMFT self-consistency cycle (Savrasov et al. 2001; Minár et al. 2005; Pourovskii et al. 2007; Grieger et al. 2012) (cf. Fig. 4) are marked by the downfolding of the full-problem Bloch space to the correlated subspace and the upfolding of the DMFT self-energy back to the original space. The main respective formulas for sites $\mathbf{R}$, local orbitals mm' and band indices $\nu\nu'$, read:

$$G_{mm'}^{\mathbf{R},\mathrm{imp}}(i\omega_n) = \sum_{\mathbf{k},(\nu\nu')\in\mathscr{W}} \bar{P}_{m\nu}^{\mathbf{R}}(\mathbf{k})\, G_{\nu\nu'}^{\mathrm{bloch}}(\mathbf{k}, i\omega_n)\, \bar{P}_{\nu'm'}^{\mathbf{R}*}(\mathbf{k}), \qquad (5)$$

$$\Delta\Sigma_{\nu\nu'}^{\mathbf{R},\mathrm{bloch}}(\mathbf{k}, i\omega_n) = \sum_{\mathbf{R},mm'} \bar{P}_{\nu m}^{\mathbf{R}*}(\mathbf{k})\, \Delta\Sigma_{mm'}^{\mathrm{imp}}(i\omega_n)\, \bar{P}_{m'\nu'}^{\mathbf{R}}(\mathbf{k}), \qquad (6)$$

with $\bar{P}$ denoting the normalized projection between Bloch space and correlated subspace (Amadon et al. 2008). The object $\Delta\Sigma_{\nu\nu'}^{\mathrm{bloch}}$ describes the k-dependent self-energy in Bloch space after double-counting correction. As for the correlated subspace, there is a choice for the range $\mathscr{W}$ of included Kohn-Sham bands in the downfolding. The double-counting correction takes care of the fact that some

correlations are already handled on the DFT level. In the upfolding operation, the charge density will also be updated with correlation effects, i.e.:

$$\rho(\mathbf{r}) = \sum_{\mathbf{k}, \nu\nu'} \langle \mathbf{r}|\Psi_{\mathbf{k}\nu}\rangle \Big(f(\tilde{\epsilon}_{\mathbf{k}\nu})\delta_{\nu\nu'} + \Delta N_{\nu\nu'}(\mathbf{k}) \Big) \langle \Psi_{\mathbf{k}\nu'}|\mathbf{r}\rangle, \tag{7}$$

where Ψ denotes Kohn-Sham states, f is the associated Fermi function, and ΔN is the DMFT self-energy correction term (Lechermann et al. 2006; Amadon et al. 2008). Thus, since a pure band picture is not vital in a many-body system and *real-space* excitations also matter, additional off-diagonal terms in the band index contribute in the correlated regime. This novel charge density accordingly then defines a new Kohn-Sham effective potential. Note finally that this first-principles many-body scheme works, at heart, at finite temperature T. Electron states are therefore subject to the full thermal impact, beyond sole occupational Fermi-function modification. For more formal and detailed accounts on the DFT+DMFT scheme, we refer to Georges (2004) and Kotliar et al. (2006).

In oxide heterostructures, as in various other multiatom unit cells, the correlated subspace invokes not only a single lattice site. For symmetry-equivalent sites, the self-energy is determined for a representative site and transferred to the remaining sites via the proper symmetry relations. A different impurity problem is defined for each symmetry-inequivalent site j through (Potthoff and Nolting 1999):

$$\mathscr{G}_0^{(j)}(i\omega_\mathrm{n})^{-1} = G^{(j)}(i\omega_\mathrm{n})^{-1} + \Sigma_\mathrm{imp}^{(j)}(i\omega_\mathrm{n}), \tag{8}$$

and the coupling is realized via the DFT+DMFT self-consistency condition invoking the computation of the complete lattice Green's function.

4 Selected Studies of the Correlated Electronic Structure of Oxide Heterostructures

There are already various applications of the DFT+DMFT approach to the problem of oxide heterostructures, and the number is expected to further grow substantially. Therefore in a short review, the choice of examples has to be highly selective and is here mainly driven by the author's interest in this field of research. We apologize for not covering many details of other interesting studies. Furthermore, albeit DFT and/or DFT+U studies of oxide heterostructures also provide relevant insight, including a discussion of such works would go beyond the limited scope of the present text.

If not otherwise stated, the materials investigations discussed in Sects. 4.1 and 4.2 involving the author were performed using charge-self-consistent DFT+DMFT based on a mixed-basis pseudopotential code (Meyer et al. 1998) and hybridization-expansion continuous-time quantum Monte Carlo (Werner et al. 2006) in the TRIQS

package (Seth et al. 2016) as an impurity solver. For more technical details on the implementation, the reader is referred to Grieger et al. (2012).

4.1   Mott-Band Insulator Architectures

Joining different bulk electronic phases across the interface of a heterostructure is appealing as a plausible route to emerging physics. Thermodynamics may select novel electronic states different from the original bulk states that cope with the intriguing phase competition.

In this respect, oxide heterostructures composed of a Mott insulator and a band insulator belong to early studied systems, initially via experimental work on the LaTiO$_3$/SrTiO$_3$ (LTO/STO) interface (Ohtomo et al. 2002). While STO is an ideal cubic perovskite at ambient temperature, LTO marks a distorted perovskite with orthorhombic crystal symmetry. First model-Hamiltonian Hartree-Fock studies of a Hubbard-model application to such an interface yielded an intricate phase diagram in the number of La layers and Hubbard U (Okamoto and Millis 2004). Simplified realistic DMFT for LTO/STO by Ishida and Liebsch (2008) emphasized the structural orthorhombic-versus-tetragonal aspect of the LTO part. Transport within a Mott-band insulator heterostructure has been studied again within a model-Hamiltonian approach by Rüegg et al. (2007).

A superlattice DFT+DMFT investigation of LTO/STO (Lechermann et al. 2013) revealed the realistic competition of both insulating systems stacked along the c-axis, giving rise to a metallic interface state. The correlated subspace can be chosen to be spanned by Ti-3$d(t_{2g})$ states, i.e., consists locally of three correlated orbitals. A Hubbard $U = 5$ eV and Hund's exchange $J_\mathrm{H} = 0.7$ eV are applied in that subspace. Orbital-dependent charge transfers lead to a strong Ti-3$d(xy)$, i.e., inplane electronic, polarization in the interface TiO$_2$ layer. Surely, on the STO side, the Ti^{4+} oxidation state with nominal 3d^0 occupation is quickly reached, while on the LTO side, the Ti^{3+}-3d^1 establishes and a sizable lower Hubbard band at ~-1.1 eV is identified in the spectral function (see also Fig. 5). In those calculations, the lattice constant was fixed to the cubic STO value, but local structural relaxations on the DFT level were allowed. It is to be noted that the lattice degrees of freedom are an important aspect in oxide heterostructures. Simplified DFT+DMFT bulk-like studies revealed, e.g., the impact of strain on the Mott-insulating state of LaTiO$_3$ (Dymkowski and Ederer 2014) and LaVO$_3$ (Sclauzero and Ederer 2015).

Even more intriguing physics may be found when starting from a doped-Mott state within the heterostructure setting. Motivated by experimental work (Moetakef et al. 2012; Jackson et al. 2014; Mikheev et al. 2015), a first-principles many-body investigation of δ-doping the rare-earth titanates LaTiO$_3$, GdTiO$_3$, and SmTiO$_3$ with a single SrO layer was undertaken (Lechermann and Obermeyer 2015; Lechermann 2017b). Especially the δ-doped SmTiO$_3$ case displays puzzling physics in experiment, namely, NFL transport that switches to FL-like characteristics upon adding further SrO layers. In the rare-earth titanate 3$d(t_{2g}^1)$ series of distorted-perovskite Mott insulators, the magnetic low-temperature state changes from antiferromagnetic (AFM) to ferromagnetic (FM) with the size of the rare-earth ion. The samarium
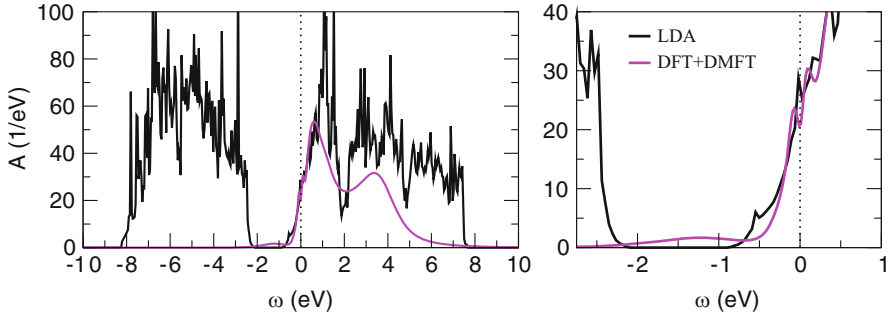
Fig. 5 Spectral information for a LaTiO$_3$/SrTiO$_3$ superlattice with 4 SrO and 4 LaO layers. (After Lechermann et al. 2013). Left, total spectrum; right, closer to Fermi level. Note that the DMFT correlated subspace includes only Ti(t_{2g}) functions; therefore, the DFT+DMFT spectra are here chosen to cover also only these contributions. However of course in the charge self-consistent cycle, all states are included

titanate is still AFM, but just on the border toward ferromagnetism. Structurally well-defined hole doping introduced by the SrO monolayer renders SmTiO$_3$ metallic. The DFT+DMFT results (see Fig. 6) reveal significant spectral-weight transfer to higher energies compared to DFT and in addition a complex layer-resolved electronic structure. While far from the doping layer, the system resides in an orbital-polarized Mott-insulating state; it is conducting in an orbital-balanced manner just at the interface TiO$_2$ layer (cf. Fig. 6c). Both regimes are joined by an orbital-polarized doped-Mott layer with a largely renormalized QP peak at the Fermi level. Detailed analysis of the layer- and orbital-dependent self-energies shown in Fig. 6d, e indeed reveals signatures of a NFL exponent for the dominant effective t_{2g} orbital in the second, i.e., orbital-polarized doped-Mott, TiO$_2$ layer. Further investigations hint toward competing AFM-FM fluctuations in the Mott-critical zone as a possible cause for NFL behavior (Lechermann 2017b). A pseudogap(-like) structure in the theoretical spectral function subject to such fluctuations has indeed been identified in experimental studies (Marshall et al. 2016). In an extension of this study, it was shown that the addition of further SrO layers establishes an extra band-insulating regime in the formed SrTiO$_3$-like region, with a stronger inplane xy-polarized metallic layer at the boundary (Lechermann 2017a).

4.2   Band-Band Insulator Architectures

The formation of a metallic two-dimensional electron system (2DES) at the n-type interface between the band insulators LaAlO$_3$ (LAO) and SrTiO$_3$ (Ohtomo and Hwang 2004) has so far been the most appreciated finding in the oxide-heterostructure context. Follow-up experimental studies furthermore revealed the 2DES delicacy, e.g., the possibility for superconductivity (Reyren et al. 2007) as well as magnetic order (Brinkman et al. 2007) in the LAO/STO interface.
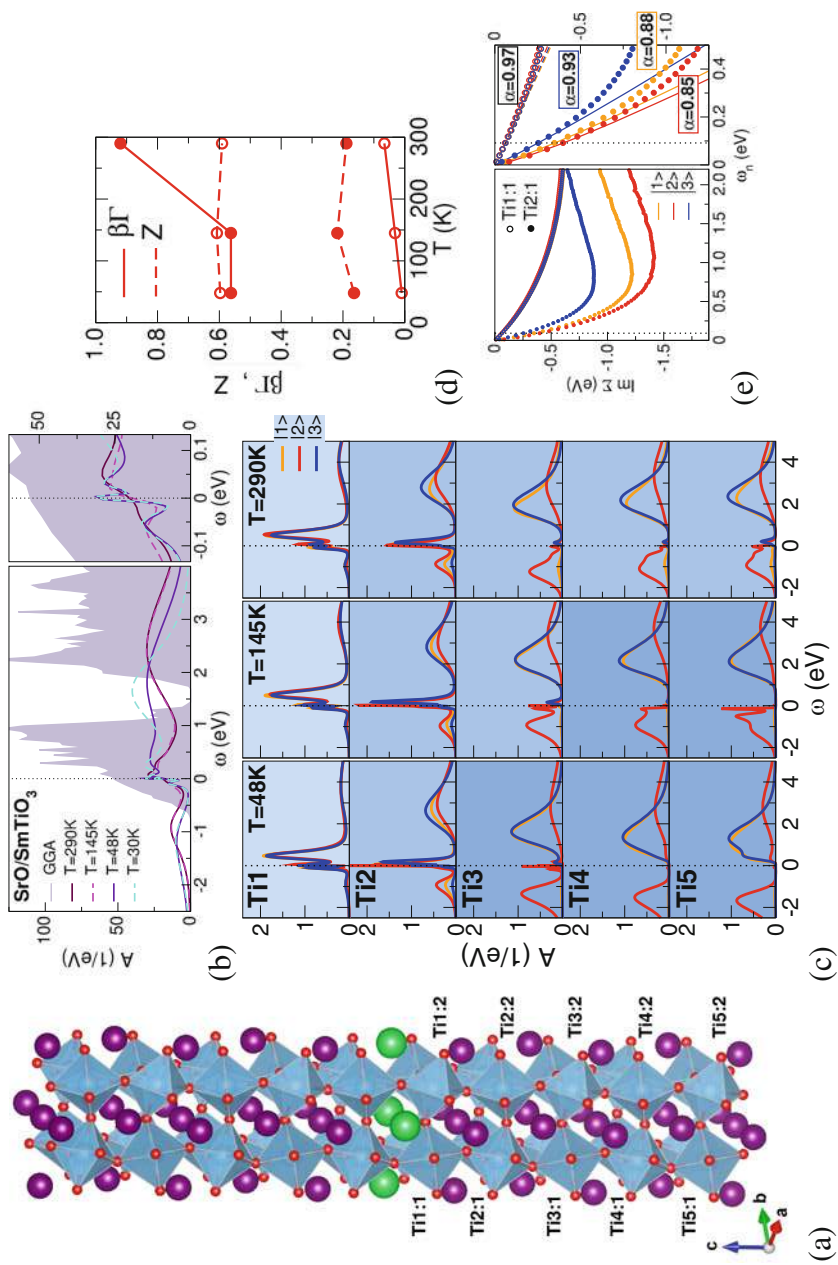
Fig. 6 (continued)

A polar-catastrophe mechanism (Nakagawa et al. 2006) is believed to be dominantly at the root of the 2DES buildup.

In view of electronic correlations, this and related interface systems (Chen et al. 2013) appear more subtle since the band-insulating constituents do not already host electrons in partially filled d- or f-states. Nonetheless, the confirmed stabilization of magnetic and superconducting order proves the existence of collective electronic behavior. Then, nonlocal effects might be more crucial than in the former Mott-band heterostructures. However, still relevant local correlations are evident from the coappearance of oxygen vacancies and ferromagnetism in LAO/STO (Salluzzo et al. 2013). Point defects are an important ingredient of the present interface physics, and oxygen vacancies (OVs) so far appear to have major impact (Pavlenko et al. 2012). In a DFT+DMFT study of the LAO/STO interface (Lechermann et al. 2014), it is shown that stable ferromagnetic order needs both, OVs *and* electronic correlations. While Ti-3$d(t_{2g})$ orbitals dominate the states directly above the STO gap, an oxygen vacancy leads to an in-gap state of Ti-3$d(e_g)$ kind, here termed $\tilde{e}_g$. In a minimal model, the correlated electronic structure at the interface may be described by the interplay between $\tilde{e}_g$ and an inplane xy orbital from the t_{2g} threefold (see Fig. 7). Therefore, a reduced Hubbard $U = 2.5\,\mathrm{eV}$ and Hund's exchange $J_H = 0.5\,\mathrm{eV}$ are appropriate in the smaller correlated subspace. Note that a dense-defect scenario is assumed in that superlattice assessment, i.e., there is an OV at every other O site in the interface TiO$_2$ plane. Still an in-gap $\tilde{e}_g$-like state at $\varepsilon_{IG} \sim -1.2\,\mathrm{eV}$ is well reproduced in agreement with experiment (Berner et al. 2013). Close to an oxygen vacancy, the formal oxidation state of titanium is close to Ti^{3+}, i.e., again locally a $3d^1$ occupation is realized with defects. But the emergent spin polarization in the given defect limit is not of purely local kind; it develops substantial dispersive behavior (cf. Fig. 7d, e). In general, such studies show that strong electron correlations, describable within DMFT, may be introduced also in band-band insulator architectures of oxide heterostructures.

Approaching significantly lower OV concentrations in a first-principles manner asks for the handling of much larger supercells. Due to the numerical heavyness of DFT+DMFT, this is yet not easily possible. Instead, a model-Hamiltonian approach, equipped with the relevant ingredients from the dense-defect limit, appears more adequate. Figure 8 displays the setting and some main results of such

---

Fig. 6 Paramagnetic DFT+DMFT data of δ-doping SmTiO$_3$ with a SrO monolayer. (After Lechermann 2017b). (**a**) 100-atom unit cell in a superlattice architecture: Sr (green), Sm (purple), Ti (blue), O (red). Two inequivalent Ti sites are handled in each TiO$_2$ layer. Structural relaxations are performed on the DFT(GGA) level. (**b**) Total spectral function at different temperatures. (**c**) Layer-, orbital-, and T-resolved spectral function. The correlated subspace consists of three effective t_{2g} orbitals at each Ti site. (**d**) QP weight Z and dimensionless electron-electron scattering rate $\beta\Gamma = -Z\,\mathrm{Im}\,\Sigma(i0^+)$ for the dominant state $|2\rangle$ in the first (open circles) and second (filled circles) TiO$_2$ layer. (**e**) Orbital-resolved imaginary part of the self-energy on the Matsubara axis. Left, larger frequency range; right, low-frequency region with fitting functions $\mathrm{Im}\,\Sigma(\omega_n) = C_0 + A\,\omega_n^\alpha$ (dashed/full lines). Exponential-fitting cutoff n_c is denoted by the dotted line. Values $\alpha = 1$ and $C_0 = 0$ mark the Fermi-liquid regime
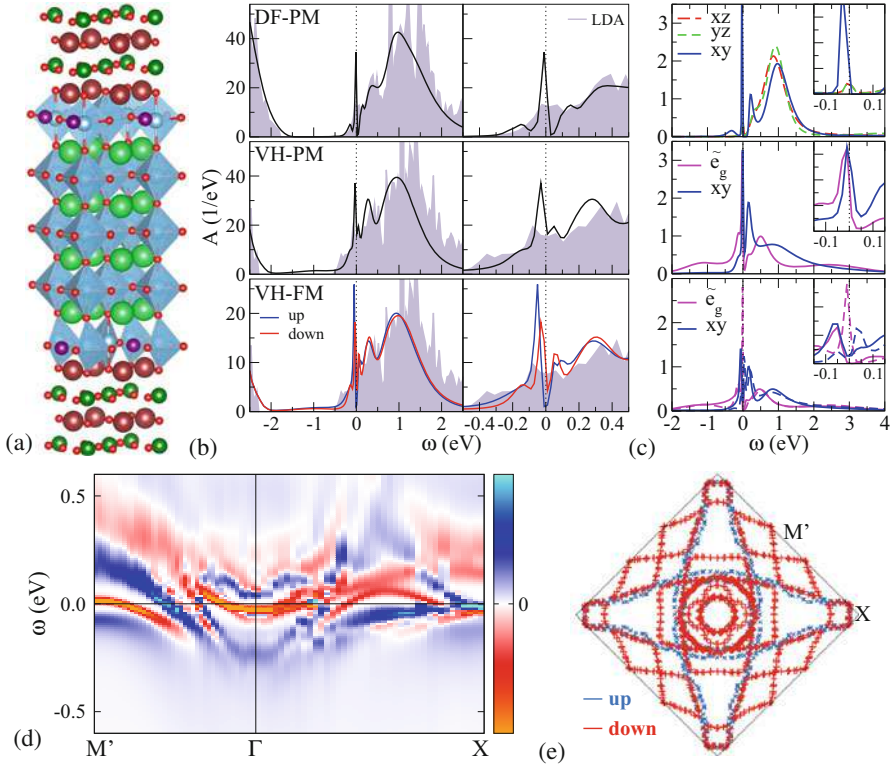
Fig. 7 Influence of oxygen vacancies in the LAO/STO interface based on DFT+DMFT calculations for a dense-defect scenario. (After Lechermann et al. 2014). (**a**) 80-atom superlattice, La (brown), Al (darkgreen), Sr (green), Ti (blue), O (red), OV (violet). (**b**) Total spectral function: left, larger window; right, smaller window. Top, defect-free paramagnetic (DF-PM); middle, vacancy-hosting paramagnetic (VH-PM); and bottom, vacancy-hosting ferromagnetic (VH-FM). (**c**) Local spectral function for selected effective Ti(3d) states (see text), vertical ordering as in (**b**). (**d**) Spin contrast of the k-dependent spectral function $A(\mathbf{k}, \omega)$ in the ferromagnetic phase. (**e**) Ferromagnetic Fermi surface

an ansatz (Behrmann and Lechermann 2015). A two-orbital $\tilde{e}_g$-xy Hubbard model is solved on a 10×10 square lattice resembling the TiO_2 interface plane. The efficient rotational-invariant slave boson (RISB) scheme (Lechermann et al. 2007; Li et al. 1989), employing a self-energy which has a linear-in-frequency term and a static term, is put into practice for a simplified treatment. Focusing on the magnetic order, it is shown that there are three regimes with growing number of OVs. At very small concentration, a Ruderman-Kittel-Kasuya-Yoshida (RKKY) coupling leads to FM order, whereas at larger concentration a double-exchange mechanism dominates a different FM phase. In-between local AFM pairs (or, in an advanced self-energy modeling, possibly singlets) result in a nearly absent net magnetic moment. This intricate OV-dependent magnetic exchange is in line with experimental findings of strongly probe-dependent magnetic response.
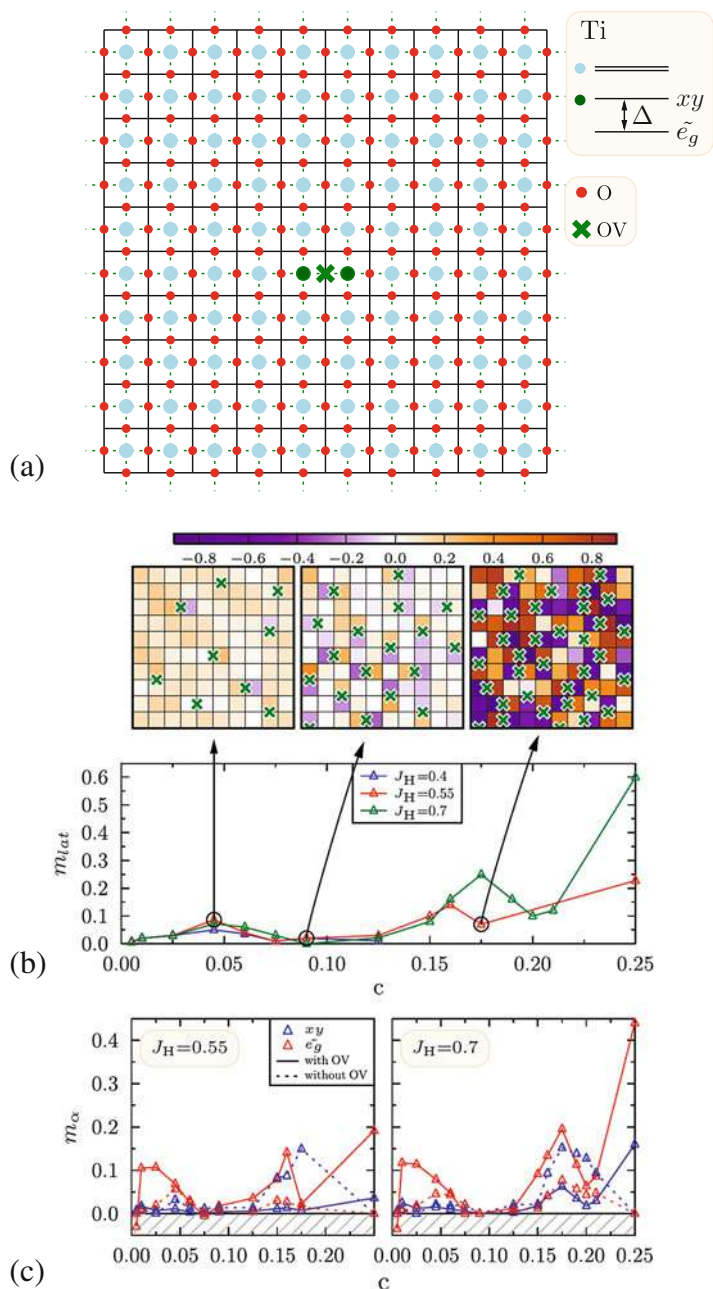
Fig. 8 Influence of oxygen vacancies in the LAO/STO interface. (After Behrmann and Lechermann (2015)). (**a**) Model setting: 10×10 square lattice, two orbitals per site, oxygen degrees of freedom that are integrated out, nearest-neighbor hopping $t = 0.2$ eV. A crystal-field splitting $\Delta = 0.3$ eV between $\tilde{e}_g$ and xy is applied, if an OV is located nearby. (**b**) Top: color-coded ordered local magnetic moments on the lattice for selected OV concentration. Bottom: lattice net moment per site for different Hund's exchange J_H. (**c**) Orbital-resolved magnetic moment, averaged per site

In parallel to the STO-based heterostructure investigations, studying the $SrTiO_3$ surface attracted significant attention. Interface and surface properties of a chosen oxide are often related and the comparison between both plane defects enables a better understanding of emergent phenomena. A 2DES was initially also found on the oxygen-deficient (001) surface (Santander-Syro et al. 2011; Meevasana et al. 2011) and soon after also confirmed for other cleavage planes, e.g., in (111) direction (McKeown Walker et al. 2014). As a difference to the LAO/STO interface, the defect-free STO surface is believed to be insulating. Due to a missing interface-driven polar-catastrophe mechanism, defects such as OVs are essential to metallize the surface. Similar to the interface spectrum, an in-gap state at a very similar position, i.e., $\sim -1.3\,eV$, has been detected on the STO surface early on Tanaka et al. (1993). Recent DFT+DMFT considerations of the STO(001) surface with OVs indeed verified this in-gap state (Lechermann et al. 2016), which is again dominantly formed by Ti-$3d(e_g)$ weight (cf. Fig. 9). Furthermore, the low-energy structure dominated by Ti-$3d(t_{2g})$ states is also in accordance with experiment. A double-vacancy defect provided the best matching with experiment; however only two distinct vacancy configurations were examined because of the large numerical effort. Nonetheless, the spectral separation of e_g at high energy and t_{2g} is a clear generic feature of the study. Conventional DFT is not sufficient to provide this orbital separation (cf. Fig. 9c, d).

4.3 Further Investigations

So far, the present review mainly focused on early-transition-metal titanate heterostructures, but theoretical work also dealt with other oxide designs. For instance, the late-transition-metal rare-earth (R) nickelate series $RNiO_3$ provides important material building blocks, too (Hwang et al. 2013). Originally, there was the idea to realize cuprate-like physics within nickelates by proper heterostructuring (Hansmann et al. 2009; Han et al. 2011); however experimental success remains absent. Bulk perovskite-like vanadates split into correlated metals, e.g., $SrVO_3$ and $CaVO_3$, and Mott insulators, e.g., $LaVO_3$ and YVO_3. Heterostructures based on $SrVO_3$ were studied in view of a possible loss of metallicity in a small-layer limit (Okamoto 2011; Zhong et al. 2015). Moreover, transition-metal oxides from the $4d$ and $5d$ series serve as further building blocks. For instance, the $4d^4$ physics of strontium and/or calcium ruthenates poses a long-standing problem, which can be tuned by heterostructuring (Si et al. 2017).

5 Outlook

The research field of oxide heterostructures will remain promising, on a basic research level, as well as in view of possible technical applications. Until now, only a few materials classes have been utilized, and there are vast ways of combinations, both in different materials as in different geometries. In addition, the opening
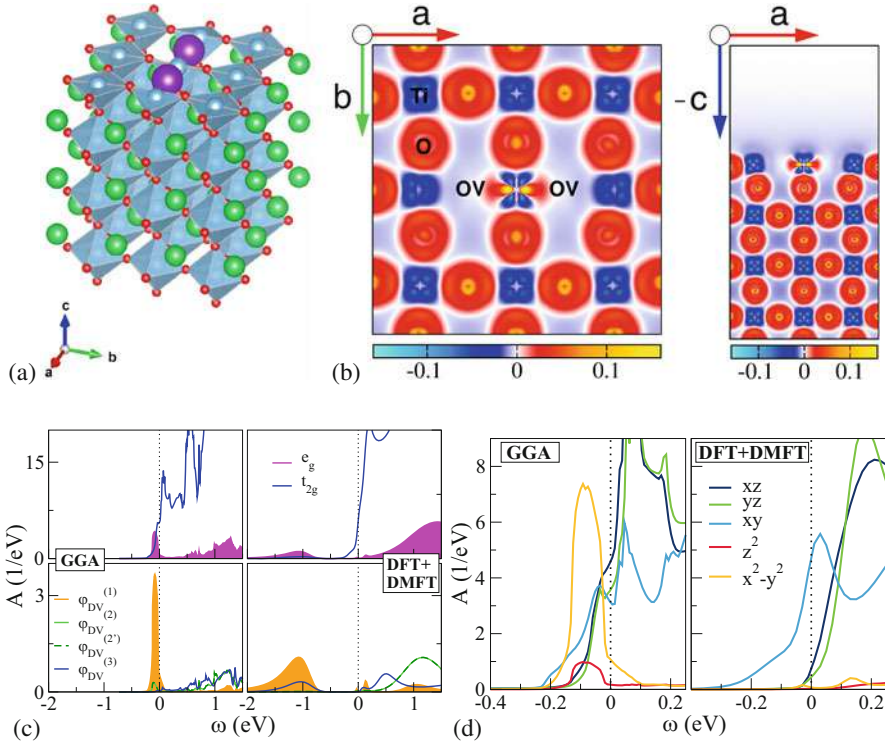
Fig. 9 Influence of a double vacancy in the SrTiO$_3$ (001) surface. (After Lechermann et al. 2016). (**a**) 180-atom supercell with a TiO$_2$-terminated surface layer of STO: Sr (green), Ti (blue), O (red), OV (violet). (**b**) Bond charge density $n_{\text{total}}(\mathbf{r}) - n_{\text{atomic}}(\mathbf{r})$: top view (left), side view (right). (**c**) Spectral-weight comparison of the summed Ti(t_{2g}, e_g) (top) and of the dominant Ti-based effective orbital between DFT(GGA) (left) and DFT+DMFT (right). The $\varphi_{\text{DV}}^{(1)}$ Wannier-like orbital is an effective inplane e_g orbital located on the Ti site between the OVs. (**d**) Low-energy spectral-weight comparison of the Ti($3d$) states

toward more general hybrid heterostructures, e.g., with Dirac materials, has just started and will lead to novel phenomena. Detailed studies of challenging spin, charge, or pairing instabilities of oxide heterostructures are still at its infancy, and expectations are high for new surprising findings. Furthermore, the recent focus on the nonequilibrium regime in strongly correlated matter will find an ideal playground in this class of designed materials.

The DFT+DMFT method, and future extensions to it, has the chance to pave the roads toward a new era in materials science, highlighting the many-body character in key functionalities. A better understanding and general appreciation of the intriguing interplay between electron correlation and materials chemistry is believed to be at the heart of this endeavor.

References

Amadon B, Lechermann F, Georges A, Jollet F, Wehling TO, Lichtenstein AI (2008) Plane-wave based electronic structure calculations for correlated materials using dynamical mean-field theory and projected local orbitals. Phys Rev B 77:205112

Andersen OK (1975) Linear methods in band theory. Phys Rev B 12:3060

Anisimov VI, Poteryaev AI, Korotin MA, Anokhin AO, Kotliar G (1997) First-principles calculations of the electronic structure and spectra of strongly correlated systems: dynamical mean-field theory. J Phys Condens Matter 9:7359

Anisimov VI, Kondakov DE, Kozhevnikov AV, Nekrasov IA, Pchelkina ZV, Allen JW, Mo SK, Kim HD, Metcalf P, Suga S, Sekiyama A, Keller G, Leonov I, Ren X, Vollhardt D (2005) Full orbital calculation scheme for materials with strongly correlated electrons. Phys Rev B 71:125119

Bednorz JG, Müller KA (1986) Possible hightc superconductivity in the Ba-La-Cu-O system. Z Physik B Condens Matter 64:189

Behrmann M, Lechermann F (2015) Interface exchange processes in $LaAlO_3/SrTiO_3$ induced by oxygen vacancies. Phys Rev B 92:125148

Berner G, Sing M, Fujiwara H, Yasui A, Saitoh Y, Yamasaki A, Nishitani Y, Sekiyama A, Pavlenko N, Kopp T, Richter C, Mannhart J, Suga S, Claessen R (2013) Direct k-space mapping of the electronic structure in an oxide-oxide interface. Phys Rev Lett 110:247601

Brinkman A, Huijben M, van Zalk M, Huijben J, Zeitler U, Maan JC, van der Wiel WG, Rijnders G, Blank DHA, Hilgenkamp H (2007) Magnetic effects at the interface between non-magnetic oxides. Nat Mater 6:493

Chakhalian J, Freeland JW, Millis AJ, Panagopoulos C, Rondinelli JM (2014) Emergent properties in plane view: strong correlations at oxide interfaces. Rev Mod Phys 86:1189

Chen YZ, Bovet N, Trier F, Christensen DV, Qu FM, Andersen NH, Kasama T, Zhang W, Giraud R, Dufouleur J, Jespersen TS, Sun JR, Smith A, Nygård J, Lu L, Büchner B, Shen BG, Linderoth S, Pryds N (2013) A high-mobility two-dimensional electron gas at the spinel/perovskite interface of γ-$Al_2O_3/SrTiO_3$. Nat Commun 4:1372

Dymkowski K, Ederer C (2014) Strain-induced insulator-to-metal transition in $LaTiO_3$ within DFT+DMFT. Phys Rev B 89:161109(R)

Eckstein JN, Bozovic I (1995) High-temperature superconducting multilayers and heterostructures grown by atomic layer-by-layer molecular beam epitaxy. Ann Rev Mater Sci 25:679

Georges A (2004) Lectures on the physics of highly correlated electron systems VIII. AIP conference proceedings, vol 715, Chap 3. American Institute of Physics, Melville

Georges A, Kotliar G (1992) Hubbard model in infinite dimensions. Phys Rev B 45:6479

Georges A, Kotliar G, Krauth W, Rozenberg MJ (1996) Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. Rev Mod Phys 68:13

Grieger D, Piefke C, Peil OE, Lechermann F (2012) Approaching finite-temperature phase diagrams of strongly correlated materials: a case study for V_2O_3. Phys Rev B 86:155121

Han MJ, Wang X, Marianetti CA, Millis AJ (2011) Dynamical mean-field theory of nickelate superlattices. Phys Rev Lett 107:206804

Hansmann P, Yang X, Toschi A, Khaliullin G, Andersen OK, Held K (2009) Turning a nickelate Fermi surface into a cupratelike one through heterostructuring. Phys Rev Lett 103:016401

Hwang HY, Iwasa Y, Kawasaki M, Keimer B, Nagaosa N, Tokura Y (2012) Emergent phenomena at oxide interfaces. Nat Mater 11:103

Hwang J, Son J, Zhang JY, Janotti A, Van de Walle CG, Stemmer S (2013) Structural origins of the properties of rare earth nickelate superlattices. Phys Rev B 87:060101(R)

Imada M, Fujimori A, Tokura Y (1998) Metal-insulator transitions. Rev Mod Phys 70:1039

Ishida H, Liebsch A (2008) Origin of metallicity of $LaTiO_3/SrTiO_3$ heterostructures. Phys Rev B 77:115350

Jackson CA, Zhang JY, Freeze CR, Stemmer S (2014) Quantum critical behaviour in confined $SrTiO_3$ quantum wells embedded in antiferromagnetic $SmTiO_3$. Nat Commun 5:4258

Keimer B, Moore JE (2017) The physics of quantum materials. Nat Phys 13:1045

Kotliar G, Savrasov SY, Haule K, Oudovenko VS, Parcollet O, Marianetti CA (2006) Electronic structure calculations with dynamical mean-field theory. Rev Mod Phys 78:865

Lechermann F (2017a) First-principles many-body investigation of correlated oxide heterostructures: few-layer-doped $SmTiO_3$. arXiv:170908875

Lechermann F (2017b) Unconventional electron states in δ-doped $SmTiO_3$. Sci Rep 7:1565

Lechermann F, Obermeyer M (2015) Towards Mott design by δ-doping of strongly correlated titanates. New J Phys 17:043026

Lechermann F, Georges A, Poteryaev A, Biermann S, Posternak M, Yamasaki A, Andersen OK (2006) Dynamical mean-field theory using Wannier functions: a flexible route to electronic structure calculations of strongly correlated materials. Phys Rev B 74:125120

Lechermann F, Georges A, Kotliar G, Parcollet O (2007) Rotationally invariant slave-boson formalism and momentum dependence of the quasiparticle weight. Phys Rev B 76:155102

Lechermann F, Boehnke L, Grieger D (2013) Formation of orbital-selective electron states in $LaTiO_3/SrTiO_3$ superlattices. Phys Rev B 87:241101(R)

Lechermann F, Boehnke L, Grieger D, Piefke C (2014) Electron correlation and magnetism at the $LaAlO_3/SrTiO_3$ interface: a DFT+DMFT investigation. Phys Rev B 90:085125

Lechermann F, Jeschke HO, Kim AJ, Backes S, Valentí R (2016) Electron dichotomy on the $SrTiO_3$ defect surface augmented by many-body effects. Phys Rev B 93:121103(R)

Li T, Wölfle P, Hirschfeld PJ (1989) Spin-rotation-invariant slave-boson approach to the Hubbard model. Phys Rev B 40:6817

Lichtenstein AI, Katsnelson M (1998) Ab initio calculations of quasiparticle band structure in correlated systems: LDA++ approach. Phys Rev B 57:6884

Marshall PB, Mikheev E, Raghavan S, Stemmer S (2016) Pseudogaps and emergence of coherence in two-dimensional electron liquids in $SrTiO_3$. Phys Rev Lett 117:046402

Marzari N, Mostofi AA, Yates JR, Souza I, Vanderbilt D (2012) Maximally localized Wannier functions: theory and applications. Rev Mod Phys 84:1419

McKeown Walker S, de la Torre A, Bruno FY, Tamai A, Kim TK, Hoesch M, Shi M, Bahramy MS, King PC, Baumberger F (2014) Control of a two-dimensional electron gas on $SrTiO_3(111)$ by atomic oxygen. Phys Rev Lett 113:177601

Meevasana W, King PDC, He RH, Mo SK, Hashimoto M, Tamai A, Songsiriritthigul P, Baumberger F, Shen ZX (2011) Creation and control of a two-dimensional electron liquid at the bare $SrTiO_3$ surface. Nat Mat 10:114

Metzner W, Vollhardt D (1989) Correlated lattice fermions in $d = \infty$ dimensions. Phys Rev Lett 62:324

Meyer B, Elsässer C, Lechermann F, Fähnle M (1998) FORTRAN 90 program for mixed-basis-pseudopotential calculations for crystals

Mikheev E, Freeze CR, Isaac BJ, Cain TA, Stemmer S (2015) Separation of transport lifetimes in $SrTiO_3$-based two-dimensional electron liquids. Phys Rev B 91:165125

Minár J, Chioncel L, Perlov A, Ebert H, Katsnelson MI, Lichtenstein AI (2005) Multiple-scattering formalism for correlated systems: a KKR-DMFT approach. Phys Rev B 72:045125

Moetakef P, Jackson CA, Hwang J, Balents L, Allen SJ, Stemmer S (2012) Toward an artificial Mott insulator: correlations in confined high-density electron liquids in $SrTiO_3$. Phys Rev B 86:201102(R)

Mott NF, Peierls R (1937) Discussion of the paper by de Boer and Verwey. Proc Phys Soc 49:72

Nakagawa N, Hwang HY, Muller DA (2006) Why some interfaces cannot be sharp. Nat Mater 5:204

Ohtomo A, Hwang HY (2004) A high-mobility electron gas at the $LaAlO_3/SrTiO_3$ heterointerface. Nature 427:423

Ohtomo A, Muller DA, Grazul JL, Hwang HY (2002) Artificial charge-modulation in atomic-scale perovskite titanate superlattices. Nature 419:378

Okamoto S (2011) Anomalous mass enhancement in strongly correlated quantum wells. Phys Rev B 84:201305(R)

Okamoto S, Millis AJ (2004) Electronic reconstruction at an interface between a Mott insulator and a band insulator. Nature 428:630

Pavlenko N, Kopp T, Tsymbal EY, Mannhart J, Sawatzky GA (2012) Oxygen vacancies at titanate interfaces: two-dimensional magnetism and orbital reconstruction. Phys Rev B 86:064431

Potthoff M, Nolting W (1999) Surface metal-insulator transition in the Hubbard model. Phys Rev B 59:2549

Pourovskii LV, Amadon B, Biermann S, Georges A (2007) Self-consistency over the charge density in dynamical mean-field theory: a linear muffin-tin implementation and some physical implications. Phys Rev B 76:235101

Reyren N, Thiel S, Caviglia AD, Kourkoutis LF, Hammerl G, Richter C, Schneider CW, Kopp T, Rüetschi AS, Jaccard D, Gabay M, Muller DA, Triscone JM, Mannhart J (2007) Superconducting interfaces between insulating oxides. Science 317:1196

Rüegg A, Pilgram S, Sigrist M (2007) Aspects of metallic low-temperature transport in Mott-insulator/band-insulator superlattices: optical conductivity and thermoelectricity. Phys Rev B 75:195117

Salluzzo M, Gariglio S, Stornaiuolo D, Sessi V, Rusponi S, Piamonteze C, DeLuca GM, Minola M, Marré D, Gadaleta A, Brune H, Nolting F, Brookes NB, Ghiringhelli G (2013) Origin of interface magnetism in $BiMnO_3/SrTiO_3$ and $LaAlO_3/SrTiO_3$ heterostructures. Phys Rev Lett 111:087204

Santander-Syro AF, Copie O, Kondo T, Fortuna F, Pailhè S, Weht R, Qiu XG, Bertran F, Nicolaou A, Taleb-Ibrahimi A, Fèvre PL, Herranz G, Bibes M, Reyren N, Apertet Y, Lecoeur P, Barthélémy A, Rozenberg MJ (2011) Two-dimensional electron gas with universal subbands at the surface of $SrTiO_3$. Nature 469:189

Savrasov SY, Kotliar G, Abrahams E (2001) Correlated electrons in δ-plutonium within a dynamical mean-field picture. Nature 410:793

Sclauzero G, Ederer C (2015) Structural and electronic properties of epitaxially strained $LaVO_3$ from density functional theory and dynamical mean-field theory. Phys Rev B 92:235112

Seth P, Krivenko I, Ferrero M, Parcollet O (2016) TRIQS/CTHYB: A continuous-time quantum monte carlo hybridisation expansion solver for quantum impurity problems. Comput Phys Commun 200:274

Si L, Janson O, Li G, Zhong Z, Liao Z, Koster G, Held K (2017) Quantum anomalous hall state in ferromagnetic $SrRuO_3$ (111) bilayers. Phys Rev Lett 119:026402

Stemmer S, Millis AJ (2013) Quantum confinement in oxide quantum wells. MRS Bull 38:1032

Tanaka H, Matsumoto T, Kawai T, Kawai S (1993) Surface structure and electronic property of reduced $SrTiO_3$(100) surface observed by scanning tunneling microscopy/spectroscopy. Jpn J Appl Phys 32:1405

Werner P, Comanac A, de' Medici L, Troyer M, Millis AJ (2006) Continuous-time solver for quantum impurity models. Phys Rev Lett 97:076405

Zhong Z, Wallerberger M, Tomczak JM, Taranto C, Parragh N, Toschi A, Sangiovanni G, Held K (2015) Electronics with correlated oxides: $SrVO_3/SrTiO_3$ as a Mott transistor. Phys Rev Lett 114:246401

Zubko P, Gariglio S, Gabay M, Ghosez P, Triscone JM (2011) Interface physics in complex oxide heterostructures. Ann Rev Condens Matter Phys 2:141

First-Principles Modeling of Interface Effects in Oxides

47

Alexander A. Demkov, Kurt D. Fredrickson, Hosung Seo, and Andrew O'Hara

Contents

Abstract

This work is an attempt to provide a self-contained introduction to the field of atomic modeling of oxide interfaces, focusing primarily on the first-principles modeling techniques based on density functional theory. By way of introduction,

A. A. Demkov (✉)
Department of Physics, The University of Texas, Austin, TX, USA
e-mail: demkov@physics.utexas.edu

K. D. Fredrickson
Patterson+Sheridan LLC, San Jose, CA, USA
e-mail: kurt_fredrickson@amat.com

H. Seo
Department of Physics, Ajou University, Suwon, Republic of Korea
e-mail: hseo2017@ajou.ac.kr

A. O'Hara
Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA
e-mail: andrew.ohara@vanderbilt.edu

a brief description of common oxides and their technological applications are included. We then discuss the first principles methods that have proven effective in dealing with oxide-based interface problems, including common techniques of building structural models. Examples of validating the atomistic model of an interface experimentally are given along with predictions of interface properties. We conclude with a brief summary of the field and the remaining challenges it faces.

1   Introduction

Metal oxides play an important role in human society. Their applications range from the mundane, such as kitchen glassware, to the very advanced, such as microelectronics and nanophotonics. Somewhat surprisingly, because the main mineral component of human bone is an oxide mineral (hydroxyapatite), oxides also have a connection to healthcare and medicine. In terms of physical properties, metal oxides exhibit a wide variety of physical phenomena: they can range from being electrically insulating to being very good conductors and can also show interesting properties such as magnetism, ferroelectricity, superconductivity, and nonlinear optical effects, to name a few. Some oxides also show catalytic activity or can be used as mechanical supports for catalytically active particles. As a result, metal oxide nanostructures find applications in a very wide range of devices including field effect transistors, light-emitting devices, photodetectors, catalysts, electromechanical devices, sensor elements, solar cells, and many more.

In most applications, oxides do not exist in isolation but are probed (contacted), manipulated, or supported by other materials. Therefore, oxide interfaces are of great practical as well as fundamental interest. Recent advances in oxide deposition techniques started a true revolution in the field of oxide thin films and heterostructures. Physical deposition methods such as molecular beam epitaxy (MBE) and pulsed laser deposition (PLD) or chemical methods such as atomic layer deposition (ALD) have been re-designed specifically to grow metal oxide films and now allow deposition control at the level of almost a single atomic layer. This has in turn enabled the use of oxide films in nanofabrication as well as creation of artificial oxide-based materials that have never been accessible in bulk form previously. The interface with vacuum is just the oxide surface, and we refer the reader to the abundant literature on the topic of oxide surfaces. An excellent starting point is, of course, the classic by Henrich and Cox (1994).

To make our discussion of oxide interfaces more specific, let us consider a device known as a FerroFET, a field effect transistor (FET) with a thin ferroelectric layer, e.g., $BaTiO_3$ (BTO), as a gate dielectric (see Fig. 1). As the ferroelectric "remembers" its polarization state and can be switched between two stable polarization states by a voltage applied to a gate electrode, the device can be used as a nonvolatile memory. Assuming a Si-based FET, putting a BTO layer on Si with MBE requires the use of a thin $SrTiO_3$ (STO) buffer layer, which grows epitaxially directly on
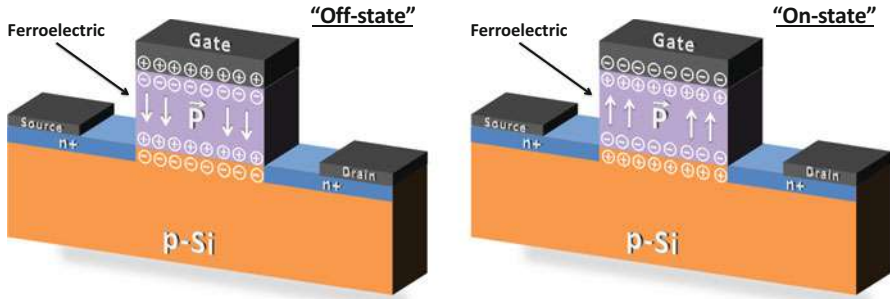
Fig. 1 The ferroelectric field-effect transistor is a MOSFET with a ferroelectric gate, achieved by replacing the dielectric material of the transistor with a ferroelectric material. This nonvolatile, low power memory device would experience reduced fatigue, due to its nondestructive readout, and more aggressive device scaling could be achieved in comparison to FeRAM where there is a separate transistor and capacitor

Si (a BTO film then grows epitaxially on such a buffer). To describe the electrical properties of this gate capacitor (Si-oxide-gate electrode), one needs to understand at least three interfaces involving an oxide material: (i) the semiconductor-oxide interface (Si-STO), (ii) the oxide-oxide interface (STO-BTO), and (iii) the oxide-metal (BTO-electrode) interface. If in the process, the top Si layer is oxidized and the SiO_2 interlayer is formed, another two interfaces need to be considered, one between Si and SiO_2 and another between SiO_2 and STO.

Note that despite all being oxide-related, these interfaces are quite different from each other, both structurally and chemically, which has implications in how one would model them. For an interface between two crystalline materials, such as Si and STO, one may hope to find a common periodicity unit that would allow the use of the periodic boundary conditions (at least laterally). For the interface involving amorphous SiO_2, this, strictly speaking, cannot be modeled precisely and one has to either use some crystalline form of SiO_2 (such as β-cristobalite) as a proxy or attempt to generate a quasi-disordered SiO_2 structure that can then be used as repeated unit in a periodic calculation. The electrostatic screening in metals and insulators also has a different nature, which places very different requirements on the simulation cell size.

Different interfaces may also require different types and levels of theory. While Si, Sr, or Ba prefer fixed oxidation states (4^+, 2^+, and 2^+, respectively) at oxide-based interfaces, transition metals such as Ti can support several different oxidation states. The existence of multiple oxidation states can be traced to electron correlation in the d-shell of the transition metal ion. A proper treatment of electron correlation requires electronic structure methods going beyond the usual mean-field approximation that is commonly used in approximate density-functional-theory-based methods.

Different applications depend on different properties of the interface. One of the most important properties of an oxide heterostructure is the band offset (the band discontinuity across the interface between two different materials). The band offset plays a key role in determining the degree of carrier confinement and electronic transport along and across the interface. This is a key property for electronic devices as well as for photocatalytic cells. Magnetic properties of the interface are important for spintronics, magnetic storage, and sensing applications. In addition, the interface controls both the electrostatic and mechanical boundary conditions in thin ferroelectrics. Charge transfer at the interface may cause structural distortions and induce metal-to-insulator transitions. Calculating such properties is the main task of the atomic modeling.

The atomistic modeling of an interface can often be performed reliably, provided the atomic structure of an interface is well known. Deciphering the interface structure at the atomic level experimentally, on the other hand, is rather difficult. The atomic structure, of course, can be generated through theoretical modeling; however, due to many uncertainties, the structural model of the interface inferred from theory needs to be properly validated before one makes conclusions about the calculated physical properties using the said model. In some cases, the atomic arrangement in certain directions can be reconstructed directly from electron microscopy. Often photoemission or optical spectroscopic methods such as infrared (IR) absorption or Raman spectroscopy are used to understand the interfacial bonding. (Note that these are difficult experiments as we are extracting information from just a few atomic planes buried under a film and in the presence of a substrate). If the results of such experiments could be predicted theoretically using the interface model, than comparison between theory and experiment can be used to validate the atomic model of the interface.

Therefore, the problem of modeling an oxide interface has two parts: first, an atomistic model of the interface needs to be generated. We emphasize that this is a complicated task as often the atomic geometry and bonding across the interface are not known a priori. Once a structural model is generated, one can compute some of its properties observable in experiment, such as the valence band density of states that can be checked with X-ray or UV photoemission spectroscopy (XPS and UPS, respectively), core-level shifts that can be probed by XPS, or interface-specific vibrational modes to be compared with IR or Raman data. These theoretical values should then be compared with the available experimental data for validation. If the model passes the validation, the second part can be accomplished by computing the device-specific properties of interest.

In this chapter, we will briefly introduce oxide materials in Sect. 2 and describe different types of technologically important oxide interfaces in Sect. 3. In Sect. 4, we will briefly introduce density functional theory and discuss specific methods of building structural models of the interface. Several recent examples of model validation will be given in Sect. 5, and in Sect. 6 we will discuss computational modeling of physical properties of several interfaces. In conclusion, we will describe various remaining challenges in modeling oxide interfaces.

2   Oxides

Oxides are compounds of oxygen with other elements. For the purpose of the following discussion, we shall divide oxides into categories based on the chemistry of the constituent metals and the degree of long-range order. Thus, we will distinguish simple oxides (e.g., SiO_2 or Al_2O_3, see Fig. 2) from transition metal oxides (e.g., VO_2 or $SrTiO_3$) and crystalline oxides (e.g., MgO or α-quartz SiO_2) from disordered or amorphous oxides (amorphous HfO_2 or amorphous SiO_2).

Traditionally, oxides are thought of in terms of the ionic model, where a positively charged metal cation is surrounded or coordinated by negative oxygen anions. This view has some limitations but captures the main feature of oxide structure: the need to minimize the electrostatic energy of a system of positively charged metal ions and negatively charged O^{2-}. The negative oxide ion O^{2-} cannot exist in free space and is only stabilized by the ionic lattice (its Madelung potential) and therefore is extremely sensitive to crystal structure. The key attributes of a crystal structure are stoichiometry and coordination. Stoichiometry describes the relative number of different atomic species and coordination describes the number of ions surrounding a given ion. For example, in the case of α-quartz SiO_2, the stoichiometry says that for each Si ion we have two oxygen ions. The coordination of Si by oxygen is 4 (each Si has 4 oxygen nearest neighbors) and that of oxygen by Si is 2 (see Fig. 2).

The structure is a (4;2) network of almost rigid, corner sharing SiO_4 tetrahedra. On the other hand, the description of the crystal structure in terms of bonding can be insightful. Figure 2 shows the ball-and-stick atomistic model for α-quartz as well as for corundum Al_2O_3 that shows how each atom is bonded to its neighbors.
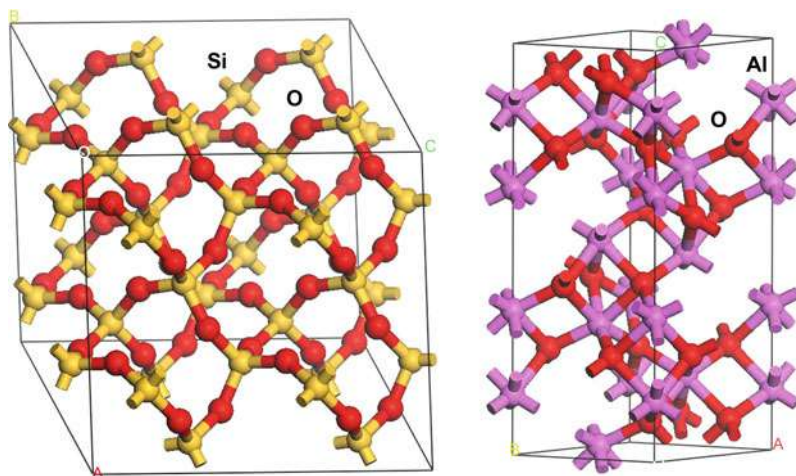
Fig. 2   Simple oxides: SiO_2 a-quartz (left) and corundum Al_2O_3 (right)
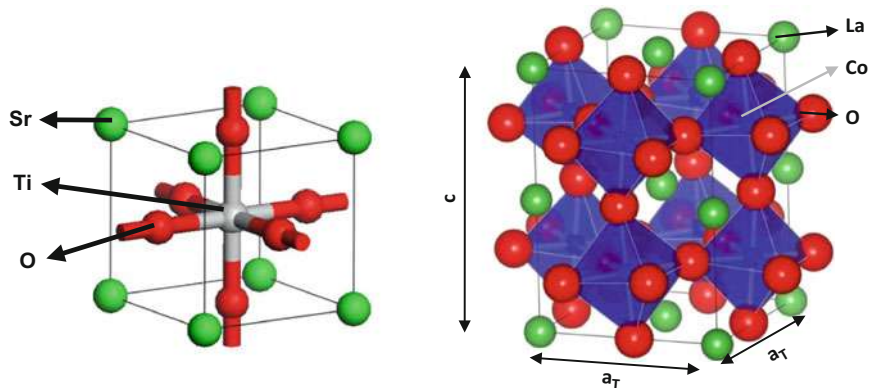
Fig. 3 The simple perovskite structure of cubic SrTiO$_3$ on the left is in sharp contrast with the distorted perovskite-related structure of LaCoO$_3$. The important physics is captured in the tilts of oxygen octahedra shown in blue

The simple perovskite structure ABO$_3$ (SrTiO$_3$, BaTiO$_3$) is shown in Fig. 3. The Ti ions are in the center of a cube and are six-fold coordinated by oxygen ions that assume the centers of cube faces, while Sr is at the cube corners. The structure has a transition metal ion on the B site (in this case Ti) octahedrally coordinated by oxygen. The A site is 12-coordinated in the ideal perovskite, but the coordination can be reduced to as low as 8 by tilting the BO$_6$ octahedra. The ideal perovskite structure is cubic, with orthorhombic and tetragonal phases being the most common noncubic distortions of the structure. A simple structure like that is well explained by a ball-and-stick model shown on the left side of Fig. 3. At the same time, LaCoO$_3$ crystallizes in a structure that is related to a perovskite but is slightly more complicated (rhombohedral distortion). On the right side of Fig. 3, we show the structure with corner-sharing octahedral units explicitly present. The octahedra are rotated and tilted with respect to the orthogonal axes but maintain connectivity. These distortions have a profound effect on the properties of LaCoO$_3$, but the transition metal coordination has not changed. Structurally related to perovskite oxides are bronzes (e.g., WO$_3$) that can be viewed as a perovskite with all A cation sites vacant.

Many common binary oxides crystallize in the cubic ***rock salt*** structure (MgO. CaO, NiO, MnO, and EuO); this structure is shown in the left panel of Fig. 4. Each atomic species is on a separate face-centered cubic Bravais sublattice. The coordination number of each atom is 6. Each metal ion is coordinated by 6 oxygen ions and each oxygen is coordinated by 6 metal ions. In terms of coordination polyhedra, the rock salt structure is composed of edge-shared octahedra. The space group is Fm$\bar{3}$m. The interatomic distance is half of the cubic unit cell length.

Another common structure of binary oxides is tetragonal ***rutile*** (TiO$_2$, SnO$_2$, and RuO$_2$). A unit cell of TiO$_2$ rutile is shown on the right side of Fig. 4. The metal
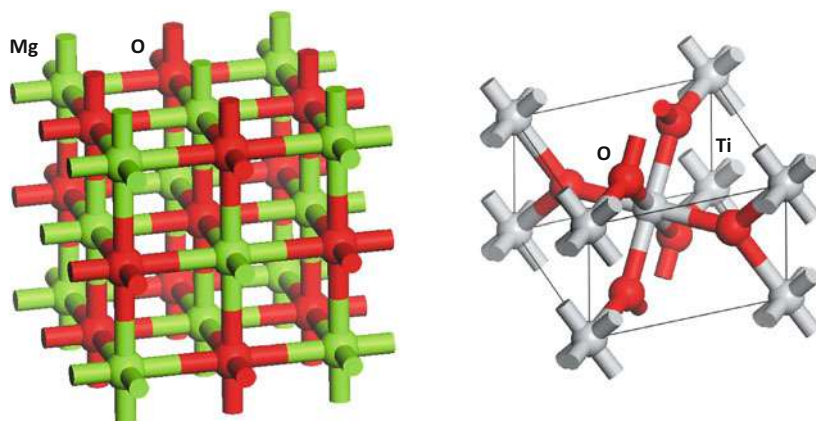
Fig. 4 The rock salt structure of MgO is shown on the left and a primitive unit cell of rutile TiO$_2$ is shown on the right

cations are six-fold coordinated by oxygen atoms, which form an octahedron around them. These octahedra are arranged in a combination of edge- and corner-sharing. The oxygen anions have a trigonal planar coordination with a coordination number of 3. The space group is $P4_2$/mnm.

Simple oxides such as MgO, SiO$_2$, or Al$_2$O$_3$ differ from the complex transition metal oxides, such as SrTiO$_3$, in their nature of bonding. Simple sp^3 metals such as Si or Al are typically found in a single dominant oxidation state of 4$^+$ or 3$^+$, respectively, and the properties of corresponding oxides are described well within the band structure picture. On the other hand, the ions of transition metals can support multiple oxidation states due to complex electronic reconstruction in the often-correlated d shell. The five-fold degeneracy of the transition metal d state is lifted by the cubic field of the oxygen octahedron. As a result, a triplet known as a t_{2g} state and a doublet known as an e_g state are separated in energy by an energy gap known as the crystal field energy 10Dq. Including into consideration the interaction between the metal d orbitals and p orbitals of the oxygen, adds another energy separation, now between the bonding and antibonding combinations of these states, which is known as a charge transfer gap Δ_{CT}. If the metallic d states are strongly localized, one can think of the problem in the atomic limit and use Hund's rules to predict the electronic configuration. If localization is weak, the itinerant or band structure picture is appropriate. In the intermediate regime, one encounters Mott-Hubbard physics and the eponymous metal-to-insulator transition. If the interaction with the lattice is strong, then Jahn-Teller physics becomes important. The surprising closeness of all the aforementioned energies is responsible for the astounding richness of phenomena observed in transition metal oxides and the reason why they are called complex oxides.

3   Types of Interfaces

As we have discussed in the introduction, depending on the technological application, an oxide film can have interfaces with different types of materials. A simple oxide can interface with a semiconductor, as in the case of the SiO_2-Si interface of the metal-oxide-semiconductor (MOS) stack. One can also have an interface between a complex and a simple oxide, like in the case of SiO_2-HfO_2 that occurs in high-k gate stacks of FETs (Sharia et al. 2007). In the same devices, HfO_2 also has an interface with metals such as Mo (Demkov 2006). The very same metal-complex oxide interface is also responsible for the operation of resistive random access memory (RRAM) devices (O'Hara et al. 2014). Complex oxides can have interfaces with each other as in $LaAlO_3$/$SrTiO_3$ (LAO/STO) interface and so can simple oxides, for example, SiO_2/Al_2O_3 in core-shell structures.

In the field of medical implants, biominerals such as hydroxyapatite (HA) have interfaces crucial to their use, for example, with Ti, water, and the TiO_2 substrate. Wetting properties at the hydroxyapatite/H_2O interface govern its stability characteristics and bioactivity as the hydroxyapatite crystallites get exposed to aqueous solutions in vivo, whereas wetting of hydroxyapatite on the TiO_2 surface dictates its adhesion characteristics to the implant backbone. In addition to the active experimental research conducted in this area (Leon and Janson 2009), a fundamental theoretical understanding of the physics and chemistry at the two relevant interfaces (TiO_2/HA and HA/H_2O) is clearly needed (Slepko and Demkov 2011, 2013). While strong adhesion of HA to the TiO_2 substrate is necessary to prevent premature deterioration of the implant's surface, water is the main medium affecting its ability to connect to biological tissue once immersed into the human body.

Complex oxides can exhibit a wide variety of novel functional properties such as ferromagnetism, ferroelectricity, and metal-insulator transitions, which have been extensively studied during the past decades (Imada et al. 1998). Furthermore, it has even become possible to create and control hybrid oxide heterostructures with abrupt epitaxial interfaces owing to recent advances in the field of oxide heteroepitaxy (Chambers 2010). Epitaxial deposition techniques with the capability of controlling interface composition, length scales, etc., have opened totally new and exciting scientific avenues and have also offered novel device applications to explore (Hwang et al. 2012). One of the most important findings in this research direction is the creation of the two-dimensional electron gas (2DEG) between two insulating complex oxides, which are LAO and STO (Ohtomo and Hwang 2004). The origin of the 2DEG has been largely discussed in the literature, ranging from the original polar-catastrophe argument to the presence of interfacial O vacancy and intermixing at the interface (Mannhart et al. 2008). Furthermore, the LAO/STO interface shows a rich variety of interesting physical properties including magnetism, superconductivity, etc. Now, the field is expanding, exploring epitaxial interfaces of diverse complex oxides, and finding new physics and new emergent phenomena (Chakhalian et al. 2014).

In addition to the interface effects for the emergent properties of oxide heterostructures, bulk properties of an oxide material can be drastically changed (and controlled) when it is integrated as a part of heterostructure (Ismail-Beigi et al. 2017). A key example is strain engineering that one can achieve by epitaxially growing one oxide material on another oxide substrate having a different lattice constant. The perovskite oxides are known to accommodate a relatively high degree of structural distortion such as rhombohedral and tetragonal distortions. In particular, the lattice has additional degrees of freedom in the oxygen octahedra, which can cooperate to tilt and rotate under external perturbations (Rondinelli et al. 2012). Therefore, for the oxide overlayer commensurate with the substrate, epitaxial strain can directly affect the lattice degrees of freedom. Several studies have shown that strain engineering of oxide heterostructures could stabilize novel electronic phases that are not present in their bulk constituents, owing to the close interplay between the lattice, orbital, and spin degrees of freedom present in the complex transition metal oxides (Ismail-Beigi et al. 2017). One exciting example is the recent demonstration of biaxial tensile strain stabilizing an insulating ferromagnetic ground state in $LaCoO_3$, which is normally nonmagnetic at low temperature (Fuchs et al. 2007a, 2008). A number of possible mechanisms have been discussed including low-spin/high-spin mixed state (Seo et al. 2012a) and formation of ordered O vacancy structures (Gazquez et al. 2013).

In the field of renewable energy, oxide-based interfaces play a crucial role, in particular, for the photocatalytic water splitting to generate hydrogen fuel. The interfaces of interest are oxide/nano-metal-clusters (Schneider et al. 2018) and oxide/liquid interfaces (Björneholm et al. 2016). Many complex oxides were identified as promising candidates for photo-catalysts for water splitting. $SrTiO_3$ is an excellent example (Carr and Somorjai 1981). The band gap of STO is 3.2 eV, and its conduction-band minimum, which is positioned about 4.0 eV below the vacuum level, matches up well with the redox potential of water. The overall quantum efficiency is, however, low and a number of limiting factors have been discussed (Chen et al. 2010). A major challenge is to decrease the electron-hole recombination rates after they are created by photon irradiation. A possible solution to this problem is to develop a suitable heterogeneous catalyst that could efficiently separate electrons and holes at the interface. Notably, noble metals, such as Pt, are frequently loaded on titania as a co-catalyst to boost the catalytic activity (Paracchino et al. 2011). The enhanced catalytic activity has been attributed in part to the fact that the rectifying behavior of the Schottky barrier formed between STO and Pt may allow Pt nanoclusters on the oxide surface to trap electrons, effectively reducing the oxidants in a photocatalytic process (Chen et al. 2010). In addition to the oxide/nano-metal-cluster interface, the oxide/water interface is also very important to understand as in a typical photo-electrochemical environment, mixed oxide catalysts are interfaced with water. Thus, the oxide/water interface is actually a crucial component of the entire device to split water. In this regard, there is an excellent review recently published to discuss atomistic modeling of oxide/water interfaces (Pham et al. 2017).

In most cases, we do not know how two materials match at the atomic level, so we need to model the interface theoretically. Usually one tries several plausible atomistic models and chooses the one with the lowest total energy, though nature uses the lowest free energy. One could, in principle, include the entropic terms due to, e.g., cation intermixing and phonons, but it is difficult and not usually done. Also, depending on the type of the interface, one can get away with a small segment replicated via the periodic boundary conditions. But if materials are not epitaxial, a larger lateral length scale is required, which puts an additional constraint on what calculation methods can be used.

4   Atomistic Modeling of Oxide Interfaces

4.1   Density Functional Theory

In this section, we describe modeling techniques to generate atomic level structural models of various oxide interfaces. Systems of interest include metal-oxide, oxide-oxide, and semiconductor-oxide interfaces. There are two basic ways to model the interface structure at the atomic level. One can either model the process of the interface creation, e.g., oxidation of Si or growth of a thin film. Then the interface structure comes out as a result of a direct simulation. Alternatively, one can make a guess of an interface by matching two materials and optimizing bonding across the resulting interface in a simulation. The process modeling or the optimization can be done using molecular dynamics (MD) or Monte Carlo (MC) methods. In turn, the forces used in MD simulations or the energy expression in MC simulation can be classical or quantum mechanical. Classical potential fields offer the advantage of scale (larger systems can be modeled because the simulation is fast), while quantum mechanical treatment offers higher reliability, but at much higher cost. Validation is usually done quantum mechanically, though some properties and, in particular, the mechanical properties can be modeled classically. And of course, magnetic, electric, or optical properties all require a quantum mechanical treatment. Thus, we start with a brief introduction to ab initio techniques.

Density functional theory (DFT) has gained wide-spread acceptance as one of the predominant ab initio or "first-principles" tools for materials modeling owing to its ability to provide relatively accurate atomic and electronic structures under the appropriate approximations. Fundamentally, the task is to solve the many body Hamiltonian for the system in question which is comprised of the kinetic energy of all electronics and nuclei as well as their interactions with each other and is simplified via application of the Born-Oppenheimer approximation (Born and Oppenheimer 1927). Although several methods exist to tackle the problem of determining the electronic structure including Hartree-Fock, configuration interaction, coupled cluster, and perturbative approaches, density functional theory remains one of the most efficient and accurate methodologies. Density functional theory would be an exact theory if the appropriate exchange-correlation (XC) and kinetic energy functionals could be found. While no such universal functional is known,

we rely on a series of approximations and improvements such as the local density approximation (LDA), generalized gradient approximation (GGA), meta-GGAs, and hybrid functionals. As an example, in LDA, the XC potential depends only on the values for the homogeneous electron gas density for a given position in space with the exchange portion calculated exactly and the correlation portion parameterized from quantum Monte Carlo simulations (Ceperly and Adler 1980). Two of the most common flavors of LDA are due to Vosko-Wilk-Nusair (VWN) (Vosko et al. 1980) and Perdew-Zunger (PZ81) (Perdew and Zunger 1981) that use different interpolations of the Ceperly and Adler data. GGA functionals introduce some nonlocality to the XC functional by introducing spatial gradients of the electron density to the functional as in the popular functionals of Perdew-Wang (PW91) (Perdew et al. 1992), Perdew-Burke-Ernzerhof (PBE) (Perdew et al. 1996), and others. Generally each successive step for the functional introduces additional computational complexity and the researcher must choose the best functional to describe the system with a reasonable computational cost. For further details on these developments of DFT, we direct the reader to Chap. 11, "Recent Developments in Density Functional Approximations," by Li and Burke in Handbook of Materials Modeling, Vol. 1, Methods: Theory and Modeling.

Beyond the choice of the functional, other considerations must be made. From a practical standpoint, the wave-functions must be determined within a basis set including real-space finite-difference approaches (such as in the PARSEC code (Chelikowsky et al. 1994a, b)), localized orbitals such as Gaussians (Boys 1950) or linear combinations of atomic orbitals (like in the SIESTA code (Soler et al. 2002)), slater-type orbitals (Slater 1930), plane-wave methods, and hybrid methods which combine localized regions and plane waves (such as (linearized) augmented plane waves (Singh and Nordstrom 2006) and localized muffin-tin orbitals (Skriver 1984)). Owing to their ease of use in systems with periodic boundary conditions, plane-waves are a common choice and implemented in the popular Vienna *ab-initio* Simulation Package (VASP) (Kresse and Furthmüller 1996) and QUANTUM ESPRESSO codes (Giannozzi et al. 2009). Regardless of the basis set used, care must be taken to ensure proper convergence (whether this is the finite-difference grid spacing, number of higher order local orbitals, or the plane-wave cutoff energy). In the case of plane-wave-based codes, portions of the calculation are performed in reciprocal space (or momentum space) and the Brillouin zone must be sampled on a well-converged grid of points (often a Monkhorst-Pack grid (Monkhorst and Pack 1976)). As the volume of the real space cell and reciprocal space Brillouin zone are inversely proportional, fewer k-points are needed for calculations of supercells and heterostructures than their primitive cell counter parts.

Finally, including all electrons in a calculation can lead to severe limits on system size. For example, in the (100) $BaTiO_3$/(100) Pt simulation cell discussed later and shown in Fig. 6 (Fredrickson et al. 2013), there are 10 Ba atoms, 11 Ti atoms, 32 O atoms, and 20 Pt atoms which would imply a total of 2618 electrons. Fortuitously, only the valence electrons participate in the formation of bonds. Hence, we can replace the total electron interaction with an effective interaction through the use of pseudopotentials which treats the core electrons as effectively screening the

nuclear interaction. In certain cases, electrons that are a few shells removed from the valence may also be included explicitly (the so-called semi-core states) to improve the description of the valence electrons. Through the use of pseudopotentials, the total number of considered electrons in the (100) BaTiO$_3$/ (100) Pt is actually only 582, cutting the number of electrons considered by roughly 80%. Common choices include norm conserving pseudopotentials (Rappe et al. 1990; Troullier and Martins 1991) with various pseudoization schemes, ultrasoft pseudopotentials (Vanderbilt 1990), and the projector-augmented wave (PAW) method (Blöchl 1994; Kresse and Joubert 1999).

For the reader wishing to explore further the theoretical and numerical underpinnings of DFT, we direct them to Li and Burke's chapter as well as several other highly detailed treatments such as in (Dreizler and Gross 1990; Engel and Dreizler 2011; Fiolhais et al. 2003; Martin 2004) as well as the manual of their chosen DFT code.

4.2   Construction of Interface Models

In the following discussion, we will assume that the interface is always perpendicular to the c axis of the considered substrate for convenience. Once the Miller indices of the substrate and the film have been decided, the next step is to consider the alignment of materials A and B. In order to do this, one must consider the most likely source of bonding at the interface. For example, one may want to line up the interface such that the O atoms at the oxide surface are able to bond with appropriate atoms at the interface (there is a strong probability of Ge-O bonds at the Ge/Oxide interface (Fredrickson et al. 2014); however, bonding with O at a noble metal/oxide interface, such as Pt, is probably not very energetically favorable (Fredrickson et al. 2013)).This requires using physical and chemical intuition and there are at best only rough guidelines to follow, and most likely several alignments of the two materials must be tried to find the lowest energy interface. To find the optimum separation distance, an efficient method is to vary the interfacial spacing between the two materials, perform several DFT calculations, and fit the energy/distance curve quadratically to find the expected minimum.

Once an appropriate interface alignment has been chosen, one needs to consider the possible "defects" or missing or extra atoms present at the interface. For example, there may be vacancies at the interface (missing atoms) from either material, or possibly both. These defects are stabilized by chemical conditions at the interface that would not otherwise be energetically favorable in the bulk or at the material's surface (defects also reduce free energy of the system via the entropy term, but this effect is rarely included in the first principles analysis). In addition, extra atoms may be present at the interface that depend heavily on the chemical conditions of the interface under consideration (e.g., extra O atoms may be present at the interface when grown under oxidizing conditions (Fredrickson et al. 2014), whereas you may not expect these same atoms for materials grown under high vacuum (Fredrickson et al. 2013)). Treatment of the substrate chemically may
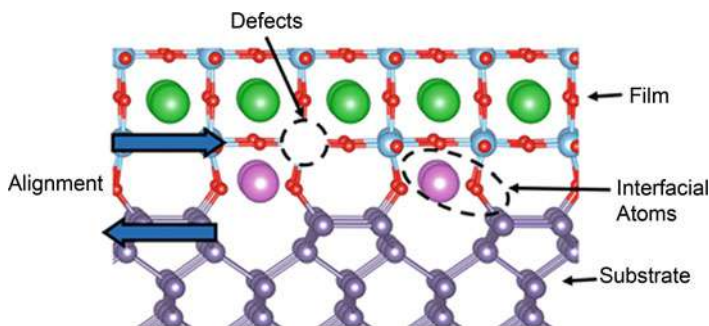
Fig. 5 An example of an interface (Ge/BaTiO$_3$) that includes considerations for interfacial alignment, interfacial atoms, and defects. Green atoms are Ba, red atoms are O, blue atoms are Ti, pink atoms are Sr, and purple atoms are Ge

introduce other atoms to the interface (e.g., for BaTiO$_3$ grown on Ge, Sr, or Ba atoms are sometimes introduced as a passivation layer, and these atoms are critical for predicting correct band gaps and electronic properties (Fredrickson et al. 2014; Ponath et al. 2015). Examples of the above extra conditions are listed below in Fig. 5.

The implicit assumption we have made so far is that there are just as many unit cells of the substrate A as there is film B, but this will not always be true. This will only work if the specific cut of both the film and the substrate have similar lattice constants, which minimizes the strain of the film. If this is not the case, and the one-to-one unit cell matching between the substrate and film induces a tensile or compressive strain greater than about 5%, the energetics of film growth will be unfavorable, and the results from the calculation will be unphysical. Thus, it is important to find whatever n-to-m (coincidence lattice matching) ratio minimizes the strain and does not result in too large of a unit cell to be unpractical to calculate (e.g., a 10-to-13 matching of substrate may give a very low strain on the film but may be hard to calculate because the system is so large!). One also must consider if the interface matches in both directions of the plane (e.g., the interface may match 1-to-1 in the x direction but (2-to-1) in the y direction). This mismatch can also depend on the specific cut of the film and substrate (Fig. 6).

Another common pitfall on interface modeling is specifically due to the electronic structure, not to the atomic structure as discussed above. Due to the underestimation in band gaps for the LDA and GGA approximations, there can be issues with band alignment due to the incorrect position of the valence band and conduction band of the substrate with respect to either the Fermi level of the metallic film, or the valence or conduction bands of the semiconducting film (e.g., the conduction band of the substrate can be lower in energy than that of the Fermi level of the metal, even though the experimental work functions and electron affinities of the substrate and metal are predicted to be ordered oppositely). In such cases there is a flow of charge from the substrate to the film (or vise-versa) which is unphysical

Fig. 6 An interface between BaTiO$_3$(001) and Pt, with the supercell model of BaTiO$_3$/Pt in a solid-line box. (**a**) The Pt is grown in the (001) direction, with 1-to-1 alignment of Pt on BaTiO$_3$; the alignment is 1-to-1 in the both x- and y-direction. The figures on the right side show the first layer of Pt (top) and the surface layer of BaTiO$_3$ (bottom) at the interface region indicated with a red dashed box in the supercell model. (**b**) The Pt is grown in the (011) direction. The interface matching condition (red dashed box) is shown on the right side in detail, in which a (2x1) BaTiO$_3$ surface cell is used to match the (3x1) Pt(011)
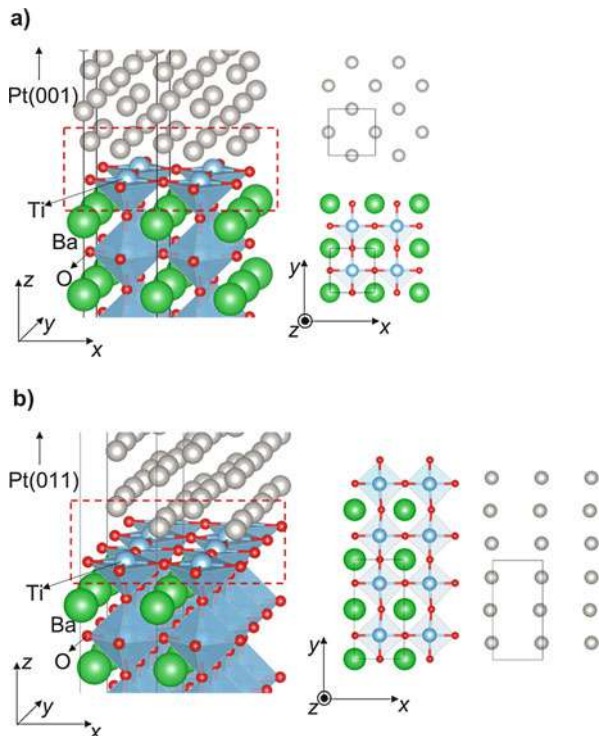
and is solely due to the incorrect energy level ordering in the simulation. To prevent this unphysical flow of charge, a higher level of theory is required that can give more accurate band gaps, such as hybrid functionals (Krukau et al. 2006) or the *GW* method (Fuchs et al. 2007b; Shishkin and Kresse 2006, 2007; Shishkin et al. 2007). A computationally less expensive solution is to use the LDA + U method (Dudarev et al. 1998), which includes a Hubbard-type U that can also increase the band gaps of materials, and which is also effective in preventing the unwanted charge transfer.

Due to the periodic boundary conditions used in many DFT codes, any calculation must enforce an equal potential at both sides of the cell, i.e., the potential at $z = 0$ and $z = l$ must be equal, where l is the total length of the simulation cell. This provides an extra complication for modeling interfaces, where, by definition, the simulated cell is not symmetric. An important note is that the code will enforce boundary conditions, meaning that if the cell is set up incorrectly, there will be an unphysical electric field across the cell that one might not expect! There are two commonly used methods to avoid this problem. The first way is to consider a symmetric cell, so that the condition of periodic potentials is automatically satisfied (Fig. 7a). For surface problems, one could include a mirror-symmetric structure of a slab model in a supercell to meet the periodic boundary condition as shown in Fig. 7b. A similar method can be used for an interface problem, for which a symmetric supercell such as one shown in Fig. 7a is not possible to construct. As
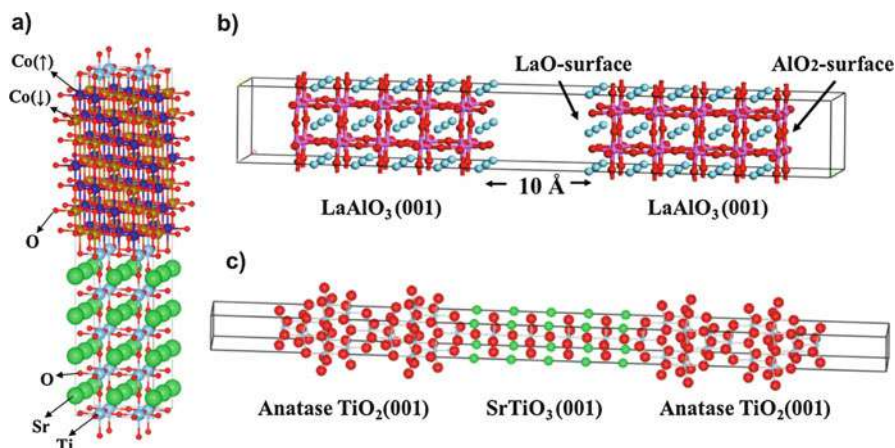
Fig. 7 A choice for the interface model between material A and B that does not satisfy periodic boundary conditions. (**a**) A superlattice structure could be constructed to study the interface between antiferromagnetic CoO(001) and SrTiO3 (001) within the periodic boundary condition. (**b**) For a surface problem where the surface dipole or an internal dipole moment play an important role, a symmetric slab model could be constructed by containing a slab model's mirror-symmetric structure in the same supercell. An example of polar $LaAlO_3$(001) is shown. (**c**) The mirror-symmetry method described in (**b**) could be used to study an interface problem, for which the superlattice approach (**a**) is not possible. An example of the anatase TiO_2/$SrTiO_3$(001) interface is shown, in which the periodic boundary condition along the c-axis is satisfied by including the TiO_2 layer and vacuum on both sides of the $SrTiO_3$(001) slab. (With permission from Seo and Demkov (2011, 2014), and Seo et al. (2012b))

shown in Fig. 7c, one could include vacuum on either side of the slab to satisfy the periodic boundary condition (Fig. 7c). However, sometimes, this first way of dealing with an artificial electric field could be computationally demanding as it would require doubling the supercell size. To be more computationally efficient, one could use an asymmetric slab model with a dipole field correction that is equal in magnitude but opposite in direction to cancel the artificial electric field induced by the use of an asymmetric slab model (Bengtsson 1999; Makov and Payne 1995). A self-consistent dipole is added to the system with a magnitude chosen to counteract the dipole created by having unequal surface termination on either side of the slab; this dipole creates a large jump in the electronic potential, which ensures that the vacuum levels remain flat. Since the dipole is placed in vacuum, it is far enough away from the surfaces of the slab to not affect the interface or bulk properties of the system. This counteracts the unphysical electric field caused by the nonequal surfaces of the system. In Fig. 8, we show the electrostatic potential calculated for a layer of water, which is highly polar, with and without the dipole correction. We see that without the dipole correction, there is an unphysical field in the vacuum which forms due to the periodic boundary conditions of the cell. With the dipole correction, this field is localized in the vacuum region of the cell, and the electrostatic potential of the water layer is unaffected.

Fig. 8 Laterally averaged electrostatic potential for the electrons in the supercell, with (solid-curve) and without (dashed-curve) dipole correction. (With permission from Bengtsson (1999))
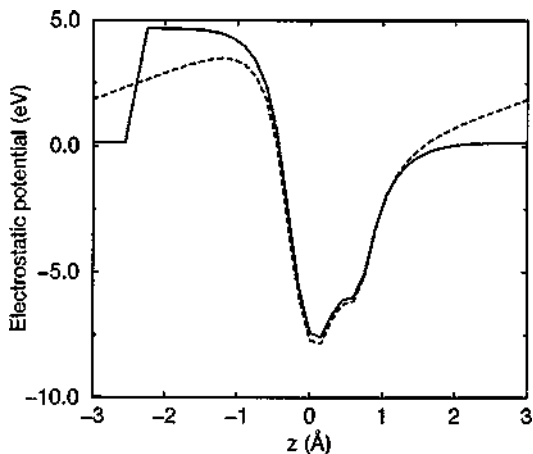
Despite their success, it is worth remembering that *ab initio* methods have some limitations. In particular, the length and time scales of such simulations are rather limited, and in many cases, classical MD and MC techniques become very useful. For example, a problem of cationic exchange at the $LaAlO_3$ and $SrTiO_3$ and its effect on the formation of the two-dimensional electron gas (2DEG) at this interface (Ohtomo and Hwang 2004) has recently received a lot of attention. Stoichiometric epitaxial $LaAlO_3$ grown on TiO_2-terminated $SrTiO_3$ (001) by off-axis pulsed laser deposition is shown to exhibit strong cation intermixing. Qiao and co-workers performed a detailed study of this system using a combination of classical and quantum mechanical calculations of the relative stabilities of abrupt and intermixed interface configurations (Qiao et al. 2010). They were able to show very convincingly that the experimentally measured valence band offset of 0.16 ± 0.10 eV cannot be accounted for theoretically without including intermixing in the physical description of the interface.

5   Validation of Interface Models

There are many experimental methods used to study interfaces, such as scanning transmission electron microscopy (STEM), electron energy loss spectroscopy (EELS), XPS, angle-resolved photoemission spectroscopy (ARPES), and they all can be used to validate the structural model. One needs to use the theoretical structure of the interface and compute a measured quantity. If several properties of the interface are well reproduced, the model is most likely reliable.

The combination of STEM and EELS is one of the most powerful methods to validate an interface model of an oxide-based heterostructure (Varela et al. 2012). The basic principle of STEM is described in detail elsewhere (Pennycook and Nellist 2011). In a STEM image, an atom with a greater atomic number produces a brighter spot than others with a lower atomic number. Using this "Z-contrast"
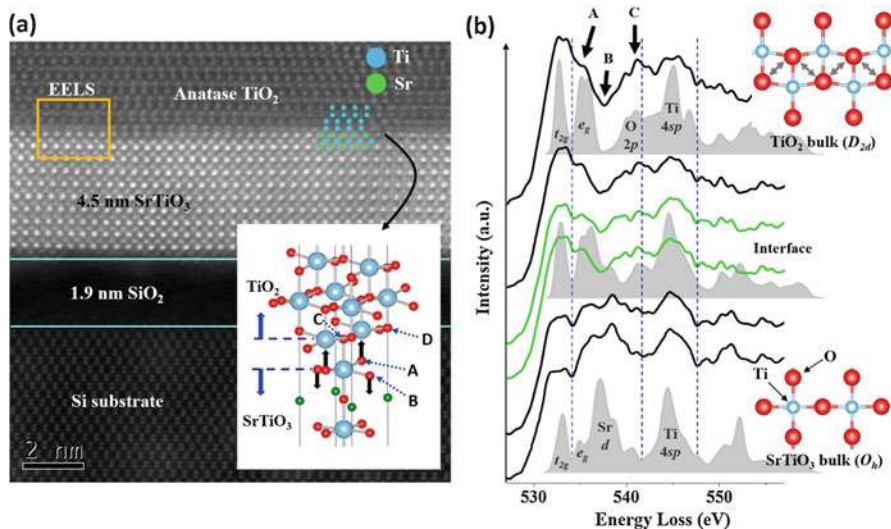
Fig. 9 (**a**) Z-contrast STEM image of the $TiO_2/SrTiO_3$ (001) structure and its atomic interface model (inset). Ti, O, and Sr are represented by large blue ball, middle-sized red ball, and small green ball, respectively. In the inset, black arrows show the relaxation pattern of the O ions at the interface. Spatially resolved EELS measurement was performed at the region indicated by the orange box. (**b**) EELS O K edge spectra taken at the interface from the STO side (two bottom spectra) to the anatase TiO_2 side (two upper spectra) through the interface (two middle green spectra). The corresponding 2p-projected DOSs calculated within the $Z + 1$ approximation are overlaid in grey. Two insets show schematic pictures of a TiO_2 plane in bulk anatase TiO_2 (top) and bulk STO (bottom). (With permission from Seo et al. (2012b))

property, one can directly image an oxide heterointerface at the atomic scale. Figure 9a shows the STEM image of the anatase TiO_2/STO (001) heterostructure grown on Si(001) (Seo et al. 2012b). The image was taken along the Si [110] zone axis. A well-ordered epitaxial relationship is shown in the image between the Si substrate, $SrTiO_3$ layer, and anatase TiO_2 film. To infer the most likely structure of the STO/TiO_2 interface, the blue and green balls were superimposed in the image, and the inset of Fig. 9a shows the deduced structural model of the interface. This atomic model can be used as input to subsequent density functional theory calculations for predicting various interface properties, which we will describe later in this chapter.

In addition to atomic-scale imaging, one can analyze the energy of beam electrons in STEM inelastically scattered from the sample (known as electron energy loss spectroscopy or EELS (Pennycook and Nellist 2011)). The inelastically scattered electrons carry information about the local electronic structure of the atom or the local region that they scattered off. Therefore, one can further verify an interface model deduced from a STEM image by comparing the EELS spectrum that is simultaneously obtained from the region of the STEM image to the electronic structure theoretically calculated with the interface model.

For the $TiO_2/SrTiO_3$ heterointerface shown in Fig. 9a, an EELS spectrum was taken across the interface and is shown in Fig. 9b. In particular, an O K-edge spectrum was taken, in which beam electrons in the STEM measurement lose their energy by exciting the 1 s core electrons of oxygen atoms. We note that the optical electronic transition observed in EELS should satisfy the dipole selection rule: the angular-momentum quantum number of a final state should differ by $+1$ or -1. Therefore, in the O K-edge spectrum in Fig. 9b, the 1 s electrons of oxygen atoms are excited into the empty O 2p states of the TiO_2/STO heterostructure. This means that the scattered beam electrons provide the conduction band structure of the TiO_2/STO heterostructure that is derived from the antibonding states of O 2p orbitals.

To track the evolution of the antibonding states of O 2p orbitals across the TiO_2/STO interface, Seo et al. performed DFT calculations (Seo et al. 2012b) at LDA level of theory, using the Ceperley-Alder data parametrized by Perdew and Zunger (1981), as implemented in the VASP code (Kresse and Furthmüller 1996). Projector-augmented wave pseudopotentials (Blöchl 1994) were employed to describe Sr, Ti, and O, along with a plane-wave cutoff energy of 600 eV. Several supercell models were constructed by varying the anatase film thickness from four to eight monolayers in order to model the evolution of the valence band as a function of the film thickness. The slab in each supercell was separated from its nearest neighboring images in the c-direction by 14 Å-thick vacuum space to avoid any artificial interactions.

At the bottom of Fig. 9b, the O K-edge spectrum taken from the STO bulk region is compared to the 2p pDOS calculated at the O site in the STO bulk region of the TiO_2/STO interface model. In experiment, three main features are found between 530 and 550 eV, which are all well reproduced in theory. The first, second, and third main peaks were identified as peaks derived from the interaction with Ti t_{2g}, Sr d, and Ti 4sp states, respectively. On the other hand, in the spectrum taken from the anatase bulk region, quite different spectral features are observed. In addition to the absence of the Sr d peak, there is an appearance of a large spectral weight between 538 and 543 eV (position C), which is not found on the STO side. This peak turns out to be derived from the interaction with the nearest-neighbor O along the c axis as shown in the inset of Fig. 9a. Furthermore, this spectral peak originally emerges at the interface, as shown in the middle of Fig. 9b. When two oxide slabs are brought into contact, chemical bonds are created that result in charge redistribution in conjunction with a structural distortion at the interface. Note that half of O at the "STO surface" is bonded to Ti of anatase, leading to three-fold coordinated O ($O_{3\text{-fold}}$, site A), while the other half of O remains two-fold coordinated ($O_{2\text{-fold}}$, site B). Ti at the STO surface is also bonded to O of anatase at site C. By comparing with the 2p pDOS at the $O_{3\text{-fold}}$ site of the interfacial layer (site A in Fig. 9a), this peak at the interface is identified as the result of the chemical bond formation between anatase TiO_2 and STO. The excellent agreement between the EELS spectrum and the theoretical electronic structure of the interface provides a strong verification of the interface model.

Core-level x-ray photoemission spectroscopy (CXPS) combined with first-principles calculations is another great tool for validation of an atomic model of

a heterostructure (Hufner 2003). When x-ray photons with sufficiently large energy are injected into a sample, they can excite several core-level electrons above the vacuum level. The excess energy of the photo-ejected electrons is transformed into kinetic energy, thus carrying information of the binding energy of the core electrons. The core levels are highly sensitive to the local electronic environment mainly through electrostatic effects and in turn, to the local atomic environment. Therefore, a combination of electronic structure calculation and core level shift measurement can provide a tool to probe the structural information of an interface at atomic scale.

Figure 10 compares the Si 2p core level shifts (CLS) of the bare (2x1) Si(001) surface with those of the Si(001) 2x1 surface in the presence of ½-ML Sr layer on top (Choi et al. 2013a). The notable difference between the two structures is the flattening of the Si dimers on the surface in the presence of the interface with the ½-ML Sr layer. When adsorbed on the Si(001) surface, Sr transfers two electrons to a Si dimer, thus removing the need of Si dimer buckling. In the Si 2p CLS of the Si(001) surface, one can observe two dimer-atom-related peaks labeled as Su and Sd, which are positioned at -0.43 eV and 0.12 eV, respectively, with respect to the Si 2p peak derived from the bulk Si. At the interface of the ½-ML Sr/Si(001) structure, only one dimer-atom-related peak is observed at -0.33 eV with respect to the bulk peak.

In Choi et al. (2013a), the CLS spectrum is also calculated theoretically using the first-principles final state theory based on density functional theory. They performed CLS calculations using the VASP code (Kresse and Furthmüller 1996) at LDA level of theory with the Ceperley-Alder data parametrized by Perdew and Zunger (1981). The PAW pseudopotentials (Blöchl 1994) were used to model Si and Sr along a plane-wave cutoff energy of 500 eV. The Si surface was modeled using a 16-layer thick slab, which was separated from its periodic images in the c-direction by 15 Å of vacuum. They considered the Si(001) p(2×2) surface and the (2×1) surface of 1/2 monolayer (ML) Sr on Si(001). They treated both surfaces of the slab as active surfaces and no hydrogen passivation was used.

When a core electron is excited by a photon with an energy of $h\nu$ and emitted from a sample with a kinetic energy, E_{kin}, the binding energy, E_B of the electron is defined by

$$E_B = h\nu - E_{kin} - \varphi = E\left(n_c - 1\right) - E\left(n_c\right),$$

where φ is the work function, $E(n_c)$ and $E'(n_c-1)$ are the system's ground and excited state energies, respectively, with n_c being the number of core electrons. The quantities on the left-hand side can be determined experimentally, and the total energies on the right-hand side are the subject of theoretical calculations. A simple theory to calculate the CLS is based on the Koopmans' theorem in the Hartree-Fock theory, stating that the binding energy is the same as the orbital energy under the frozen orbital approximation (Hufner 2003). Thus, in this theory, the CLS is the same as the shifts of the orbital energy in the ground state due to the change of the local electrostatic potential. This contribution to the CLS is known as the initial state effect. In addition, when a core hole is created, the system tends to screen it
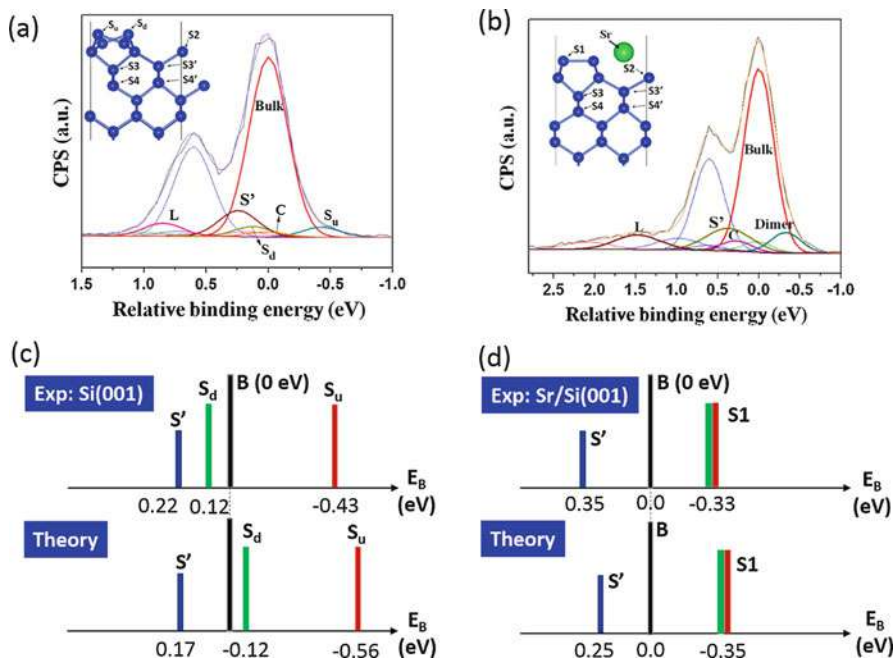
Fig. 10 (**a**, **b**) Si 2p core-level spectrum of clean 2x1 Si(001) (**a**) and ½-ML Sr/Si(001) (**b**) by in situ XPS at room temperature. The energy positions of all components are expressed in terms of the binding energy relative to the binding energy of the well-resolved bulk component. The insets in (**a**) and (**b**) are theoretical structures of Si(001) and ½-ML Sr/Si(001), respectively, obtained from the DFT calculations. Si atoms in the top four layers relax significantly and symmetrically different atoms are labeled accordingly. (**c**, **d**) Experimental (top) and theoretical (bottom) peak positions in the CLS spectrum of Si(001) (**c**) and ½-ML Sr/Si(001) (**d**). (With permission from Choi et al. (2013a))

(the final state effect) as much as possible, providing a relaxation energy gain for the emitted electron and shifting the core level binding energy toward lower binding energy from the initial state result.

Figure 10c and d compare the experimental and theoretical CLS of the bare Si(001) surface and the ½-ML Sr/Si(001) heterostructure, respectively. We observe that the final state theory calculations well reproduce the dimer-related peak positions found in experiment for both cases. For the bare Si(001) surface, the up-atoms of Si surface dimers create a peak at -0.43 eV in experiment, which is well compared to the theoretical value of -0.56 eV. For the ½-ML Sr/Si(001) structure, the position of the dimer peak was theoretically calculated to be at -0.35 eV with respect to the bulk 2p peak, which is in excellent agreement with the experimental value of -0.33 eV. The agreement between theory and experiment strongly supports the atomic model of the Si(001)'s structural reconstruction at the interface of the ½-ML Sr/Si(001) heterostructure.

X-ray CLS is also a powerful tool to probe a buried interface (Himpsel et al. 1988). One of the most studied oxide-based interface systems for applications of CXPS was the Si/SiO_2 interface due to its paramount importance in silicon technology. For the Si/SiO_2 interface, Si 2p core level shifts were mainly used to probe the chemical and structural properties of the interface (Himspel et al. 1988). Himpsel et al. used CXPS to suggest that there is a transition region, in which three distinct intermediate oxidation states of Si are present at the Si/SiO_2 interface. The oxidation state of the atom is heavily dependent on the local O structural environment. Using this property, Himpsel et al. suggested interface models based on the density and distribution of intermediate-oxidation states. However, the underlying structural information obtained solely from the CXPS is indirect and its verification requires a reliable theoretical calculation. Later, Pasquarello et al. used density functional theory to reproduce the important features observed in the previous XPS experiments (Pasquarello et al. 1996). In addition, they combined the DFT-based CLS calculations and the experimental results to suggest and validate structural models of the Si/SiO_2 interface.

One can also use optical spectroscopy to validate interfacial models. Unlike electron microscopy, it is nondestructive and not nearly as expensive. Both IR and Raman spectroscopies are useful in probing interface or defect-related bonds. For example, understanding the atomistic details of how platinum surfaces are oxidized under electrochemical conditions is of importance for many electrochemical devices such as fuel cells and electrolysers. Recently, Huang and co-authors examined the intermediate stages of electrochemical oxidation of single-crystalline platinum using a combination of theoretical modeling and in situ Raman spectroscopy (Huang et al. 2016). The authors used in situ shell-isolated nanoparticle-enhanced Raman spectroscopy to identify the intermediate stages of the electrochemical oxidation of Pt(111) and Pt(100) single crystals in perchloric acid. DFT calculations were used to determine the vibrational frequencies of possible intermediates and products. These in turn were used to assign the experimental Raman bands. The perchlorate anion was suggested to interact with a hydroxyl phase formed on the surface. Peroxo-like and superoxo-like two-dimensional (2D) surface oxides and amorphous 3D a-PtO_2 are sequentially formed during the anodic polarization. The work provided evidence for the structure-sensitive formation of a 2D platinum-(su)peroxide phase and provided a fundamental understanding of the mechanism of degradation of platinum electro-catalysts.

Even more recently, the effect of strain and oxygen deficiency on the Raman spectrum of monoclinic HfO_2 films that serve as gate dielectrics in modern FETs was investigated using first principles calculations (Gao et al. 2017). The results provided a more detailed understanding of vibrational modes in HfO_2, which besides technological applications in transistors is important in resistive random-access memory (RRAM) whose operation relies on oxygen-deficient HfO_x. The simulations of strained oxygen deficient films were compared with experimental results of Raman frequencies for 50 nm HfO_2 thin films. It was possible to identify a localized vibrational mode at $96.4 \ cm^{-1}$ as being introduced by threefold-coordinated oxygen vacancies.

6   Band-Alignment at Oxide Interfaces from First Principles

One of the most important interface properties of an oxide heterostructure is the band offset at the interface. The band offset plays a key role in determining the carrier confinement and electronic transport along and across the interface. The discovery of novel electronic phases emerging at oxide interfaces increased interest in the band offsets of oxide heterostructures (Ohtomo and Hwang 2004; Mannhart et al. 2008). Furthermore, the understanding of band offsets is crucial for many renewable energy applications such as heterogeneous photo-catalyst for water splitting (Chen et al. 2010). However, the band offsets at oxide heterostructures are not as well understood as those between semiconductors and metals. The first step to understand the band offset of an oxide heterostructure is to understand the interfacial chemistry and atomic structure at the interface. Once a representative interfacial atomic model is obtained, one could calculate its band alignment using first-principles methods as we describe below.

In Sect. 5, we discussed an atomic model of the TiO_2/$SrTiO_3$ heterointerface, which was validated by using a combination of the STEM/EELS measurements and the DFT-based theoretical spectroscopic calculations. With this atomic model, one could proceed to calculate the band alignment of the heterostructure in two different ways using density functional theory. The first method is to use layer-by-layer projected densities of states (pDOS). Figure 11a shows the layer-by-layer pDOS of the TiO_2/STO(001) heterostructure, for which the total valence band DOS is projected to each [001] atomic plane: SrO or TiO_2 layer for $SrTiO_3$ and TiO_2 layer for anatase TiO_2. The description of the calculation setups is given in Sect. 5. To extract the valence band offset, one can compare the valence band edge on the STO side (green line in Fig. 11) to that on the TiO_2 bulk region, which was determined to be 0.7 eV.

The second method of calculating the band offset at an oxide interface is to use an averaged electrostatic potential across the interface and those of the constituent bulk materials. Figure 11b shows the local electrostatic potential profile across the TiO_2/STO(001) interface, which was obtained by calculating the planar average of the local electrostatic potential and plotting it as a function of distance in the direction normal to the TiO_2/STO(001) interface. Here, "local" in the local electrostatic potential means that we only use the sum of Hartree and ionic potentials. After calculating the oscillating potential profile, one can macroscopically average the potential in the direction normal to the [001] interface plane in order to obtain the reference potential levels at the STO and TiO_2 bulk regions (black lines in Fig. 11b). Then, the next step is to position the valence band edge with respect to this reference potential level on both the STO and TiO_2 side. For this step, one can perform two separate calculations with the STO and TiO_2 bulk unit cells, in which one can directly access the position of the bulk valence band edge with respect to the average of the local electrostatic potential. Using this information, one can determine the valence band edge position on the STO and TiO_2 bulk regions of the supercell (green lines in Fig. 11b). The offset between the green lines is the band offset of the heterostructure, which is 0.76 eV. Note that this result is consistent with the result obtained using the layer-by-layer pDOS in Fig. 11a.
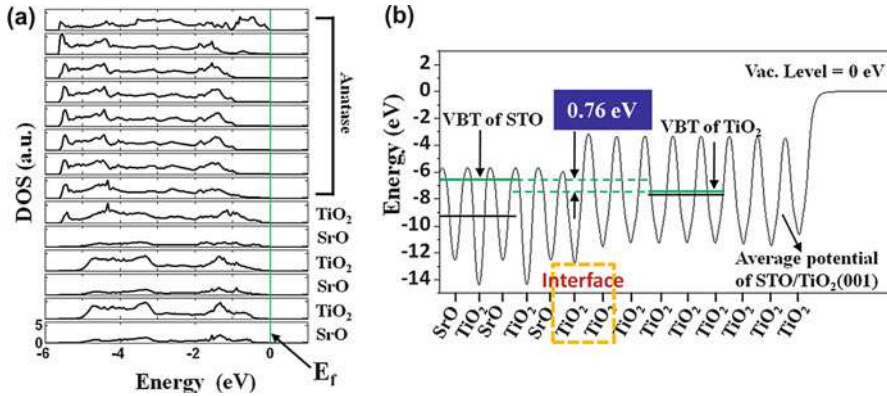
Fig. 11 (**a**) Layer-by-layer projected density of states (pDOS) of the $TiO_2/SrTiO_3(001)$ heterostructure. The green line is the Fermi level, which is determined by the valence band edge of the $SrTiO_3$ bulk (**b**) planar-averaged electrostatic potential of the heterostructure along the (001) direction. The straight black lines indicate the reference electrostatic energy positions with respect to the vacuum level (0 eV) in the bulk region of STO and TiO_2, respectively. The green lines indicate the relative positions of valence band edges of STO and TiO_2 with respect to their corresponding reference energy positions. (With permission from Seo et al. (2012b))

The formation of an interface and its effect on its band offset can be better understood by considering the Schottky-limit band-offset, which is an offset in a situation that no interaction is present between the constituent materials: TiO_2 and STO slabs in the case of TiO_2/STO(001) interface. Then, the valence band offset of the heterostructure can be written as $\Delta E_H = \Delta E_S + \Delta V$, where ΔE_H is the band offset of the actual heterostructure, ΔE_S is the band offset in the Schottky limit, and ΔV is the electrostatic potential drop at the interface due to the heterojunction formation. Therefore, ΔV includes all the important chemical and structural effects due to the formation of the interface. Using the same technique described above, Seo, et al. calculated the Schottky-limit valence band offset of the TiO_2/STO(001) interface to be 0.95 eV, thus ΔV is -0.2 eV (Seo et al. 2012b). The origin of ΔV is often described as the creation of an interface dipole or double layer involving charge transfer and subsequent dielectric screening at the interface.

Another important property that can be obtained from interface calculations is the presence and concentration of conducting states at the interface. These states can either be present due to interfacial states (Fredrickson and Demkov 2016), defects (Kalabukhov et al. 2007), electronic reconstruction like the polar catastrophe as in $LaAlO_3/SrTiO_3$ (Ohtomo and Hwang 2004; Lee and Demkov 2008), or due to charge transfer from ferroelectric films and interfaces (Fredrickson and Demkov 2015; Yin et al. 2015). By examining the pDOS of atoms at the interface, one can see if there are states near the Fermi level that do not exist in the bulk of either the film or the substrate. The atomic and orbital character of these states can offer more clues to the source of the interfacial states.

In certain cases, construction of a realistic interface may prove difficult due to factors such as lateral size, undesired artificial strain, or functional choice. An

alternative to the above described methods can be used to obtain band offsets (and also the relative defect alignments for point defects) between the two materials in question by alignment of the charge neutrality levels of the materials (Flores and Tejedor 1979; Robertson 2000; Tejedor et al. 1977; Tejedor and Flores 1978). There are a variety of ways to do this including calculation of the zero of the Green's function of the density of states (Robertson 2000; Tersoff 1984a, b):

$$G\left(E_0\right) = \int \frac{N\left(E'\right)dE'}{E - E'} = 0$$

However, a more straightforward method for estimating the charge neutrality level can be performed by calculating the $\varepsilon_H(+/-)$ for interstitial hydrogen (Van de Walle and Neugebauer 2003) using the Zhang-Northrup formalism for calculating defect energy levels in bulk materials (Zhang and Northrup 1991):

$$E_{form}\left(H_i^q\right) = E_{tot}\left(H_i^q\right) - E_{tot}(bulk) + \mu_H + q\left(E_F + \epsilon_V + \Delta V\right)$$

Essentially, a reasonably sized (i.e., well converged, but computationally tractable) supercell is constructed with an interstitial hydrogen atom. The atomic position of the hydrogen is then optimized in the H^+ state and the H^- charged states and the formation energy plotted as a function of the Fermi energy between the valence and conduction bands. Since we only care about the crossing point and not absolute energies, the potential correction (ΔV) can in the present case be neglected as the correction is usually similar for oppositely charged defects of the same type (this is not true generally in defect calculations). The point at which the formation energy for interstitial H^+ and H^- cross is the $\varepsilon_H(+/-)$ level and provides a surface independent estimate for the charge neutrality level. As the charge neutrality level in each material is given relative to the valence and conduction bands of the materials, the relative valence and conduction band offsets can then be determined. Furthermore, since this methodology is essentially an extension of the calculation of point defect transition energy levels, it can provide an easy extension from simply calculating formation energy of vacancies, interstitials, or dopants in oxides to explain their roles as electron or hole "traps" when in contact with semiconductors. In Fig. 12, this process is illustrated for the alignment of dielectric oxide HfO_2 and semiconductor black phosphorus (BP) using hybrid functional calculations as performed in (Liang et al. 2018). The calculations therein were performed using the VASP code (Kresse and Furthmüller 1996) with the HSE06 hybrid functional (Heyd et al. 2003, 2006), PAW pseudopotentials (Blöchl 1994), and a plane-wave cutoff of 400 eV. The hydrogen impurity was placed in a 2x2x2 supercell of HfO_2 and 3x2x1 supercell of BP with experimental lattice constants and optimized in both the H^+ and H^- state by modifying the supercell electron count. We point out that the value obtained for the CNL in BP is similar to that obtained slab models (Cai et al. 2014). Additionally, this method has been
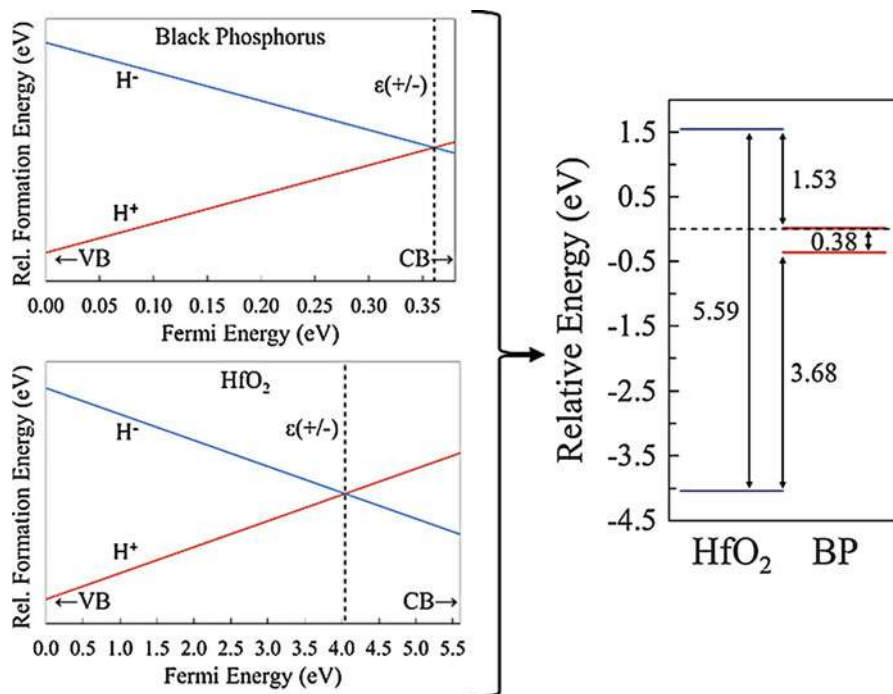
Fig. 12 Calculation of the $\varepsilon_H(+/-)$ level for black phosphorus (top left) and HfO$_2$ (bottom left) and their relative band alignments generated by aligning this level (right). (Adapted from Liang et al. (2018))

used to predict and validate band alignment and relative defect levels of HfO$_2$ and ZrO$_2$ to Si (Lyons et al. 2011), LaAlO$_3$ to Si and III-V semiconductors (Choi et al. 2013b), and chalcogenide semiconductors with ZnO (Varley and Lordi 2013), among others.

While relatively simple and potentially less computationally costly than the local potential method, there are also drawbacks to this method. Principally, the method can only provide band alignments and relative defect alignments for defects that would not be immediately at the interface. By neglecting the specific interface construction, band bending is also neglected. Such methods are thus more suitable to amorphous or polycrystalline interfaces than to high quality epitaxial interfaces or computationally challenging interfaces due to lattice mismatch. Additionally, it neglects the ability to theoretically study the formation of the 2DEG formed in cases like the SrTiO$_3$/LaAlO$_3$ interface (Lee and Demkov 2008) or atomic migration across interfaces like in the case of interfacial hydrogen shuttling (Marinopoulous et al. 2007) or oxygen gettering in resistive RAM applications (O'Hara et al. 2014). Therefore, the decision as to which method is most useful should be made based on the nature of the questions being investigated.

7   Conclusions

From this brief overview, we hope the reader has gained a better understanding of how one can model interfaces involving oxide materials and what kind of information can be obtained from such models. In view of their wide ranging technological applications, a detailed understanding of the physical properties of these interfaces is very important. At the same time, the complexity of the materials involved makes the study of oxide interfaces a highly challenging scientific endeavor. First principles methods have been used with great success in describing and predicting the properties of oxide interfaces, assuming the interface structure is known, but much is left to do. The time and length scales accessible with these first principles methods are rather limited and multiscale computational schemes are often needed for the prediction of many properties. In addition, the approximations made to make the *ab initio* calculations tractable often are inadequate in the context of oxides and specifically the oxide interfaces. Electron correlation dominates the properties of many oxides, particularly those containing transition metals of the iron group (3d metals). As the more accurate quantum mechanical techniques are much more expensive computationally, much of the work relies on semi-empirical methods such as LDA + U. As a result, the predictive power of this approach is somewhat limited. The local approximation causes severe underestimation of the band gap and may cause problems with unphysical charge transfer at the interface if the empty states of the oxide become available for the electrons of the metal because they are positioned too low in energy. But this field of study is extremely active and there is no doubt that in the not so distant future, better computers will allow the use of more accurate theoretical methods for systems large enough to be technologically relevant. The continuous improvement in computing technology as well as techniques to more precisely characterize interface structure and properties will ensure that atomistic modeling of oxide interfaces will in the near future be capable of much more accurate predictions, enabling one to fabricate designer heterostructures with properties tailored for specific applications.

References

Bengtsson L (1999) Dipole correction for surface supercell calculations. Phys Rev B 59:12301–12304. https://doi.org/10.1103/PhysRevB.59.12301

Björneholm O, Hansen MH, Hodgson A, Liu L-M, Limmer DT, Michaelides A, Rossmeisl J, Shen H, Tocci G, Tyrode E, Walz M-M, Werner J, Bluhm H (2016) Water at interfaces. Chem Rev 116:7698–7726. https://doi.org/10.1021/acs.chemrev.6b00045

Blöchl P (1994) Projector augmented-wave method. Phys Rev B 50:17953–17979. https://doi.org/10.1103/PhysRevB.50.17953

Born M, Oppenheimer R (1927) Zur quantentheorie der molekeln. Ann Phys 20:457–484. https://doi.org/10.1002/andp.19273892002

Boys SF (1950) Electronic wave function I. A general method of calculation for the stationary states of any molecular system. Proc R Soc Lond A200:542–554. https://doi.org/10.1098/rspa.1950.0036

Cai Y, Zhang G, Zhang Y-W (2014) Layer-dependent band alignment and work function of few-layer phosphorene. Sci Rep 4:6677. https://doi.org/10.1038/srep06677

Carr RG, Somorjai GA (1981) Hydrogen production from photolysis of steam adsorbed onto platinized $SrTiO_3$. Nature 290:576–577. https://doi.org/10.1038/290576a0

Ceperley DM, Alder BJ (1980) Ground state of the electron gas by a stochastic method. Phys Rev Lett 45:566–569. https://doi.org/10.1103/PhysRevLett.45.566

Chakhalian J, Freeland JW, Millis AJ, Panagopoulos C, Rondinelli JM (2014) Emergent properties in plane view: strong correlations at oxide interfaces. Rev Mod Phys 86:1189–1202. https://doi.org/10.1103/RevModPhys.86.1189

Chambers SA (2010) Epitaxial growth and properties of doped transition metal and complex oxide filmes. Adv Mater 22:219–248. https://doi.org/10.1002/adma.200901867

Chelikowsky JR, Troullier N, Saad Y (1994a) Finite-difference-pseudopotential method: electronic structure calculations without a basis. Phys Rev Lett 72:1240–1243. https://doi.org/10.1103/PhysRevLett.72.1240

Chelikowsky JR, Troullier N, Wu K, Saad Y (1994b) Higher-order finite-difference pseudopotential method: an application to diatomic molecules. Phys Rev B 50:11355–11364. https://doi.org/10.1103/PhysRevB.50.11355

Chen X, Shen S, Guo L, Mao SS (2010) Semiconductor-based photocatalytic hydrogen generation. Chem Rev 110:6503–6570. https://doi.org/10.1021/cr1001645

Choi M, Posadas AB, Seo H, Hatch RC, Demkov AA (2013a) Charge transfer in Sr Zintl template on Si(001). Appl Phys Lett 102:031604. https://doi.org/10.1063/1.4788916

Choi M, Janotti A, Van de Walle CG (2013b) Native point defects in $LaAlO_3$: a hybrid functional study. Phys Rev B 88:214117. https://doi.org/10.1103/PhysRevB.88.214117

Demkov AA (2006) Thermodynamic stability and band alignment at a metal high-k dielectric interface. Phys Rev B 74:085310. https://doi.org/10.1103/PhysRevB.74.085310

Dreizler RM, Gross EKU (1990) Density functional theory: an approach to the quantum many-body problem. Springer, Berlin Heidelberg

Dudarev SL, Button GA, Savrasov SY, Humphreys CJ, Sutton AP (1998) Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA+U study. Phys Rev B 57:1505–1509. https://doi.org/10.1103/PhysRevB.57.1505

Engel E, Dreizler RM (2011) Density functional theory: an advanced course. Springer, New York

Fiolhais C, Nogueira F, Marques M (eds) (2003) A primer in density functional theory. Springer, Berlin Heidelberg

Flores F, Tejedor C (1979) Energy barriers and interface states at heterojunctions. J Phys C Solid State 12:731–749. https://doi.org/10.1088/0022-3719/12/4/018

Fredrickson KD, Demkov AA (2015) Switchable conductivity at the ferroelectric interface: nonpolar oxides. Phys Rev B 91:115126. https://doi.org/10.1103/PhysRevB.91.115126

Fredrickson KD, Demkov AA (2016) Spin-polarized, orbital-selected hole gas at the EuO/Pt interface. J Appl Phys 119:095309. https://doi.org/10.1063/1.4942837

Fredrickson KD, Posadas AB, Demkov AA, Dubourdieu C, Bruley J (2013) Wetting at the $BaTiO_3$/Pt interface. J Appl Phys 113:184102. https://doi.org/10.1063/1.4803705

Fredrickson KD, Ponath P, Posadas AB, McCartney MR, Aoki T, Smith DJ, Demkov AA (2014) Atomic and electronic structure of the ferroelectric $BaTiO_3$/Ge(001) interface. Appl Phys Lett 104:242908. https://doi.org/10.1063/1.4883883

Fuchs D, Pinta C, Schwartz T, Schweiss P, Nagel P, Schuppler S, Schneider R, Merz M, Roth G, Löhneysen H (2007a) Ferromagnetic order in epitaxially strained $LaCoO_3$ thin films. Phys Rev B 75:144402. https://doi.org/10.1103/PhysRevB.75.144402

Fuchs D, Arac E, Pinta C, Schuppler S, Schneider R, Löhneysen H (2008) Tuning the magnetic properties of LaCoO$_3$ thin films by epitaxial strain. Phys Rev B 77:014434. https://doi.org/10.1103/PhysRevB.77.014434

Fuchs F, Furthmüller J, Bechstedt F, Shishkin M, Kresse G (2007b) Quasiparticle band structure based on a generalized Kohn-Sham scheme. Phys Rev B 76:115109. https://doi.org/10.1103/PhysRevB.76.115109

Gao L, Yalon E, Chew AR, Deshmukh S, Salleo A, Pop E, Demkov AA (2017) Effect of oxygen vacancies and strain on the phonon spectrum of HfO$_2$ thin films. J Appl Phys 121:224101. https://doi.org/10.1063/1.4984833

Gazquez J, Bose S, Sharma M, Torija MA, Pennycook SJ, Leighton C, Varela M (2013) Lattice mismatch accommodation via oxygen vacancy ordering in epitaxial La$_{0.5}$Sr$_{0.5}$CoO$_{3-\delta}$ thin films. APL Mater 1:012105. https://doi.org/10.1063/1.4809547

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502. https://doi.org/10.1088/0953-8984/21/39/395502

Henrich VE, Cox PA (1994) The surface science of metal oxides. Cambridge University Press, Cambridge, UK

Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened Coulomb potential. J Chem Phys 118:8207–8215. https://doi.org/10.1063/1.1564060

Heyd J, Scuseria GE, Ernzerhof M (2006) Erratum: "Hybrid functionals based on a screened Coulomb potential" [J. Chem. Phys. 118, 8207 (2003)]. J Chem Phys 124:219906. https://doi.org/10.1063/1.2204597

Himpsel FJ, McFeely FR, Taleb-Ibrahimi A, Yarmoff JA, Hollinger G (1988) Microscopic structure of the SiO$_2$/Si interface. Phys Rev B 38:6084–6096. https://doi.org/10.1103/PhysRevB.38.6084

Huang Y-F, Kooyman PJ, Koper MTM (2016) Intermediate stages of electrochemical oxidation of single-crystalline platinum revealed by *in situ* Raman spectroscopy. Nat Commun 7:12440. https://doi.org/10.1038/ncomms12440

Hüfner S (2003) Photoelectron spectroscopy: principles and applications. Springer, Berlin Heidelberg

Hwang HY, Iwasa Y, Kawasaki M, Keimer B, Nagaosa N, Tokura Y (2012) Emergent phenomena at oxide interfaces. Nat Mater 11:103–113. https://doi.org/10.1038/nmat3223

Imada M, Fujimori A, Tokura Y (1998) Metal-insulator transitions. Rev Mod Phys 70:1039–1263. https://doi.org/10.1103/RevModPhys.70.1039

Ismail-Beigi S, Walker FJ, Disa AS, Rabe KM, Ahn CH (2017) Picoscale materials engineering. Nat Rev Mater 2:17060. https://doi.org/10.1038/natrevmats.2017.60

Kalabukhov A, Gunnarsson R, Börjesson J (2007) Effect of oxygen vacancies in the SrTiO$_3$ substrate on the electrical properties of the LaAlO$_3$/SrTiO$_3$ interface. Phys Rev B 75:121404. https://doi.org/10.1103/PhysRevB.75.121404

Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186. https://doi.org/10.1103/PhysRevB.54.11169

Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. Phys Rev B 59:1758–1775. https://doi.org/10.1103/PhysRevB.59.1758

Krukau AV, Vydrov OA, Izmaylov AF, Scuseria GE (2006) Influence of the exchange screening parameter on the performance of screened hybrid functionals. J Chem Phys 125:224106. https://doi.org/10.1063/1.2404663

Lee J, Demkov AA (2008) Charge origin and localization at the n-type SrTiO$_3$/LaAlO$_3$ interface. Phys Rev B 78:193104. https://doi.org/10.1103/PhysRevB.78.193104

Leon B, Janson JA (eds) (2009) Thin calcium phosphate coatings for medical implants. Springer, New York

Liang CD, Ma R, Su Y, O'Hara A, Zhang EX, Alles ML, Zhao SE, Pantelides ST, Koester J, Schrimpf RD, Fleetwood DM (2018) Defects and low-frequency noise in irradiated black phosphorus MOSFETs with HfO_2 gate dielectrics. IEEE Trans Nucl Sci 65:1227–1238. https://doi.org/10.1109/TNS.2018.2828080

Lyons J, Janotti A, Van de Walle C (2011) The role of oxygen-related defects and hydrogen impurities in HfO_2 and ZrO_2. Microelectron Eng 88:1452–1456. https://doi.org/10.1016/j.mee.2011.03.099

Makov G, Payne MC (1995) Periodic boundary conditions in *ab initio* calculations. Phys Rev B 51:4014–4022. https://doi.org/10.1103/PhysRevB.51.4014

Mannhart J, Blank DHA, Hwang HY, Millis AJ, Triscone J-M (2008) Two-dimensional electron gases at oxide interfaces. MRS Bull 33:1027–1034. https://doi.org/10.1557/mrs2008.222

Marinopoulous AG, Batyrev I, Zhou XJ, Shcrimpf RD, Fleetwood DM, Pantelides ST (2007) Hydrogen shuttling near Hf-defect complexes in $Si/SiO_2/HfO_2$ structures. Appl Phys Lett 91:233503. https://doi.org/10.1063/1.2820380

Martin RR (2004) Electronic structure: basic theory and practical methods. Cambridge University Press, New York

Monkhorst HJ, Pack JD (1976) Special points for Brillouin-zone integrations. Phys Rev B 13:5188–5192. https://doi.org/10.1103/PhysRevB.13.5188

O'Hara A, Bersuker G, Demkov AA (2014) Assessing hafnium on hafnia as an oxygen getter. J Appl Phys 115:183703. https://doi.org/10.1063/1.4876262

Ohtomo A, Hwang HY (2004) A high-mobility electron gas at the $LaAlO_3/SrTiO_3$ heterointerface. Nature 427:423–426. https://doi.org/10.1038/nature02308

Paracchino A, Laporte V, Sivula K, Grätzel M, Thimsen E (2011) Highly active oxide photocathode for photoelectrochemical water reduction. Nat Mater 10:456–461. https://doi.org/10.1038/nmat3017

Pasquarello A, Hybertsen MS, Car R (1996) Theory of Si 2p core-level shifts at the Si(001)-SiO_2 interface. Phys Rev B 53:10942–10950. https://doi.org/10.1103/PhysRevB.53.10942

Pennycook SJ, Nellist PD (eds) (2011) Scanning transmission electron microscopy: imaging and analysis. Springer, New York

Perdew JP, Zunger A (1981) Self-interaction correction to density-functional approximations for many-electron systems. Phys Rev B 23:5048–5079. https://doi.org/10.1103/PhysRevB.23.5048

Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) Atoms, molecules solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. Phys Rev B 46:6671–6687. https://doi.org/10.1103/PhysRevB.46.6671

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865

Pham TA, Ping Y, Galli G (2017) Modelling heterogeneous interfaces for solar water splitting. Nat Mater 16:401–408. https://doi.org/10.1038/nmat4803

Ponath P, Fredrickson K, Posadas AB, Ren Y, Wu X, Vasudevan RK, Okatan BO, Jesse S, Aoki T, McCartney MR, Smith DJ, Kalinin SV, Lai K, Demkov AA (2015) Carrier density modulation in a germanium heterostructure by ferroelectric switching. Nat Commun 6:6067–6074. https://doi.org/10.1038/ncomms7067

Qiao L, Droubay TC, Shutthanandan V, Zhu Z, Sushko PV, Chambers SA (2010) Thermodynamic instability at the stoichiometric $LaAlO_3/SrTiO_3$(001) interface. J Phys Condens Matter 22:312201. https://doi.org/10.1088/0953-8984/22/31/312201

Rappe AM, Rabe KM, Kaxiras E, Joannopoulos JD (1990) Optimized pseudopotentials. Phys Rev B 41:1227–1230. https://doi.org/10.1103/PhysRevB.41.1227

Robertson J (2000) Band offsets of wide-band-gap oxides and implications for future electronic devices. J Vac Sci Technol B 18:1785–1791. https://doi.org/10.1116/1.591472

Rondinelli JM, May SJ, Freeland JW (2012) Control of octahedral connectivity in perovskite oxide heterostructures: an emerging route to multifunctional materials discovery. MRS Bull 37:261–270. https://doi.org/10.1557/mrs.2012.49

Schneider W-D, Heyde M, Freund H-J (2018) Charge control in model catalysis: the decisive role of the oxide-nanoparticle interface. Chem Eur J 24:2317–2327. https://doi.org/10.1002/chem.201703169

Seo H, Demkov AA (2011) First-principles study of polar $LaAlO_3$ (001) surface stabilization by point defects. Phys Rev B 84:045440. https://doi.org/10.1103/PhysRevB.84.045440

Seo H, Demkov AA (2014) Band alignment in visible-light photo-active $CoO/SrTiO_3$ (001) heterostructures. J Appl Phys 116:245305. https://doi.org/10.1063/1.4905112

Seo H, Posadas A, Demkov AA (2012a) Strain-drive spin-state transition and superexchange interaction in $LaCoO_3$: *Ab intio* study. Phys Rev B 86:014430. https://doi.org/10.1103/PhysRevB.86.014430

Seo H, Posadas AB, Mitra C, Kvit AV, Ramdani J, Demkov AA (2012b) Band alignment and electronic structure of the anatase $TiO_2/SrTiO_3$(001) heterostructure integrated on Si(001). Phys Rev B 86:075301. https://doi.org/10.1103/PhysRevB.86.075301

Sharia O, Demkov AA, Bersuker G, Lee BH (2007) Theoretical study of the insulator/insulator interface: band alignment at the SiO_2/HfO_2 junction. Phys Rev B 75:035306. https://doi.org/10.1103/PhysRevB.75.035306

Shishkin M, Kresse G (2006) Implementation and performance of the frequency-dependent *GW* method within the PAW framework. Phys Rev B 74:35101. https://doi.org/10.1103/PhysRevB.74.035101

Shishkin M, Kresse G (2007) Self-consistent *GW* calculations for semiconductors and insulators. Phys Rev B 75:235102. https://doi.org/10.1103/PhysRevB.75.235102

Shishkin M, Marsman M, Kresse G (2007) Accurate quasiparticle spectra from self-consistent *GW* calculations with vertex corrections. Phys Rev Lett 99:246403. https://doi.org/10.1103/PhysRevLett.99.246403

Singh DJ, Nordstrom L (2006) Planewaves, pseudopotentials, and the LAPW method, 2nd edn. Springer, New York

Skriver H (1984) The LMTO method. Springer, Berlin Heidelberg

Slater JC (1930) Atomic shielding constants. Phys Rev 36:57–64. https://doi.org/10.1103/PhysRev.36.57

Slepko A, Demkov AA (2011) First-principles study of the biomineral hydroxyapatite. Phys Rev B 84:134108. https://doi.org/10.1103/PhysRevB.84.134108

Slepko A, Demkov AA (2013) First principles study of hydroxyapatite surface. J Chem Phys 139:044714. https://doi.org/10.1063/1.4813828

Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, Sánchez-Portal D (2002) The SIESTA method for ab initio order-N materials simulation. J Phys Condens Matter 14:2745. https://doi.org/10.1088/0953-8984/14/11/302

Tejedor C, Flores F (1978) A simple approach to heterojunctions. J Phys C Solid State Phys 11:L19–L23. https://doi.org/10.1088/0022-3719/11/1/005

Tejedor C, Flores F, Louis E (1977) The metal-semiconductor interface: Si (111) and zincblende (110) junctions. J Phys C Solid State Phys 10:2163–2177. https://doi.org/10.1088/0022-3719/10/12/022

Tersoff J (1984a) Schottky barrier heights and the continuum of gap states. Phys Rev Lett 52:465–468. https://doi.org/10.1103/PhysRevLett.52.465

Tersoff J (1984b) Theory of semiconductor heterojunctions: the role of quantum dipoles. Phys Rev B 30:4874–4877. https://doi.org/10.1103/PhysRevB.30.4874

Troullier N, Martins JL (1991) Efficient pseudopotentials for plane-wave calculations. Phys Rev B 43:1993–2006. https://doi.org/10.1103/PhysRevB.43.1993

Van de Walle CG, Neugebauer J (2003) Universal alignment of hydrogen levels in semiconductors, insulators and solutions. Nature 423:626–628. https://doi.org/10.1038/nature01665

Vanderbilt D (1990) Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Phys Rev B 41:7892–7895. https://doi.org/10.1103/PhysRevB.41.7892

Varela M, Gazquez J, Pennycook SJ (2012) STEM-EELS imaging of complex oxides and interfaces. MRS Bull 37:29–35. https://doi.org/10.1557/mrs.2011.330

Varley JB, Lordi V (2013) Electrical properties of point defects in CdS and ZnS. Appl Phys Lett 103:102103. https://doi.org/10.1063/1.4819492

Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Can J Phys 58:1200–1211. https://doi.org/10.1139/p80-159

Yin B, Aguado-Puente P, Qu S, Artacho E (2015) Two-dimensional electron gas at the PbTiO$_3$/SrTiO$_3$ interface: an *ab initio* study. Phys Rev B 92:115406. https://doi.org/10.1103/PhysRevB.92.115406

Zhang SB, Northrup JE (1991) Chemical potential dependence of defect formation energies in GaAs: application to Ga self-diffusion. Phys Rev Lett 67:2339–2342. https://doi.org/10.1103/PhysRevLett.67.2339

Design of New Multiferroic Oxides

48

Xue-Zeng Lu and James M. Rondinelli

Contents

Abstract

Multiferroics materials, displaying magnetic, polar, and elastic order parameters simultaneously, have garnered intense interests because they may host electric-field controllable magnetism, which can be utilized in devices for information storage and processing. In this chapter, based on recent progress in understanding microscopic mechanisms supporting multiferroism and advances in materials synthesis, we describe the basic routes by which to design new multiferroic oxides using electronic structure based calculations to guide materials selection and discovery.

X.-Z. Lu · J. M. Rondinelli (✉)
Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA
e-mail: jrondinelli@northwestern.edu

1   Introduction

Multiferroic materials, as defined in the early days of the field, simultaneously exhibit two or more of the three primary ferroic orders, including spontaneous ferroelectricity, ferromagnetism, and ferroelasticity in a single phase (Schmid 1994). Nowadays, the term is usually used to refer to any material with coexisting (anti)ferroelectricity and (anti)ferromagnetism. Uniting an electric polarization and magnetization in a single material makes these compounds ideally suited to explore magnetoelectric (ME) coupling, i.e., the coupling between the two ferroic orders (Khomskii 2006, 2009; Erenstein et al. 2006; Cheong and Mostovoy 2007; Ramesh and Spaldin 2007; Van den Brink and Khomskii 2008; Picozzi et al. 2008; Wang et al. 2009, 2016; Spaldin et al. 2010; Picozzi and Stroppa 2012; Johnson and Radaelli 2014; Lu et al. 2015a; Fiebig et al. 2016).

The multiferroic field and ME interactions can be traced back to the 1950s, when Landau and Lifshitz proposed the concept of magnetoelectric coupling whereby a material would exhibit a magnetization and electric polarization under the application of an external electric field and magnetic field, respectively, i.e., arising from a nonconjugate field (Landau and Lifshitz 1959). Soon after, Dzyaloshinskii (1960) predicted and Astrov (1960) observed ME coupling in the compound Cr_2O_3. In the early days, this field grew slowly (Ascher et al. 1966; Fox et al. 1980; DiDomenico et al. 1969; Kato et al. 1983; Ismailza and Kizhaev 1965; Teague et al. 1970), on the one hand, because the techniques producing high-quality multiferroics were not accessible and, on the other hand, the weak observed ME coupling did not attract much attention. The situation changed quickly after the development of multiple advanced thin film synthesis techniques, including pulsed laser deposition (PLD) and molecular beam epitaxy (MBE), and improved first-principles computational techniques. For example, $BiFeO_3$, which adopts the $R3c$ symmetry, was the first discovered room-temperature multiferroic. It exhibits ferroelectric (FE) and antiferromagnetic (AFM) transitions at 1103 K and 643 K, respectively (Teague et al. 1970; Sosnowska et al. 1982; Johnson et al. 2013). The polarization is along the [111] pseudo-cubic direction, and the AFM magnetic structure takes the form of a long-wavelength (62 nm) spiral. Early experiments showed that $BiFeO_3$ possesses a small polarization ($\sim$6.1μC/cm^2) in bulk (Kiselev et al. 1963); however, this value was significantly enhanced by almost an order of magnitude ($\sim$100μC/cm^2) in thin film by using a state-of-the-art technique, i.e., PLD, in 2003 (Wang et al. 2003), and the larger polarization value was demonstrated by electronic structure calculations on the $R3c$ structure utilizing the modern theory of polarization (Neaton et al. 2005).

Although transition metal perovskite compounds are among the most promising materials to search for new multiferroics, few were reported to be multiferroic prior to 2003. This is to be contrasted with the abundance of perovskite oxides exhibiting exclusively ferroelectricity or magnetism (Goodenough and Longo 1970; Mitsui et al. 1982). To explain the mutual exclusion between ferroelectricity and long-range magnetic order in transition metal perovskites, Hill proposed a "d^0 rule" in

a seminal paper reported in Hill (2000). It is paraphrased as follows: In most FE materials, the transition metal ion (e.g., represented by B in ABO_3) will be formally a d^0 ion because this electronic configuration can favor hybridization between the B cation antibonding d states that form the conduction band and the oxygen 2p states that form the valence band. Polar B cation displacements are stabilized by the second-order Jahn-Teller effect (Pearson 1975) and result in a spontaneous electric polarization as in $BaTiO_3$. In contrast, when the d shell is occupied (i.e., d^n, n > 0), the energetic gain provided by the SOJT is suppressed and eliminates the driving force for inversion symmetry lifting displacements. In addition, other factors may suppress the FE distortion, such as a first-order Jahn-Teller distortion in $LaMnO_3$ (Mn: $3d^4$ configuration) and $YTiO_3$ (Ti: $3d^1$ configuration) and the strong Hund's exchange energy between occupied t_{2g} and empty e_g orbitals, e.g., in $CaMnO_3$ and $RCrO_3$ (R is rare-earth ion) (Khomskii 2006). In addition, increased electronic Coulomb repulsion can also compete with the hybridization in bond formation leading to a centrosymmetric structure with d^n (n > 0) configuration. However, this d^0 principle can be circumvented when allowing for ferroelectricity and magnetism to arise from different origins; for instance, the magnetism can arise from the magnetic B cations that form one sublattice of the structure, while ferroelectricity can be induced by the A-site cation. This combination of properties arising from coexisting cation sublattices is the basic principle active in $BiFeO_3$ and $BiMnO_3$. In this case, polar Bi displaces arise from Bi-O hybridization due to stereochemically active $6s^2$ lone-pair electrons of the Bi cations (Seshadri and Hill 2001).

The multiferroic boom arrived in 2003 owing to both the discovery of a large electric polarization in $BiFeO_3$ and the discovery of another class of multiferroics in the manganite family. $TbMnO_3$ is an exciting material that violates the d^0 rule owing to spin-induced ferroelectricity (Kimura et al. 2003a). The spiral spin structure with modulation wave vector (0,0.28,0) (Kenzelmann et al. 2005) present in the orthorhombic *Pbnm* crystal structure permits an electric polarization along the c axis. In addition, a moderate magnetic field applied along the b direction can switch the polarization from the c axis to the a axis of the crystal. Since then, numerous multiferroics based on the transition metal perovskites were reported, and they may all generally be clustered into two groups as proposed by Khomskii (2009):

- *Type I Multiferroics*: The ferroelectricity and magnetism arise from the different sources, and thus magnetoelectricity is weak. The critical temperature of the spontaneous polarization is higher than that of the spontaneous magnetization, and the magnitude of the polarization is usually very large ($>1\mu C/cm^2$). The representative Type I multiferroic is $BiFeO_3$.
- *Type II Multiferroics*: The ferroelectricity is caused by complex (non)collinear spin order due to competing exchange interactions and therefore the magnetoelectric coupling is stronger; however, the magnitude of the polarization is usually very small ($<1\mu C/cm^2$). The representative material of this class is $TbMnO_3$.

In both multiferroic types, the critical temperature of the spontaneous magnetic order is usually lower than room temperature, and the magnetic order often leads to no macroscopic magnetization or a weak ferromagnetism.

Both types of the multiferroics have their own drawbacks, which preclude their direct application in new technologies, for example, those based on four-state memory devices and in the writing and reading of magnetic information with low power by exploiting electric field control of magnetism through the magnetoelectric response (Lu et al. 2015a; Fiebig et al. 2016). Beyond these single-phase multiferroic materials, researchers have devoted significant efforts to find alternatives solutions, for examples, artificial materials engineered with epitaxial strain, superlattices comprised of multiferroic constituents or non-multiferroic constituents, ferromagnetic-multiferroic heterostructures, ferromagnetic-ferroelectric interfaces, and domain walls. Recent reviews given in Khomskii (2006, 2009), Erenstein et al. (2006), Cheong and Mostovoy (2007), Ramesh and Spaldin (2007), Van den Brink and Khomskii (2008), Picozzi et al. (2008), Wang et al. (2009, 2016), Spaldin et al. (2010), Picozzi and Stroppa (2012), Johnson and Radaelli (2014), Lu et al. (2015a), and Fiebig et al. (2016) provide a complete description of advances made within these areas. More recently the search for single-phase multiferroics has been buoyed by the formation of new design strategies based on hybrid improper ferroelectrics (or broadly, geometric ferroelectricity) and asymmetric multiferroics.

In this chapter, we describe the design of multiferroic compounds according to the mechanisms permitting the combination of a spontaneous electric polarization and magnetism. The main mechanisms are lone-pair-induced ferroelectricity, charge-order-induced ferroelectricity, spin-driven ferroelectricity, geometric ferroelectricity, asymmetric multiferroics, polymorphic materials, superlattices, ferromagnetic-multiferroic heterostructures, ferromagnetic-ferroelectric interfaces, and domain walls. We conclude the chapter by briefly discussing frontiers and functionalities in multiferroics.

2 Strategy and Considerations for Multiferroics by Computational Design

When beginning to design a new multiferroic, the crystal structure and chemistry require assessment. It was proposed that any functional material is defined by three features, referred to as ACS (Rondinelli et al. 2015), and that property realization through materials-by-design relies on exploiting atomic identities involved (elemental species), composition (integer stoichiometry in compounds or continuous variable, "subscript engineering," in alloys and ceramics), and structure (i.e., where the atoms reside with regard to crystallographic structure). ACS features in this chapter are based on the transition metal oxides exhibiting or derived from the ABO_3 perovskite composition. Owing to the multiferroic behavior, the atomic species of the A and/or B cations must have partially filled d and/or f orbitals. Usually, B magnetic ions are selected to have open d shells, which is achieved by selecting among Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Nb, Mo, Tc, Ru, W, Os, etc.

No matter what is the mechanism inducing the simultaneous electric and magnetic polarization, the material must be insulating. For many transition metal oxides consisting of the aforementioned cations, the electrons in the d orbitals may be considered weakly or strongly correlated, which means that there will be strong repulsion when two electrons with opposite spins occupy the same orbital. This repulsion arises from the Coulomb interaction between the two electrons, which based on a single-site Hubbard model is referred to as the "Hubbard U" (Hubbard 1963). The strength of the U interaction is related to the insulating state and in the solid is screened by the dielectric constant of the material. In practice, this on-site repulsion U can be treated within the density functional theory (DFT) framework by including an on-site interaction for the correlated orbitals when investigating correlated compounds to correctly describe the insulating state (Anisimov et al. 1997, 1991; Solovyev et al. 1994). Note that the U correction depends on the type of the exchange-correlation functional and is not identical to the U that may be extracted from spectroscopic measurements.

To assess if a designed material is insulating, the band structure and density of states can be computed by carrying out DFT calculations. Although the size of the band gap may vary with U, an appropriate U value in the calculations can be identified by examining the robustness of critical physical properties (e.g., computed band gap, magnetic transition temperature, structural and magnetic structure parameters) with respect to U and comparing these results to available experiments on analogous phases. Inaccuracies in the magnitude of the band gap are well-known consequences of the absence of the derivative discontinuity in the exchange-correlation functionals, which are usually the local-density approximations (LDA) (Ceperley and Alder 1980; Vosko et al. 1980), generalized gradient approximations (GGA), and hybrid functionals. At the GGA level, the forms of Perdew-Burke-Ernzerhof (PBE) (Perdew et al. 1996) and Perdew-Burke-Ernzerhof revised for solids (PBEsol) (Perdew et al. 2008) are usually chosen for most studies. In hybrid calculations, the HSE06 functional is most frequently adopted (Paiera et al. 2006) owing to its reduced computational overhead without loss of accuracy. It also performs well for many magnetic oxides and multiferroic materials (Stroppa and Picozzi 2010). The LDA and GGA functionals usually underestimate the band gap, and HSE06 functional overestimates the band gap. Alternative approaches to obtain accurate band gap energies include the GW (or G_0W_0) method, which works well for semiconductor (Hybertsen and Louie 1986; Rohlfing and Louie 2000; Deslippe et al. 2012), and many transition metal oxides (Ergönenc et al. 2018). MetaGGA functionals include the local density (as in LDA) and the gradient of the density (as in GGA) but also the kinetic energy density (Sun et al. 2012, 2013, 2015, 2016). These functionals produce results often much more consistent with experiment and at times perform similarly to a hybrid functional but require less computational resources than hybrid functionals or the GW method. The exchange-correlation functional can also affect equilibrium crystal structures and hence lattice dynamics, electric polarizations, magnetic structure, etc. (Stroppa and Picozzi 2010; Krukau et al. 2006; Seixas et al. 2016; Diéguez and Iñiguez 2015). Thus the selection of a suitable exchange-correlation functional for the study of a known or predicted multiferroic material should be properly justified.

For a potential multiferroic compound, one can determine the equilibrium magnetic structure by using a spin Hamiltonian to model its ordered state, which can be expressed as:

$$\widehat{H}_{spin} = \sum\nolimits_{i<j} J_{ij} \vec{S}_i \cdot \vec{S}_i + \sum\nolimits_{i<j} \vec{D}_{ij} \cdot \left(\vec{S}_i \times \vec{S}_i \right) + \sum\nolimits_{i,\alpha\beta} A_{i,\alpha\beta} \vec{S}_{i\alpha} \vec{S}_{i\beta}$$

$$(1)$$

where J_{ij}, $\vec{D}_{ij}$, and $A_{i,\alpha\beta}$ are the symmetric spin exchange interaction parameters, antisymmetric Dzyaloshinskii-Moriya (DM) interaction vectors, and single-ion anisotropic parameters, respectively, and $\vec{S}_i$ is the spin vector. The DM interaction vector and single-ion anisotropic parameter are a consequence of spin-orbital interactions and in practice require DFT calculations that include relativistic spin-orbital coupling. All of the coefficients can be computed using the four-states mapping analysis and DFT energies (Xiang et al. 2011a, 2013a).

In a material with a simple cubic arrangement of magnetic cations with spin S, for example, one can obtain J_{12}, using four states: (I) $S_1 = S$ and $S_2 = S$; (II) $S_1 = S$ and $S_2 = -S$; (III) $S_1 = -S$ and $S_2 = S$; and (IV) $S_1 = -S$ and $S_2 = -S$. In the calculation, the other spins are kept fixed for each of the four states, and all spins are aligned along the z direction. Then, $J_{12} = \frac{E_I + E_{IV} - E_{II} - E_{III}}{4S^2}$, where E_i ($i = $ I–IV) is the energy for each state. For the collinear spin orders, there are four simple cases, which are FM, A-type, C-type, and G-type orders, respectively (see Fig. 1). After obtaining the coefficients in the spin Hamiltonian, one can also simulate the magnetic transition by performing Monte Carlo simulations based on the Hamiltonian.

To assess the ferroelectric properties, a DFT lattice dynamical analysis can be performed for proper FE materials within a high-symmetry structure (e.g., which may but not necessarily correspond to an observed high temperature paraelectric structure), in which the zone-center FE phonon mode is unstable and exhibits a double-well energy surface versus phonon mode amplitude (see Fig. 2). In addition, phonon calculations can also identify other unstable antiferrodistortive (AFD) lattice modes such as BO_6 oxygen octahedral-rotational modes in corner-shared ABO_3-based perovskites. Such instabilities typically occur at the Brillouin zone boundaries of the high-symmetry phase.

After condensing various combinations of the polar and the AFD phonon modes followed by structural relaxations using DFT, a final stable structure can be regarded as the ground state for the composition. The electric polarization in the ground state structure can be computed by either using the Berry phase method (Resta 1992, 1994; King-Smith and Vanderbilt 1993; Vanderbilt and King-Smith 1993) or by combing the Born effective charges with the ionic displacements found in the ground state structure with respect to a reference structure (i.e., paraelectric state) (Meyer and Vanderbilt 2002). In the Berry phase method, the total polarization (P) consists of electronic (P_e) and ionic (P_i) contributions (i.e., $P = P_e + P_i$), in which P is well-defined modulo $\frac{eR}{\Omega}$, where e, R, and Ω are the charge of the electron, a real-space lattice vector, and the unit cell volume, because of a
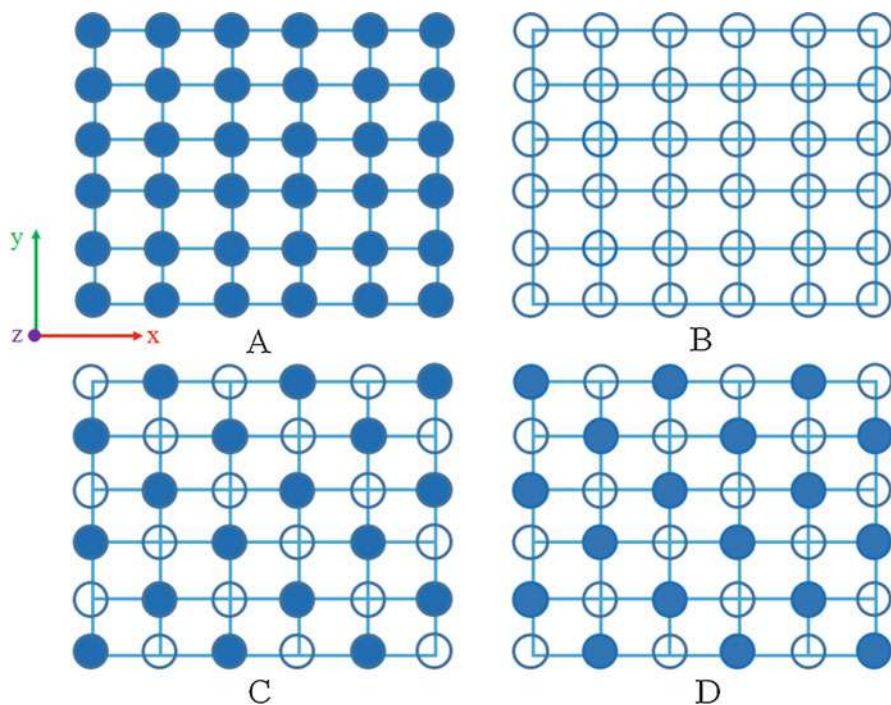
Fig. 1 (**a**, **b**) show the FM spin alignment in the *xy* plane; (**c**, **d**) show the checkboard AFM spin alignment in the *xy* plane, where solid circles indicate spin-up and open circles indicate spin-down. In the FM spin structure, there is an $\cdots$A-A-A-A$\cdots$ sequence along the *z* axis; in the A-type AFM spin structure, there is a $\cdots$B-A-B-A$\cdots$ sequence along the *z* axis; in the G-type AFM spin structure, there is a $\cdots$C-D-C-D$\cdots$ sequence along the *z* axis; and in the C-type AFM spin structure, there is a $\cdots$D-D-D-D$\cdots$ sequence along the *z* axis
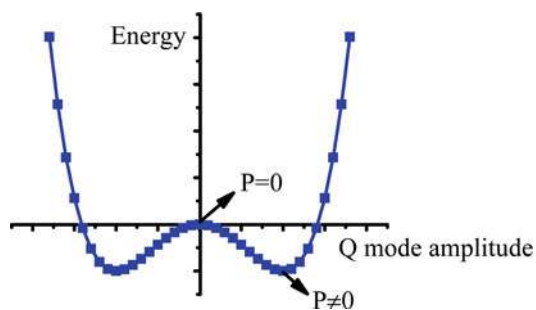
Fig. 2 The energy versus Q mode amplitude for a proper ferroelectric material. According to a Landau phenomenological model, the energy curve can be described by $E = \alpha Q^2 + \beta Q^4 + higher$ *even-order terms in Q*. When the polar (P) mode is unstable, then the coefficient α becomes negative and results in a decrease of the energy as shown by the double well with negative curvature about the origin (P $=$ 0)

gauge-phase factor in the Bloch function. Then the final polarization is obtained by subtracting any polarization obtained from the paraelectric phase using the same procedures. By using the Born effective charges, $P_\alpha = \frac{1}{\Omega} \sum_{i\beta} Z^*_{i\alpha\beta} \Delta u_{i\beta}$, where Z^*_i and Δu are the Born effective charge tensor and ionic displacement vector of ion i, respectively, such that i and α/β run over all the ions in the cell and the three Cartesian components, respectively. The transition temperature related to the FE orders can also be simulated by Monte Carlo simulations with an effective Hamiltonian including a local-mode self-energy, a long-range dipole-dipole interaction, a short-range interaction between soft modes, an elastic energy, and an interaction between local modes and local strain (Zhong et al. 1994, 1995).

In addition to examining intrinsic FE instabilities, the effect of epitaxial strain on ferroic properties should be assessed because strain engineering is a powerful tool to induce FE instabilities in perovskite oxides (Schlom et al. 2014). The sensitivity of the ferroelectric state in these materials on the strain state of a film coherently grown on a substrate is due to a strain-phonon coupling mechanism. Typically, compressive strain favors an out-of-plane electric polarization, whereas tensile strain supports an in-plane polarization.

In this section, we introduced the basic strategy to investigate the magnetic and FE properties of candidate multiferroic compounds by using different computational tools. Applying this scheme to design new multiferroics relies on harnessing known mechanisms for ferroelectricity and magnetic order. In the next section, we describe how different atomic specifies, compositions, and structures enable or activate the necessary mechanism for multiferroism through specific compounds.

3 Mechanisms to Achieve Coexisting Ferroelectricity and Magnetism

3.1 Lone-Pair-Induced Ferroelectricity

The representative compounds are $BiFeO_3$ and $BiMnO_3$ (Hill and Rabe 1999). By performing DFT calculations, Seshadri and Hill (2001) demonstrated that the Bi-s orbitals can mix with O-p orbitals (see Fig. 3c) in the monoclinic $C2$ structure (see Fig. 3a), which results in a lobe-like shape of the unbonded $6s^2$ lone-pair electrons about the Bi^{3+} ion (see Fig. 3b). The arrangement of this anisotropic distribution of $6s^2$ electrons in the unit cell is the origin for ferroelectricity in the $C2$ structure of $BiMnO_3$. It may also be used to explain the ferroelectricity that arises in the rhombohedral $R3c$ structure of $BiFeO_3$. In $BiFeO_3$, the long-wavelength (62 nm) spiral magnetic structure can be suppressed in thin film where the spin alignment is G-type with weak ferromagnetism within the (111) plane of the $R3c$ structure. In $BiMnO_3$ (Faqir et al. 1999; Moreira dos Santos et al. 2002a, b; Kimura et al. 2003b; Montanari et al. 2005, 2007; Belik and Takayama-Muromachi 2006; Belik et al. 2007), the spin alignment is ferromagnetic (FM), which leads to a magnetization of $\sim 3.6\mu_B$ per formula unit that is close to the nominal magnetic moment of Mn^{3+} ($4\mu_B$). The FM transition temperature occurs at ~ 105 K. Compared with the easy synthesis of the $BiFeO_3$ in bulk and thin film at ambient condition, high pressure is
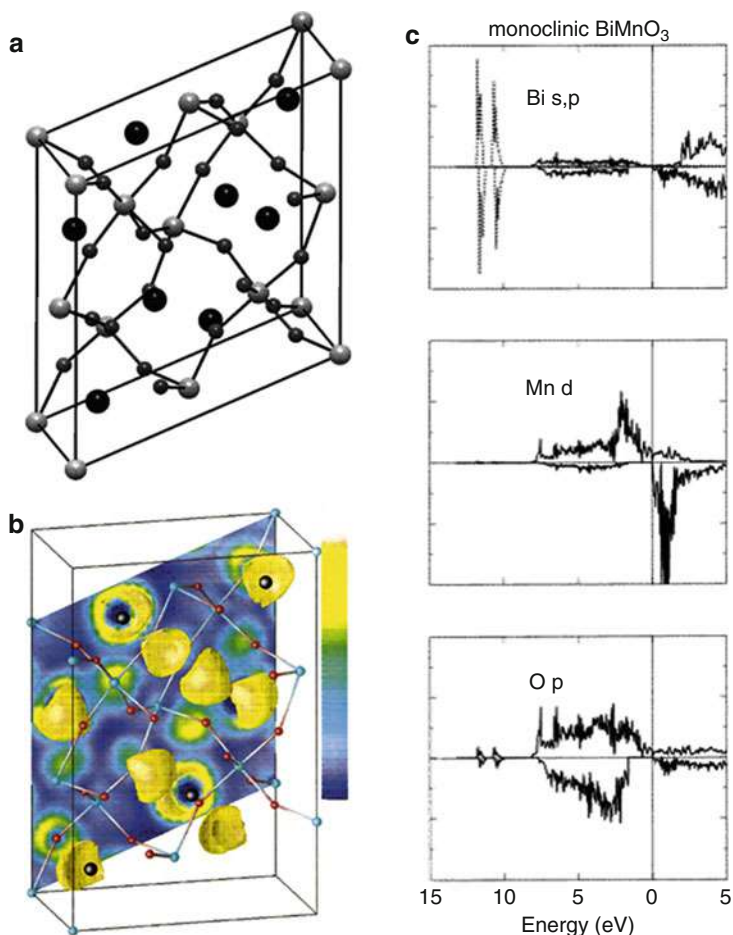
Fig. 3 (**a**) Monoclinic $C2$ structure of $BiMnO_3$ projected nearly down the b axis. The large black sphere is Bi, and the smaller white sphere is Mn centered in an oxygen octahedron. (**b**) Isosurface with a value of 0.75 of the valence electron localization function of the structure shown in (**a**). Blue color corresponds to nearly no electron localization, and white color corresponds to complete localization. (**c**) Orbital resolved density of states (DOS) in FM $C2$ structure of $BiMnO_3$. The DOS have been rescaled to a single spin direction and to one $BiMnO_3$ formula unit. (Figure reproduced from Seshadri and Hill (2001), Copyright American Chemical Society, 2001)

needed to synthesize isoelectronic bulk $BiMnO_3$ single crystals. In thin film form, relatively strict growth conditions are required even without high pressure (Lu et al. 2015a).

It is noted that the ferroelectricity in $BiMnO_3$ is still under debate (Belik et al. 2007; Montanari et al. 2007; Goian et al. 2012), because a first-principles study showed that the ground state adopts $C2/c$ symmetry in ideal, stoichiometric $BiMnO_3$ where the anisotropic distribution of $6s^2$ electrons around different Bi^{3+} ions cancel perfectly leading to no net polarization (Baettig et al. 2007).

3.2 Charge-Ordering-Induced Ferroelectricity

In 2004, Efremov et al. (2004) proposed ferroelectricity could appear by inducing a charge ordering (CO) state in $Pr_{1-x}Ca_xMnO_3$. The CO can exist for $0.3 < x < 0.8$. When $x = 0.5$, the CO is site-centered, in which the Mn^{3+} and Mn^{4+} alignments take the form of checkboard ordering as shown in Fig. 4a. Upon decreasing x to 0.4, the CO becomes bond-centered as demonstrated by experiments (see Fig. 4b), in which the extra electron is shared by the two nearest-neighboring Mn ions. In both types of charge-ordered state, ferroelectricity is absent owing to the presence of inversion symmetry compatible with the CO state. Theoretically when x is between 0.4 and 0.5, and including $x = 0.4$, the site-centered and bond-centered COs should coexist (see Fig. 4c), which will lead to the loss of inversion and a small electric polarization ($\sim 1mC/m^2$) in the monoclinic $P11m$ structure. (As a point of reference, $BaTiO_3$ exhibits an electric polarization of 300 mC/m^2.) Furthermore, the actual valence difference between the nearest-neighboring Mn ions is less than 0.1e even with $x = 0.5$.

Although the CO-induced ferroelectricity was demonstrated by a degenerate double exchange model, the direct observation of the ferroelectricity has been hindered by the finite conductivity displayed by the compound. Support for the appearance of the ferroelectricity in $Pr_{1-x}Ca_xMnO_3$ is given by the anomaly in the dielectric constant observed in vicinity of the CO transition (Jardon et al. 1999; Mercone et al. 2004).

Another compound proposed to exhibit CO-induced ferroelectricity is $LuFe_2O_4$ (Ikeda et al. 2000, 2005) in which the average valence of iron is $Fe^{2.5+}$. Fe^{2+} and Fe^{3+} undergo a three-dimensional CO under 330 K, where the ratio of Fe^{2+} and Fe^{3+} in one triangular layer comprised of corner-sharing FeO_5 trigonal bipyramids is 1:2, while it is 2:1 in the other triangular layer (see Fig. 5b). With this CO, the structure exhibits a $\sqrt{3}$x$\sqrt{3}$ supercell of the unordered $R\bar{3}m$ variant (see Fig. 5a), and ferroelectricity appears below 330 K. By performing DFT total-energy
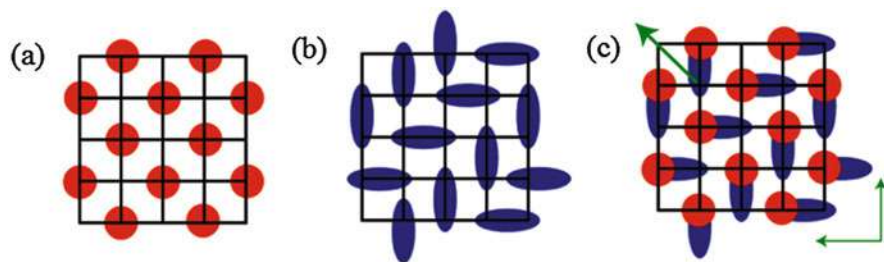
Fig. 4 (**a**) Site-centered charge order. (**b**) Bond-centered charge order. (**c**) A FE intermediate state. The charge-ordered structure in (**c**) lacks inversion symmetry. The bottom green arrows indicate the dipole direction in the horizontal and vertical dimers, and the diagonal green arrow indicates the direction of the electric polarization in the structure. (Figure reproduced from Efremov et al. (2004), Copyright Springer Nature, 2004)
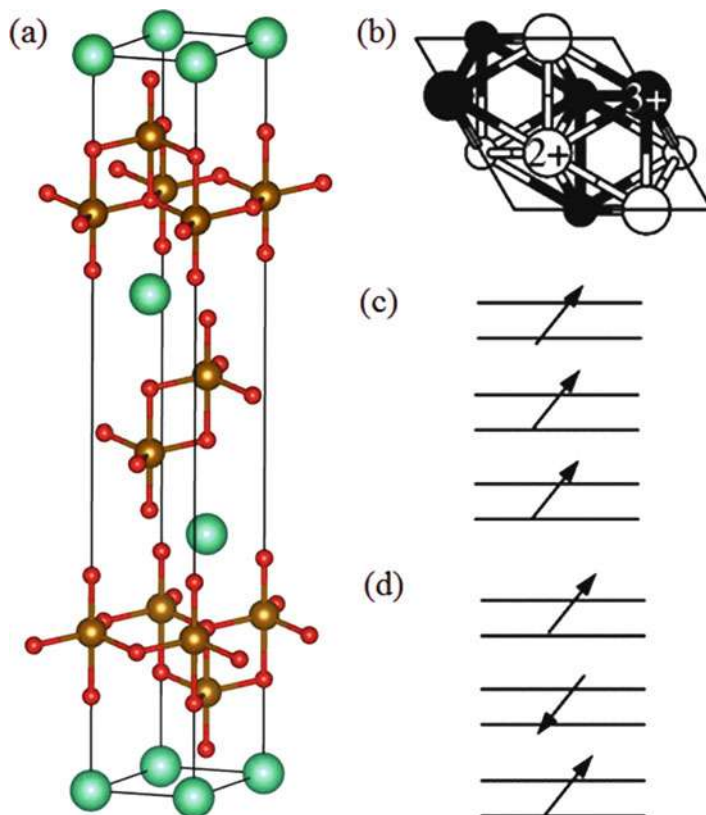
Fig. 5 (**a**) Hexagonal $R\overline{3}m$ structure of $LuFe_2O_4$, in which the large cyan, medium orange, and small red circles represent the Lu, Fe, and O atoms, respectively. (**b**) Schematic illustration of the CO structure, in which the small and large circles indicate the Fe ions in the upper and lower triangular sheets of an iron bilayer shown in (**a**), respectively. The open and closed circles represent Fe^{2+} and Fe^{3+} ions, respectively. (**c**) and (**d**) schematically display ferroelectric and ferrielectric arrangements of the dipole moments in the iron bilayers in the CO structure shown in (**b**). (Figure reproduced from Xiang and Whangbo (2007), Copyright American Physical Society, 2007)

calculations with the GGA exchange-correlation functional in PBE form, Xiang and Whangbo (2007) identified that the CO in $LuFe_2O_4$ leads to a ferrielectric arrangement of the dipole moments among the FeO_5 bilayers by comparing the values of the polarizations calculated in the ferroelectric ($\sim$52.7μC/cm^2) (see Fig. 5c) and ferrielectric ($\sim$26.3μC/cm^2) (see Fig. 5d) states with that from experiment (Ikeda et al. 2005) ($\sim$25μC/cm^2 measured at 77 K). Also in the vicinity of a two-dimensional ferrimagnetic order below 240 K, an increase in the magnitude of the polarization was observed, which indicated a coupling between the ferroelectric and magnetic orders in this compound (Ikeda et al. 2005). It should be noted that in a later experiment (de Groot et al. 2012), $LuFe_2O_4$ was shown not to exhibit ferroelectricity because its equilibrium structure is monoclinic $C2/m$

symmetry (centrosymmetric and the inversion center is located in-between the Fe bilayers) with a new CO configuration in the iron bilayer. Therefore, the CO-induced ferroelectricity in $LuFe_2O_4$ is still controversial although the material has more recently been used in oxide superlattices to achieve an artificial multiferroic (Mundy et al. 2016).

3.3 Spin-Driven Ferroelectricity

3.3.1 Microscopic Models for Spin-Driven Ferroelectricity

KNB Model

In 2005, Katsura et al. (2005) developed a microscopic electronic model for magnetic materials exhibiting a noncollinear spin structure at low temperatures. This so-called KNB model is based on a three-atom system, in which it is assumed that inversion symmetry exists between the two magnetic ions M1 and M2 as shown in Fig. 6. Owing to inversion, there is no Dzyaloshinskii-Moriya (DM) interaction, and the noncollinear spin alignment is due to the competing symmetric spin-exchange interactions among the magnetic ions. By only considering the t_{2g} orbitals with the spin-orbital coupling (SOC) and on-site Coulomb repulsion on the magnetic ions (i.e., M1 and M2) and relying on a process of electron transfer along the M1-O-M2 bond due to orbital hybridization, the KNB model predicts a polarization $\boldsymbol{P} \propto -\boldsymbol{e_{12}} \times (\boldsymbol{e_1} \times \boldsymbol{e_2})$, where $\boldsymbol{e_{12}}$ is the unit vector pointing from M1 to M2 ions and $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$ are the directions of the two noncollinear spins on M1 and M2, respectively (see Fig. 6). It should be noted that the KNB model is a
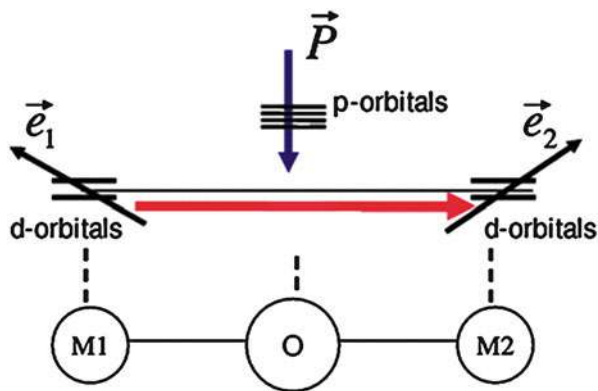
Fig. 6 The model with the two transition metal ions M1 and M2 with the oxygen atom O between sites. The noncollinear spin directions on M1 and M2 are represented by $\mathbf{e_1}$ and $\mathbf{e_2}$. $\boldsymbol{P}$ indicates the direction of the polarization, which is proportional to $-\boldsymbol{e_{12}} \times (\boldsymbol{e_1} \times \boldsymbol{e_2})$ where $\boldsymbol{e_{12}}$ is the unit vector pointing from M1 to M2 ions. (Figure reproduced from Katsura et al. (2005), Copyright American Physical Society, 2005)
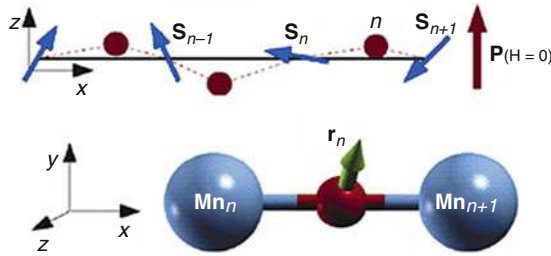
Fig. 7 Top: The *xz* projection of the spin structure and the O ions (solid circles). Bottom: The Mn_2O "molecule" in the cubic perovskite structure, in which $\mathbf{r_n}$ indicates the displacement of the O ion. (Figure reproduced from Sergienko and Dagotto (2006), Copyright American Physical Society, 2006)

purely electronic mechanism for ferroelectricity and there are no ions displacements required or considered in the contribution to the macroscopic polarization.

Inverse Dzyaloshinskii-Moriya Interaction Model

The so-called inverse DM interaction model (Sergienko and Dagotto 2006) arises from the DM interaction in the M1-O-M2 subsystem (M1 and M2 are magnetic ions) *without* inversion symmetry as shown in Fig. 7. By assuming that only the position of the oxygen is allowed to move and those of M1 and M2 are fixed and the DM interaction induced energy is expressed as $\mathbf{D}(r_n) \cdot (\mathbf{S}_n \times \mathbf{S}_{n+1})$, there will be an additional displacement of the oxygen along the *z* direction with the spiral spin structure. The modulation wave vector (0,0.28,0) for the spiral spins was obtained from experiment and $\mathbf{D}^x(r_n) = \gamma(0, -z_n, y_n)$ and $\mathbf{D}^y(r_n) = \gamma(z_n, 0, -x_n)$ in the orthorhombic *Pbnm* symmetry where γ is the DM interaction constant. The result is a polarization along the *z* direction. Although this model was derived based on the spiral spin structure in $TbMnO_3$ compound, the basic physics can be extended to other compounds that permit DM interactions (e.g., those that possess the spiral spin structure) with induced ionic displacements that may significantly contribute to the spin-driven ferroelectricity.

Unified Model for the Spin-Driven Ferroelectricity

Xiang (Wang et al. 2016; Xiang et al. 2011b, 2013b; Lu et al. 2015b) developed a unified model with the pure electronic contribution, ionic displacement contribution, and lattice-deformation contribution together. It is captured by a single formula by expressing the polarization driven by the magnetic structure in the general form: $\boldsymbol{P}_t = \boldsymbol{P}_t(\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_m; \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_n; \eta_1, \eta_2, \ldots, \eta_6)$ where $\boldsymbol{S}_i$ ($i = 1, 2, \ldots, m$), $\boldsymbol{\mu}_j$ ($j = 1, 2, \ldots, n$), and η_k ($k = 1$–6) are the spin vectors, ion displacement vectors, and homogenous strains, respectively. Here, the pure electronic contribution, ion displacement contributions, and lattice deformation contributions arise from the electron density redistribution, ion displacements relative to a paramagnetic (PM)

structure, and changes of the lattice vectors relative to a paramagnetic structure, respectively, due to the appearance of a spin order that occurs upon cooling below the magnetic ordering temperature.

Usually, the spin-order-induced μ_j ($j = 1, 2, \ldots, n$) and η_k ($k = 1$–6) effects are rather small, which allows one to rewrite $P_t \approx P_e(S_1, S_2, \ldots, S_m) + P_{lattice + ion}$ ($\mu_1, \mu_2, \ldots, \mu_n; \eta_1, \eta_2, \ldots, \eta_6$), where the pure electronic contribution (P_e) is separated from the ionic displacement and lattice deformation contributions ($P_{lattice + ion}$). The pure electronic contribution to the polarization (Xiang et al. 2011b) is obtained by considering in the same way a spin dimer (see Figs. 6 and 7), such that $P_e = P_1(S_1) + P_2(S_2) + P_{12}(S_1, S_2)$ where $P_i(S_i) = \sum_{\alpha\beta} P_i^{\alpha\beta} S_{i\alpha} S_{i\beta}$ ($i = 1, 2$) and $P_{12}(S_1, S_2) = \sum_{\alpha\beta} P_{12}^{\alpha\beta} S_{1\alpha} S_{2\beta}$ are the intra-site and inter-site polarization terms, respectively. The inter-site term can be expressed in matrix form as:

$$
P_{12}(S_1, S_2) = \begin{bmatrix} S_{1x} & S_{1y} & S_{1z} \end{bmatrix} \begin{bmatrix} P_{12}^{xx} & P_{12}^{xy} & P_{12}^{xz} \\ P_{12}^{yx} & P_{12}^{yy} & P_{12}^{yz} \\ P_{12}^{zx} & P_{12}^{zy} & P_{12}^{zz} \end{bmatrix} \begin{bmatrix} S_{2x} \\ S_{2y} \\ S_{2z} \end{bmatrix}. \tag{2}
$$

Specifically, if SOC effects are neglected in the P_e contribution, then $P_{12}(S_1, S_2)$ is reduced to be $P_{es}(S_1 \cdot S_2)$ where P_{es} indicates the polarization is induced by exchange striction and $P_{es} = P_{12}^{xx} = P_{12}^{yy} = P_{12}^{zz}$, because any global rotations of all the spins will not change the total electric polarization. If there is inversion symmetry at the center of a spin dimer, then $P_{12}(S_1, S_2)$ can be reduced to be

$$
M(S_1 \times S_2) \quad \text{where} \quad M = \begin{bmatrix} (P_{12}^{yz})_x & (P_{12}^{zx})_x & (P_{12}^{xy})_x \\ (P_{12}^{yz})_y & (P_{12}^{zx})_y & (P_{12}^{xy})_y \\ (P_{12}^{yz})_z & (P_{12}^{zx})_z & (P_{12}^{xy})_z \end{bmatrix} \tag{3}
$$

because $P_{12}^{\alpha\beta} = -P_{12}^{\beta\alpha}$ with the inversion symmetry in the dimer. To extract the coefficients in P_e by using DFT calculations, a so-called four-states mapping method analogous to that used to extract the coefficients in the Hamiltonian model as discussed in Session II is applied. Here, taking the coefficient for the exchange striction as an example, we can write $P_{es} = \frac{P_I - P_{II} - P_{III} + P_{IV}}{4}$ in which P_I, P_{II}, P_{III}, and P_{IV} are the polarizations with the spin arrangements of S_1 and S_2 being (+,+), (+,-), (−,+), and (−,-), respectively, and the other spins are kept the same in the calculations of P_I, P_{II}, P_{III}, and P_{IV}.

For the ionic displacement (Xiang et al. 2013b) and lattice deformation contributions (Lu et al. 2015b), one first considers the total energy based on a paramagnetic reference structure expanded as $E_{PM} = E_0 + \sum_m A_m u_m + \sum_j A_j \eta_j + \frac{1}{2} \sum_{mn} B_{mn} u_m u_n + \frac{1}{2} \sum_{jk} B_{jk} \eta_j \eta_k + \sum_{jk} B_{mj} u_m \eta_j + \ldots$ where A_m, A_j, B_{mn}, B_{jk}, and B_{mj} are the forces, stresses, force constants, frozen-ion elastic constants, and internal displacement tensor, respectively. The elements u_m and η_j correspond to the ionic displacement and the homogenous strain in Voigt notation, respectively, for the paramagnetic reference structure. Second, by considering a magnetic transition

that occurs upon cooling to stabilize a magnetically order state, an additional energy $E_{spin} = E_H + E_{DM} + E_{SIA}$ is required, where E_H, E_{DM}, and E_{SIA} are the Heisenberg symmetric exchange interaction energy, the DM interaction energy, and single-ion anisotropic energy, respectively. Then by considering only u_m and η_j to be induced from the E_H contribution, a simplified form of the symmetric exchange energy can be obtained: $E_H = E_H^0 + \sum_{i,i'} \frac{\partial J_{i,i'}}{\partial u_m} \boldsymbol{S}_i \cdot \boldsymbol{S}_{i'} u_m + \sum_{i,i'} \frac{\partial J_{i,i'}}{\partial \eta_j} \boldsymbol{S}_i \cdot \boldsymbol{S}_{i'} \eta_j + \ldots$ in which $\frac{\partial J_{i,i'}}{\partial u_m}$ and $\frac{\partial J_{i,i'}}{\partial \eta_j}$ are the first order derivatives of the Heisenberg symmetric exchange interaction ($J_{i,i'}$) between spins $\boldsymbol{S}_i$ and $\boldsymbol{S}_{i'}$ with respect to u_m and η_j, respectively. (Here an implied sum notation is used for m and j in E_H hereafter.) Then, by minimizing $E_{PM} + E_H$ with respect to u_m and η_j, one obtains that $\begin{cases} B_{mn}u_n + B_{mj}\eta_j = -\sum_{i,i'} \frac{\partial J_{i,i'}}{\partial u_m} \boldsymbol{S}_i \cdot \boldsymbol{S}_{i'} \\ B_{jk}\eta_k + B_{mj}u_m = -\sum_{i,i'} \frac{\partial J_{i,i'}}{\partial \eta_j} \boldsymbol{S}_i \cdot \boldsymbol{S}_{i'} \end{cases}$. The ionic displacements u_m and strains η_j induced by the spin order are readily obtained by solving the two coupled linear equations. Density functional perturbation theory can be used to compute the force constant (B_{mn}) and the internal-displacement tensor (B_{mj}). The frozen-ion elastic constant (B_{jk}) can be straightforwardly obtained by calculating the strain-stress relation within DFT. To compute the first-order derivatives of the symmetric spin exchange parameter $J_{ii'}$ with respect to η_j, we propose a four-states mapping approach: $\frac{\partial J_{ii'}}{\partial \eta_j} = \frac{1}{4} \left(\frac{\partial E_I}{\partial \eta_j} - \frac{\partial E_{II}}{\partial \eta_j} - \frac{\partial E_{III}}{\partial \eta_j} + \frac{\partial E_{IV}}{\partial \eta_j} \right) = \frac{1}{4} \left(\sigma_j^I - \sigma_j^{II} - \sigma_j^{III} + \sigma_j^{IV} \right)$ (see Fig. 8). Here, I–IV refer to the four spin states with different spin orientations for sites i and i' (see Fig. 8 for an example), E and σ denote the total energy and stress, respectively. We note that the stress can be computed without doing additional DFT calculations due to the celebrated Hellmann-Feynman theorem. The first-order derivatives of the symmetric spin

Fig. 8 Schematic illustration of the four spin states in the four-states mapping method to compute the derivative of the exchange parameter with respect to strain $\frac{\partial J_{ii'}}{\partial \eta_j}$. In the four spin states, only the spins at sites i and i' change the orientation. (Figure reproduced from Lu et al. (2015b), Copyright American Physical Society, 2015)

$$\frac{\partial J_{ii'}}{\partial \eta_j} = \frac{1}{4} \left(\frac{\partial E_I}{\partial \eta_j} + \frac{\partial E_{IV}}{\partial \eta_j} - \frac{\partial E_{II}}{\partial \eta_j} - \frac{\partial E_{III}}{\partial \eta_j} \right)$$
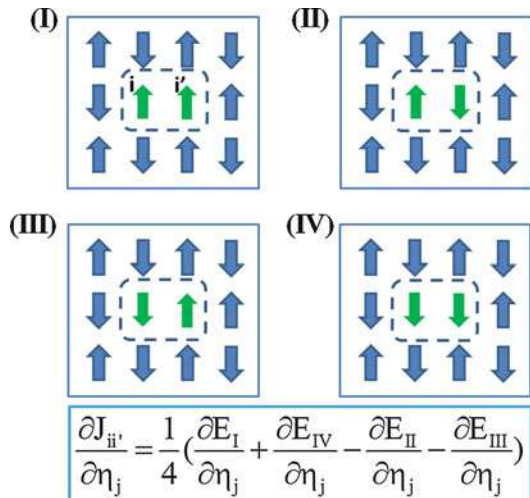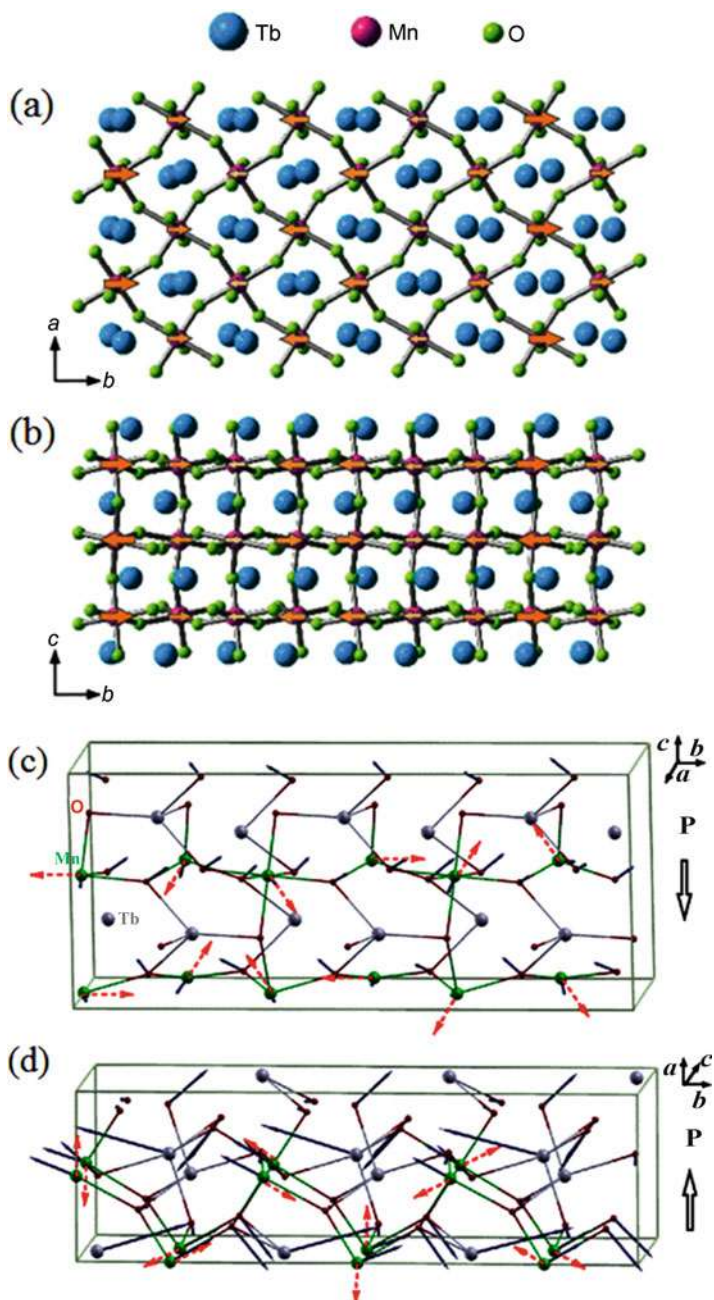
Fig. 9 (continued)

exchange parameter $J_{ii'}$ with respect to u_m can be also efficiently evaluated by using a four-states mapping approach (Xiang et al. 2011a, 2013a). Finally, the electric polarization from the ionic displacements and lattice deformation is given as $P_\alpha = Z_{\alpha m}u_m + e_{\alpha j}\eta_j$ ($\alpha = x,y,z$) in which $Z_{\alpha m}$ and $e_{\alpha j}$ are the Born effective charges and frozen-ion piezoelectric tensors, respectively. These later quantities are directly accessible from a DFT calculation of the non-magnetic or magnetically ordered phase. If one wants to separate P_α into the ionic displacement and lattice deformation contributions, then $\sum_{i,i'} \frac{\partial J_{i,i'}}{\partial u_m} S_i \cdot S_{i'}$ should be set to zero, and the lattice deformation contribution is $P_{lattice} = d_{\alpha j}\sigma_j$ where $\sigma_j = -\sum_{i,i'} \frac{\partial J_{i,i'}}{\partial \eta_j} S_i \cdot S_{i'}$ is the total stress due to the spin structure and $d_{\alpha j} = S_{jk}e_{\alpha k}$ in which S_{jk} and $e_{\alpha k}$ are the relaxed-ion elastic compliance tensor and the relaxed-ion piezoelectric tensor, respectively, and $P_{ion} = P_\alpha - P_{lattice}$.

3.3.2 Multiferroic Materials with Spin-Driven Ferroelectricity

Spin-Orbital Coupling Effects

Cycloidal Spin Structure

TbMnO$_3$ (Kimura et al. 2003a) is the first reported multiferroic material with spin-driven ferroelectricity. Below 41 K, the first magnetic transition occurs, with spins aligned along the b direction of the *Pbnm* structure. The spin structure exhibits a sinusoidal spin-density wave with an incommensurate wave vector (0,0.295,0) as shown in Fig. 9a and b. The experiment showed that there was no ferroelectricity with this magnetic structure. Upon decreasing temperature below 27 K, a second magnetic transition occurs, and the spins rotate in the *bc* plane and form a transverse spiral structure with the wave vector (0,0.28,0) (Kenzelmann et al. 2005) as shown in Fig. 9c. Interestingly, a polarization appears in the vicinity of this magnetic structure. By applying the KNB model (Katsura et al. 2005) and inverse DM interaction model (Sergienko and Dagotto 2006), both approaches can correctly predict the direction of the polarization parallel to the c axis, which was also demonstrated by a phenomenological Ginzburg-Landau model (Mostovoy 2006) and first principles DFT calculations (Xiang et al. 2008; Malashevich and Vanderbilt 2008) with both LDA and GGA (in PBE form) exchange-correlation functionals. Furthermore, the theoretical results (Xiang et al. 2008; Malashevich and Vanderbilt

---

Fig. 9 Schematic illustration of the crystal structure with *Pbnm* symmetry at room temperature (**a**) and spatial variation of Mn magnetic moments along the b axis and atomic displacements ($\Delta z//c$) below 41 K (**b**). Orange arrows indicate Mn magnetic moments below 41 K. The atomic displacements (indicated by solid arrows) of the spin-spiral states after geometry optimization for the *bc* plane spiral (**c**) and the *ab* plane spiral (**d**). The dashed arrows indicate the directions of the Mn spins in the spiral states. The directions of the electric polarizations are also shown. (Figures (**a**) and (**b**) reproduced from Kimura et al. (2003a), Copyright Springer Nature, 2003. Figures (**c**) and (**d**) reproduced from Xiang et al. (2008), Copyright American Physical Society, 2008)

2008) also showed that the polarization in the experiments arises from both the pure electronic contribution from the KNB model and the ionic displacement contribution from the inverse DM interaction model. Moreover, the experiment observed that a moderate magnetic field ($\sim$5 T) along b direction can induce a spin flop (spin-rotational plane is changed from bc to ab planes), which leads to the polarization change from the direction to a direction (see Fig. 9d). This is because the spin-rotational plane is either the ab or bc plane and results from the Mn spin anisotropy (Xiang et al. 2008). The polarization direction with the spin rotating in the ab plane was also correctly predicted by the theoretical works with both models (Xiang et al. 2008). In DyMnO$_3$ (Kimura et al. 2005) and Eu$_{0.75}$Y$_{0.25}$MnO$_3$ (Cheong and Mostovoy 2007; Noda et al. 2006; Hemberger et al. 2007), similar phenomena were observed, and the mechanism is the same as that in TbMnO$_3$. One may find other multiferroics as TbMnO$_3$ with the spiral spin structure in a *Pbnm* crystal structure, such as AMnO$_3$ with A being a rare-earth element.

Helical Spin Structure

MnI$_2$ crystallizes in a centrosymmetric $P\bar{3}m1$ structure (Kurumaji et al. 2011) in which each species forms a triangular lattice stacking along the [001] direction in the sequence of –(I-Mn-I)-(I-Mn-I)- (see Fig. 10a). Upon cooling below 3.45 K, a helical spin structure occurs with spins rotating in the plane perpendicular to the modulation wave vector $\boldsymbol{k}$ = (0.181, 0, 0.439). When a magnetic field of 2.3 T is applied along the [1$\bar{1}$0] direction, the helical spin structure changes to have a propagation wave vector ($\boldsymbol{q}$) along [1$\bar{1}$0] simultaneously with the appearance of an electric polarization along the [110] direction. If the magnetic field is increased to 5 T along the [110] direction, then the helical spin structure will have $\boldsymbol{q}$ along [110] and an electric polarization along (Picozzi et al. 2007). It should be noted that the KNB model, inverse DM interaction model, and the phenomenological Ginzburg-Landau model proposed by Mostovoy cannot describe the aforementioned ferroelectricity induced by the helical spin structure, because $\boldsymbol{q}$ is aligned along the same direction as $\boldsymbol{S}_i \times \boldsymbol{S}_j$, meaning there is no DM interaction between the two nearest-neighboring Mn^{2+} ions.

Later, Xiang et al. (2011b) developed a general KNB (gKNB) model to explain the polarization arising from the helical spin structure. Because of the presence of the inversion symmetry between the two nearest-neighboring Mn^{2+} ions (see the right inset in Fig. 10a), the inter-site terms in $\boldsymbol{P}_e$ can be simplified into the form given by Eq. (3). The DFT-LDA calculations showed that $\mathbf{M}$ has the form of $\begin{bmatrix} M_{11} & 0 & 0 \\ 0 & M_{22} & M_{23} \\ 0 & M_{32} & M_{33} \end{bmatrix}$. Therefore, by considering all the Mn^{2+} sites have the same polarization and the contributions from the intra-site terms to the total polarization are zero, $\boldsymbol{P}_{tot} = \boldsymbol{P}_0 = \sum_{k=1}^{6} \boldsymbol{M}_{0k} (\boldsymbol{S}_0 \times \boldsymbol{S}_k)$ where $\boldsymbol{M}_{0k}$ ($k = 1 - 6$) can be obtained from $\boldsymbol{M}$ by using the mirror symmetry of the structure (see Fig. 10b). In the case of the propagation wave vector ($\boldsymbol{q}$) along [Q00] direction, which is redefined in the coordinate system in Ref. (Xiang et al. 2011b) (see Fig. 10b)
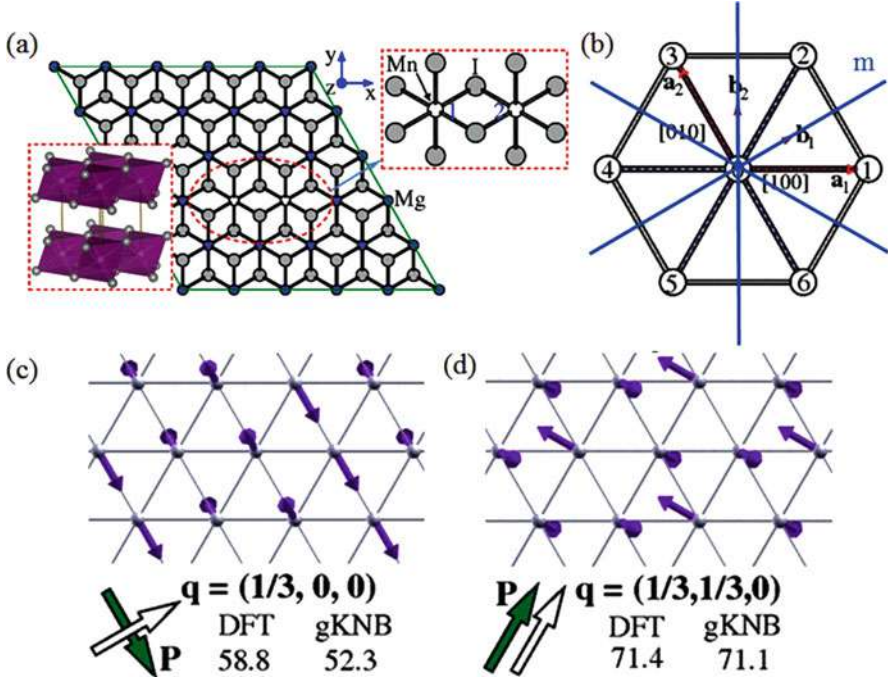
Fig. 10 (**a**) The $5 \times 5 \times 1$ supercell of MnI_2 in which all Mn ions except for an isolated nearest-neighboring Mn-Mn pair are replaced by nonmagnetic Mg ions for computing the coefficients in **M** (see section "Unified Model for the Spin-Driven Ferroelectricity"). The left inset displays the layered structure of MnI_2. The right inset shows the top view of the Mn_2I_{10} cluster. (**b**) The triangular lattice of Mn ions, in which the in-plane lattice vectors a_1 and a_2 and the corresponding reciprocal lattice vectors b_1 and b_2 are shown. The blue lines indicate the mirror (m) planes. (**c**) and (**d**) display the spin orientations of two proper-screw spirals with modulation vectors $q = (1/3,0,0)$ and $q = (1/3,1/3,0)$, respectively. The modulation vector q and the polarization vector P are represented by the white and green arrows, respectively. The numbers in $\mu C/cm^2$ indicate the magnitudes of the polarizations obtained from the direct DFT calculations and the gKNB model. (Figure reproduced from Xiang et al. (2011b), Copyright American Physical Society, 2011)

and corresponds to the $[1\bar{1}0]$ direction at the low magnetic field in the experiment (Kurumaji et al. 2011) (see Fig. 10c), $\boldsymbol{P}_{tot} = \left(\frac{\sqrt{3}}{2}A, -\frac{3}{2}A, 0\right)$ in the coordinate system as depicted in Fig. 10a, with $A = (-M_{11} + M_{22}) \sin(2\pi Q)$. And for $\boldsymbol{q}$ along [QQ0] direction (corresponding to [110] direction at the high magnetic field in the experiment (Kurumaji et al. 2011)) (see Fig. 10d), $\boldsymbol{P}_{tot} = \left(\frac{1}{2}B, \frac{\sqrt{3}}{2}B, 0\right)$ with $B = (-M_{11} - 3M_{22} + 4M_{11} \cos 2\pi Q) \sin 2\pi Q$. It is noted that $\boldsymbol{P}_{tot}$ with [Q00] is perpendicular to the propagation vector and $\boldsymbol{P}_{tot}$ with [QQ0] is parallel to the propagation vector. Thus, the gKNB model obtained the correct polarization directions in the helical spin structures at the low and high magnetic fields, respectively. Furthermore, the polarization values from the gKNB model are also in good agreement with those from the direct DFT calculations (see

Fig. 10c and d). Thus, the gKNB model both predicts correctly the direction of the polarization with a certain spin arrangement and gives an accurate value of the polarization. The gKNB model is then useful for predicting the polarization with arbitrary spin arrangements and for investigating the microscopic mechanism on the spin-ordering-induced polarization, such as elucidating the important path and the important coefficients.

In $CuFeO_2$ (Kimura et al. 2006) and $ACrO_2$ (A = Cu,Ag) with the delafossite structure (Seki et al. 2008), similar helical spin structures with propagation wave vectors $(q,q,0)$ where q is incommensurate were also observed. In $CuFeO_2$, the helical spin structure only appears by applying the magnetic field along the out-of-plane axis with a magnitude between 7 and 13 T below 11 K. For small Al substitution values on the B site in $CuFeO_2$ [i.e., $CuFe_{1-x}Al_xO_2$ ($x = 0.01, 0.02$)] (Seki et al. 2007; Nakajima et al. 2008), the helical spin structure can be stabilized without any magnetic field at low temperature (6–9 K for $x = 0.01$ and below 7 K for $x = 0.02$). In $ACrO_2$ (A = Cu,Ag), below 24 K, the helical spin structure appears.

Exchange Striction Effects

Helical Spin Structure

$CaMn_7O_{12}$ (Zhang et al. 2011; Johnson et al. 2012) was reported to exhibit improper ferroelectricity ($0.287\mu C/cm^2$) with a helical spin structure having propagation wave vector $(0,1,0.963)$, where both the microscopic mechanisms for the polarization and magnetic structure can be elucidated by using computational methods. Experiments showed that when the temperature is below 440 K, the structure transforms into a $R\bar{3}$ phase, where there are one Mn^{4+} and six Mn^{3+} ions per formula unit as can be seen in Fig. 11. Figure 11a and b shows the spiral channel comprised of every three chains in which Mn^{3+} ions (Mn1 and Mn2) align along the crystallographic c axis. Moreover, the spiral chains share one Mn^{3+} chains to form hexagonal channels, in which the Mn^{4+} (Mn3) and Ca^{2+} are at the center of the cavities aligned along the c direction (see Fig. 11c and d)]. Experiments showed that a helical spin order appears with the temperature between 48 K and 90 K, in which the spins on the Mn^{3+} ions in one Mn^{3+} chain are ferromagnetic (FM) and the spins between the Mn^{3+} chain in the spiral chain are frustrated and form a $120°$ arrangement (see Fig. 11e and f). The Mn^{4+} spins make an angle of $90°$ with the Mn^{3+} spins in one set of the three FM chains (dotted triangle in Fig. 11f) and $30°$ with the Mn^{3+} spins in another set of the three FM chains (dashed triangle in Fig. 11f). In addition, all the spins lie in the plane nearly perpendicular to the c direction.

Here, the DFT-PBE + U calculations with the on-site repulsion $U = 3$ eV and the exchange parameter $J = 1$ eV were carried out (Lu et al. 2012), in which the U and J assess the Hubbard U and Hund's energy for the correlated orbitals, respectively, and U was chosen to make the experimental magnetic structure more stable and the system insulating. It was found that the FM order in one Mn^{3+} chain is due to the strong FM spin exchange interaction (J_1) between Mn1 and Mn2. The $120°$ frustrated spin alignment in the spiral chain arises from the strong AFM spin exchange interaction (J_2) between Mn1 ions in the adjacent Mn^{3+} chains (see
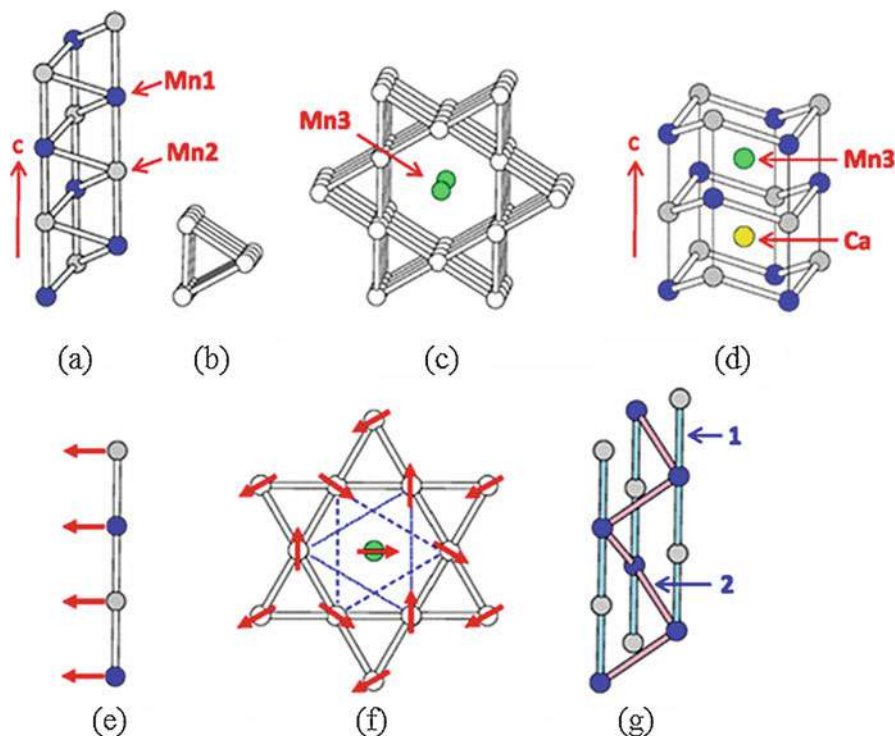
Fig. 11 (**a**) Three adjacent chains of Mn^{3+} ions along the c direction, in which the blue and gray circles represent the Mn1 and Mn2 atoms, respectively. (**b**) A spiral chain comprised of three adjacent chains of Mn^{3+} ions viewed approximately along the c direction. (**c**) Three-dimensional arrangement of the Mn^{3+} and Mn^{4+} (Mn3) ions in $CaMn_7O_{12}$ viewed approximately along the c direction. For simplicity, the Ca^{2+} ions are not shown. (**d**) Arrangements of the Mn^{3+} ions surrounding two adjacent Mn^{4+} and Ca^{2+} ions, where green and yellow circles represent the Mn^{4+} and Ca^{2+} ions, respectively. (**e**) Arrangement of the Mn^{3+} spins in a single chain along the c direction. (**f**) Projection view of the Mn^{3+} and Mn^{4+} spin arrangements along the c direction, where each triangle represents three Mn^{3+} chains forming a spiral chain. The spins of each FM chain along the c direction (see (**e**)) are represented by a single spin. (**g**) Two important spin exchange paths in a spiral chain comprised of three adjacent chains along the c direction, where the numbers 1 and 2 indicate J_1 (FM) and J_2 (AFM), respectively. (Figure reproduced from Lu et al. (2012), Copyright American Physical Society, 2012)

Fig. 11g). The DFT $+ U$ plus SOC effects (DFT $+ U + SOC$) calculations (Lu et al. 2012) also showed that there exists an easy axis for the Mn1 spin and an easy plane for the Mn2 spins, which are both much closer to the ab plane than to the c axis, thus leading to the nearly in-plane alignment of all the spins.

It was also found that if only symmetric spin exchange interactions are considered, then the Mn^{4+} spin will align along the direction of $\mathbf{S}_1 + \mathbf{S}_4$, because the energy of the hexagonal channel (see Fig. 12a and b) $E_{hex} \propto (J_3 + J_4)\cos\alpha$, where J_3 and J_4 are the spin exchange interactions between Mn^{4+} and Mn1 and Mn^{4+} and Mn2, respectively, and they are both negative. Therefore, α will be zero
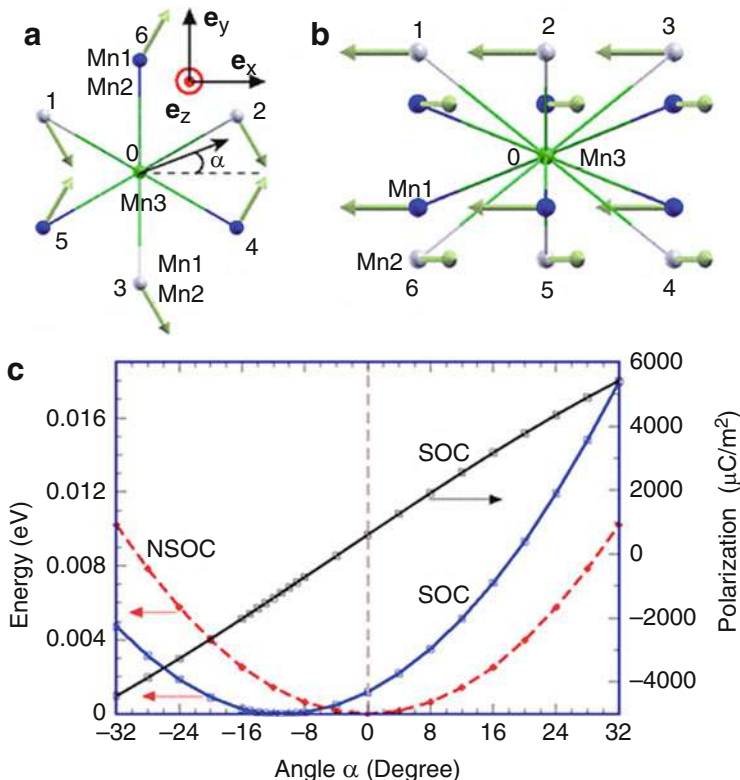
Fig. 12 (**a**) and (**b**) display the top and side views of six Mn1 and six Mn2 ions around a Mn3 ion, respectively. The directions of the Mn1 and Mn2 spins observed experimentally are also shown. (**c**) The total energy and electric polarization as a function of the Mn3 spin direction α (see the text for the definition). NSOC indicates no SOC involved in the calculations. (Figure reproduced from Lu et al. (2012), Copyright American Physical Society, 2012)

in order to lead to a maximized energy lowering, but this is inconsistent with the experiment showing $\alpha \approx -30°$. To reconcile this discrepancy, the DM interactions between Mn^{4+} and Mn1 and Mn^{4+} and Mn2 were examined, and it was found that the DM interaction ($\mathbf{D}_4$) between Mn^{4+} and Mn2 is very large ($|\mathbf{D}_4|/|J_4| \sim 0.54$, compared with the usual ratio of ~ 0.1). Thus, by adding a DM interaction energy arising from $\mathbf{D}_4$ into E_{hex}, $E_{hex}^{new} \propto (J_3 + J_4)\cos\alpha + D_4^z\sin\alpha$ was obtained. By minimizing E_{hex}^{new}, $\alpha = \arctan\frac{D_4^z}{J_3+J_4} \approx -11.2°$, which is in good agreement with the direct DFT + U + SOC calculations. The quantitative difference in the magnitude of the angle compared with experiment can be improved by decreasing U value. Therefore, a particular spin exchange path between Mn^{4+} and Mn^{3+} ions, in which the DM antisymmetric interaction (Moriya 1960a, b) between them is unusually strong for magnetic insulators (i.e., $|D/J| \approx 0.54$), was uncovered to play a key role in determining the helical spin structure.

Although SOC effects are important to obtain the experimental spin structure, the DFT $+ U +$ SOC calculations with the helical spin structure with the propagation wave vector $(0,1,1)$ in the experimental structure indicated that the SOC contribution to the total polarization is minor. Thus, it becomes important to investigate the role of exchange striction using the $\boldsymbol{P}_{es}(\boldsymbol{S}_1 \cdot \boldsymbol{S}_2)$ term. The model results demonstrated that the polarization induced by the exchange striction arises mainly from the path that includes Mn^{4+} and Mn2 (i.e., P_{es}^4), where the polarization on one Mn^{4+} ion can be written as $\boldsymbol{P} \propto P_{es,4}^z \sin\alpha \operatorname{sign}[\boldsymbol{e}_z \cdot (\boldsymbol{S}_4 \times \boldsymbol{S}_1)]$ because of the threefold rotational symmetry and inversion symmetry between the upper set and bottom set of the Mn^{4+} and Mn2 path. From the form of $\boldsymbol{P}$, it should also be noted that the direction of $\boldsymbol{P}$ is determined by the sign of the scalar chirality $\sigma \sim \boldsymbol{e}_z \cdot (\boldsymbol{S}_4 \times \boldsymbol{S}_1) \sim \boldsymbol{r}_{41} \cdot (\boldsymbol{S}_4 \times \boldsymbol{S}_1)$, and the magnitude is also related to the finite angle α that is due to the large DM interaction ($\mathbf{D}_4$) between Mn^{4+} and Mn2. Moreover, the results from the exchange model are also in good agreement with the direct DFT $+ U$ calculations where the magnitude of the polarization is dependent of α and a large electric polarization of $-0.4 \mu C/cm^2$ is found at $\alpha = -30°$. Therefore, it was confirmed that the large improper polarization in $CaMn_7O_{12}$ with the helical spin structure arises from the combined effects of the exchange striction and strong DM interaction in a particular spin exchange path. To design new multiferroics based on these principles, *A*-site cation (i.e., Ca) substitution (Belik et al. 2016) or doping (Zhang et al. 2015) may be a promising way to optimize the structural and magnetic properties.

Collinear Spin Structure

$HoMnO_3$ with *Pbnm* symmetry was predicted theoretically (Sergienko et al. 2006; Picozzi et al. 2007) to adopt an E-type collinear spin structure in its ground state. This spin structure consists of spins that are aligned ferromagnetically in a zigzag chain, and the spins in different chains are coupled antiferromagnetically as shown in the left panel of Fig. 13. With the E-type spin configuration, the atomic structure adopts a noncentrosymmetric state owing to a net displacement of Mn and O ions along the *c* axis, leading to a relatively large polarization along the *c* direction ($\sim 6 \mu C/cm^2$) (right panel, Fig. 13). The DFT-PBE results also showed that the ferroelectricity is related to the difference ($\Delta\alpha = \alpha_p - \alpha_{ap}$) between the Mn-O-Mn angle (α_p) with parallel spins on Mn ions and the angle (α_{ap}) with antiparallel spins as can be seen in Fig. 14. Here both P and $\Delta\alpha$ decrease with increasing U. Furthermore, the total polarization can be divided into the pure electronic contributions and ionic contribution with almost equal values from each ($3.47 \mu C/cm^2$ and $3.28 \mu C/cm^2$, respectively) (Xiang et al. 2013b; Picozzi et al. 2007).

Experimentally, upon cooling to 26 K in polycrystalline $HoMnO_3$, the E-type spin structure sets in and a spontaneous polarization arises ($\sim 0.001 \mu C/cm^2$) (Lorenz et al. 2007), but it is smaller than the predicted from DFT calculations. Although the polarization increases upon further cooling below 15 K due to a complex magnetic structure arising from the spins ordering of Ho, it does not change significantly ($\sim 0.009 \mu C/cm^2$). Subsequent experimental studies have found that the magnitude of the polarization can be increased significantly when samples are
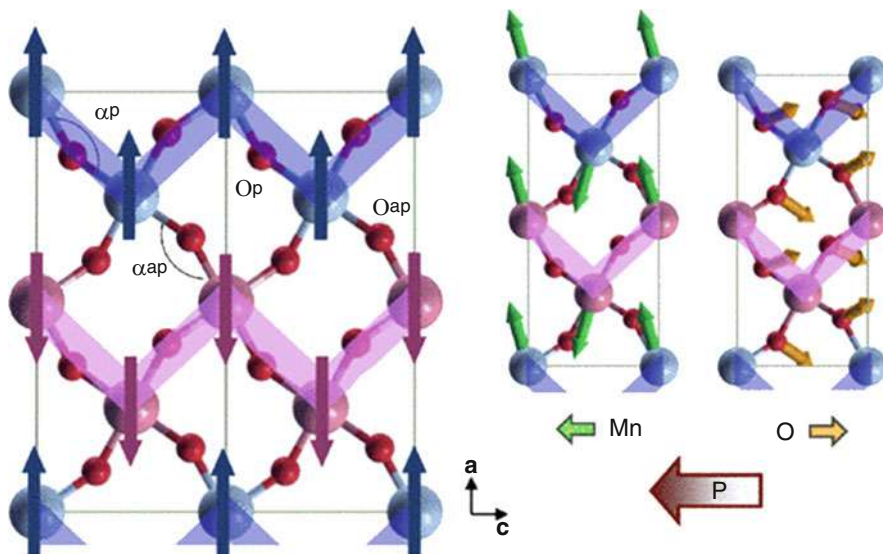
Fig. 13 Left panel: Projection of orthorhombic HoMnO$_3$ in the *ac* plane. Arrows indicate the direction of spins and the AFM zigzag spin chains are denoted by shaded areas. Mn-O-Mn angle (in degrees) for parallel and antiparallel Mn spins are represented by α_p and α_{ap}, respectively. Right panel: Arrows show the directions of the ionic displacements of Mn (left) and O (right) with the AFM-E spin alignment. The thick arrows at the bottom show the direction of the net displacements of Mn and O ions and the resulting $\boldsymbol{P}$. (Figure reproduced from Picozzi et al. (2007), Copyright American Physical Society, 2007)
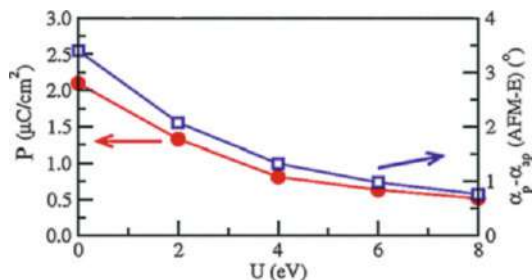
Fig. 14 P$_{PCM}$ and α_p-α_{ap} versus U with the AFM-E spin alignment. P is computed by using the point charge model (PCM) with nominal charges (i.e., $-2e$ for O and $+3e$ for Ho and Mn). (Figure reproduced from Picozzi et al. (2007), Copyright American Physical Society, 2007)

grown in thin film geometries, including YMnO$_3$ with $\sim$0.8μC/cm^2 (Nakamura et al. 2011) and TmMnO$_3$ with $\sim$0.45μC/cm^2 (Han and Chao 2010). Both of these materials exhibit the same E-type spins order as reported in HoMnO$_3$. One may use small A-site cation substitution and/or strain engineering to obtain a large polarization in such type of the multiferroics. Before experiment, the DFT

Fig. 15 A schematic description of the spin alignment in TbMn$_2$O$_5$ in the *ab* plane. The angle of spin direction is given in parenthesis, where q$_x$ is the propagation wave vector of the spin order in *x* direction. The dashed, solid, and double lines represent J_3, J_4, and J_5 exchange interaction paths, respectively. (Figure reproduced from Cao et al. (2009), Copyright American Physical Society, 2009)
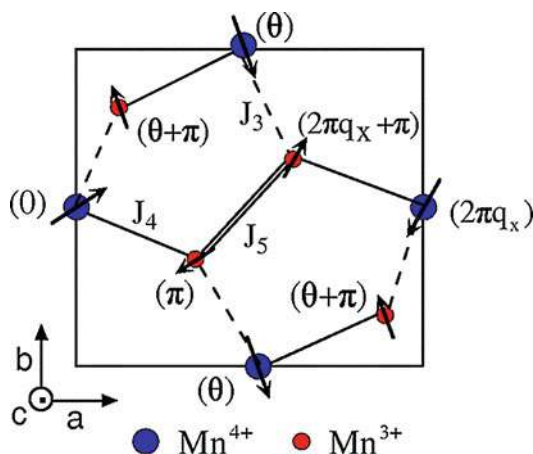
calculations can be used to model the crystal and spin structures and formulate a design strategy based on the spin Hamiltonian model.

TbMn$_2$O$_5$ (Hur et al. 2004a; Chapon et al. 2004) containing Mn^{4+}O$_6$ octahedra and Mn^{3+}O$_5$ square pyramids exhibits a complex sequence of magnetic phase transitions. An incommensurate spin order with propagation wave vector (0.48,0,0.245) appears below 44 K followed by an incommensurate-to-commensurate spin structure transition, with the propagation wave vector (0.5,0,0.25), at 38 K. The lower temperature transition is accompanied by the appearance of ferroelectricity. Another magnetic transition occurs from the commensurate phase to a different incommensurate spin structure with propagation wave vector (0.48,0,0.28) at 18 K and both dielectric anomaly and significant decrease in the electric polarization arising from the commensurate magnetic order that lifts inversion symmetry.

Most computational investigations (Wang et al. 2008, 2007; Cao et al. 2009) of the ferroelectricity in TbMn$_2$O$_5$ focus on the commensurate spin order depicted in Fig. 15. The spins aligned in the *ab* plane in the zigzag chain along the *a* direction are coupled antiferromagnetically with a small spin canting as are the spins in different zigzag chains along the *b* direction. The spins are arranged along the *c* direction with an order described by a modulation of $\frac{\pi}{2}$ in phase factor. With this magnetic spin structure, the symmetry is reduced from *Pbam* to polar *Pb2$_1$m*, enabling a polarization along the *b* direction.

First-principles calculations (Cao et al. 2009) of TbMn$_2$O$_5$ with this spin configuration (GGA exchange-correlation functional in PBE form was used) indicate that the polarization consists of a pure electronic contribution (minor contribution) from the exchange striction term in $\boldsymbol{P}_e$ and an ionic contribution (main effect) owing to exchange striction. Calculations of the electric polarization also indicated that the ion displacements are mainly from the Mn^{3+}-Mn^{4+}-Mn^{3+} segments along the *b* direction that are connected through the Mn^{3+}-Mn^{3+} spin exchange interaction between the segments. In the Mn^{3+}-Mn^{4+}-Mn^{3+} segment, the spins on the Mn^{3+} ions above and below the Mn^{4+} align in an opposite direction, either parallel or

antiparallel to the spin on Mn^{4+}. Because of this spin order, the Mn^{3+}-Mn^{4+} path in the Mn^{3+}-Mn^{4+}-Mn^{3+} segment exhibits parallel spin alignment that allows the Mn^{3+} and Mn^{4+} cations to move closer to each other in order to lower the exchange energy in the *Pbam* structure. Finally, this exchange striction effect, expressed as $-\sum_{i,i',m} \frac{\partial J_{i,i'}}{\partial u_m} u_m S_i \cdot S_{i'}$ where $J_{i,i'}$ runs over all the spin exchange interactions between nearest-neighboring Mn^{3+} and Mn^{4+} and u_m is the ion displacement (where m includes all the ions in the unit cell), leads to the symmetry reduction to $Pb2_1m$.

Furthermore, owing to the existence of the exchange striction effects in the Mn^{3+}-Mn^{4+} path, the incommensurate magnetic phase competes with commensurate magnetic phase, resulting in the appearance of the incommensurate magnetic phase at 18 K with the anomalous magnetodielectric response. The exchange striction-induced ferroelectricity in $TbMn_2O_5$ was also demonstrated in experiment (Chapon et al. 2004; Kagomiya et al. 2003; Hur et al. 2004b; Higashiyama et al. 2005; Radaelli et al. 2009; Stroppa et al. 2010) by investigating the incommensurate spin order, in which the spins in the different zigzag chains along the *b* direction were found to be almost orthogonal to each other, leading to the loss of the exchange striction effects and the strong reduction of the polarization in the incommensurate phase.

The exchange striction-induced ferroelectricity were also found in $DyFeO_3$ (Stroppa et al. 2010), CdV_2O_4 (Giovannetti et al. 2011), and $Ca_3(Co,Mn)_2O_6$ (Choi et al. 2008) compounds. Because the exchange striction effects are related to $S_i \cdot S_j$, the corresponding ferroelectricity is usually larger than that obtained from the SOC effects ($S_i \times S_j$), because of the fact that the symmetric spin exchange interaction $\sim S_i \cdot S_j$ is larger than the antisymmetric DM interaction $\sim S_i \times S_j$. This trend can be readily observed from pressure-induced magnetic phase transitions in $TbMnO_3$, i.e., upon transitioning from the cycloidal spin structure to the E-type spin structure whereby an enhancement of the polarization to $\sim 1 \mu C/cm^2$ occurs (approximately one that is an order-of-magnitude increase) (Aoyama et al. 2014).

Magnetoelasticity-Induced Ferroelectricity

If a material exhibits a paramagnetic structure that is also piezoelectric, then a magnetic phase transition may induce a polarization that arises from magnetoelastic coupling induced strain on the lattice constants (hereafter denoted by lattice deformation) and the piezoelectricity. And the polarization, as discussed in section "Unified Model for the Spin-Driven Ferroelectricity," can be expressed as $P_{lattice} = d_{\alpha j}\sigma_j$ where $\sigma_j = -\sum_{i,i'} \frac{\partial J_{i,i'}}{\partial \eta_j} S_i \cdot S_{i'}$ is the total stress due to the spin-ordered structure and piezoelectric coefficient $d_{\alpha j} = S_{jk}e_{\alpha k}$ in which S_{jk} and $e_{\alpha k}$ are the relaxed-ion elastic compliance tensor and the relaxed-ion piezoelectric tensor, respectively. Here, we take $BiFeO_3$ compound as an example to understand this mechanism and indeed describe how a large polarization arises from the lattice deformation (spontaneous strain) (Lu et al. 2015b). A G-type spin order was imposed in the DFT-PBE study, and only nearest-neighboring Fe-Fe spin exchange interactions were used in computing the stress σ. This later assumption

Table 1 The different contributions to the electric polarization (in unit of $\mu C/cm^2$) induced by the G-AFM order in $BiFeO_3$ from model and DFT calculations. $P_{lattice}$, P_e, and P_{ion} refer to the lattice deformation, pure electronic, and ion displacement contributions, respectively. Table reproduced from Lu et al. (2015b), Copyright American Physical Society, 2015

Polarization	$P_{lattice}$	P_e	P_{ion}
Model	1.32	0.53	0.56
DFT	1.22	0.40	0.54

proved to be accurate enough to reproduce σ from direct DFT calculations. The $P_{lattice}$ contribution was obtained to be as large as $\sim 1.32 \mu C/cm^2$. It is even larger than the sum of the pure electronic and ion displacement contributions (Table 1). This is an unprecedented result in which a previously unknown spin-order-induced contribution ($P_{lattice}$) to the electric polarization is found to be even larger than the typical contributions (the pure electronic and ion displacement contributions). The result suggests that when one is trying to design new multiferroics, materials large piezoelectricity and spin ordering is a promising candidate with large spin-ordering-induced polarization for strong magnetoelectricity. Table 1 also shows that the result obtained from the model used in this study is in agreement with the direct DFT calculations. Summing up all the three spin-order-induced contributions, the total polarization calculated for the G-type AFM order in $BiFeO_3$ is $\sim 2 \mu C/cm^2$, which is approximately 2% of the total polarization ($\sim 100 \mu C/cm^2$) that is induced by the ionic contribution from the Bi off-centering. The spin-induced polarization in $BiFeO_3$ is also comparable with that of $HoMnO_3$ (Sergienko et al. 2006; Picozzi et al. 2007).

3.4   Geometric Ferroelectricity

3.4.1   Single Primary Mode Coupled to a Polar Mode (Improper Ferroelectricity)

HexagonalYMnO$_3$ (h-YMnO$_3$) (Van Aken et al. 2004) is a ferroelectric (space group: $P6_3cm$) with a T_c as high as ~ 1270 K and an electric polarization of $\sim 5.6 \mu C/cm^2$. Below approximately 70 K, an AFM transition occurs and $YMnO_3$ becomes multiferroic (Lottermoser et al. 2004; Fiebig et al. 2000). Density functional theory calculations within the LDA approximation demonstrated that the polar nature of h-YMnO$_3$ arises from condensation of a zone-boundary phonon in the first Brillouin zone, labeled as K_3, which results in the buckling (trimerization) of MnO_5 bipyramids (Van Aken et al. 2004). The K_3 mode couples to a stable polar lattice mode, which shorten and elongate the Y-O_{eq} distances along the c direction (see Fig. 16) to produce a net ferroelectricity. As a consequence, the polarization P is proportional to either $Q_{K_3}^3$ or Q_{K_3}, depending on the magnitude of mode Q_{K_3} (Fennie and Rabe 2005). The A-site substitution with the elements in the lanthanides can also lead to this type of the improper ferroelectricity (Young et al. 2015).
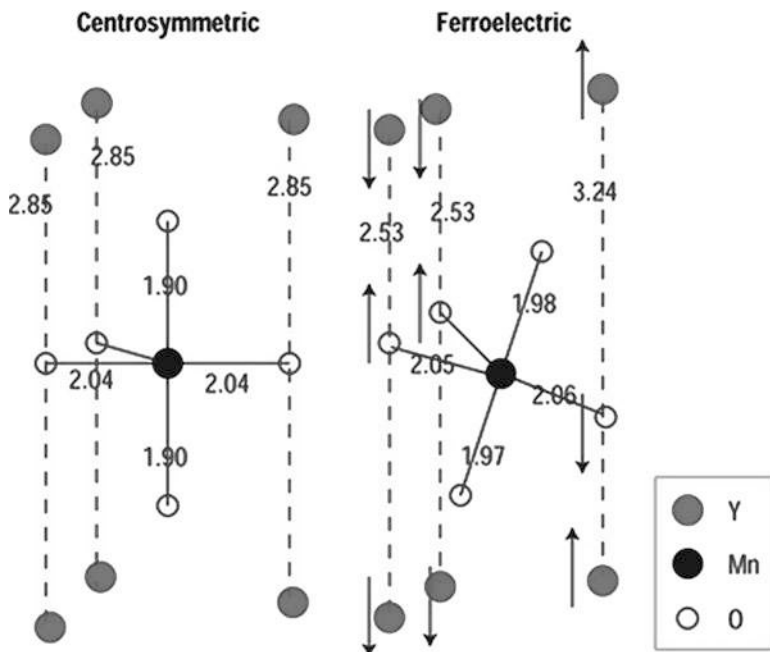
Fig. 16 The DFT calculated atomic positions in the centrosymmetric and ferroelectric structures. The numbers display the bond lengths in Å. The arrows indicate atomic displacements in the ferroelectric structure with respect to the centrosymmetric structure. (Figure reproduced from Van Aken et al. (2004), Copyright Springer Nature, 2004)

Recently, first-principles DFT-PBE calculations also demonstrated that the K_3 mode may also lead to magnetoelectric coupling in hexagonal LuBO_3 (B = Mn,Fe) (Das et al. 2014), which is isostructural to h-YMnO$_3$. Based on a phenomenological model in which there exists a trilinear term $F_{tri} \sim (\mathbf{Q}_I \cdot \mathbf{L}_I + \mathbf{Q}_{II} \cdot \mathbf{L}_{II}) M_z$ where $\mathbf{Q}_i$ and $\mathbf{L}_i$ (i = I,II) are the K_3 mode vector and AFM vector, respectively, as depicted in Fig. 17b, and M_z is the total magnetization in the two magnetic layers I and II, it was found that if $\mathbf{Q}_i$ changes its tilting direction, then either $\mathbf{L}_i$ or M_z change accordingly in order to lower the energy of the trilinear term F_{tri}. For example, when $\mathbf{Q}_i$ changes from configuration α^+ to β^- (Fig. 17a), the direction of the polarization will reverse, and according to F_{tri}, either $\mathbf{L}_i$ or M_z will be reversed accompanying the reversal of the polarization. The interesting phenomenon here is the switching of the magnetization upon electric polarization switching, while it may not be always true, because if $\mathbf{L}_i$ is reversed, M_z will be unchanged. It should be noted even if $\mathbf{L}_i$ is changed, ME coupling also occurs via the linear ME coefficient $\alpha_{zz} \approx \frac{\partial M_z}{\partial Q_{K_3}} * \frac{\partial P_z}{\partial Q_{K_3}}$ where P_z is the electric polarization along the c axis in hexagonal $P6_3cm$. More details about such ME effects are discussed in the Sect. 3.10.

3.4.2 Two Primary Modes Coupled to the Polar Mode (Hybrid Improper Ferroelectricity)

Hybrid improper ferroelectrics were first proposed to occur in ultrashort period $PbTiO_3/SrTiO_3$ superlattices (Bousquet et al. 2008; Aguado-Puente et al. 2011). Bousquet et al. reported a new mechanism for ferroelectricity, whereby a stable polar mode (P) can be coupled to *two* primary antiferrodistortive (AFD) modes (most unstable modes at the phase transition), i.e., an in-phase oxygen octahedral-rotational mode ($Q_{M_2^+}$) and an out-of-phase oxygen octahedral-rotational mode ($Q_{M_4^-}$). In this particular compound, both rotational modes occur about the out-of-plane axis (c axis) as shown in Fig. 18. This coupling among multiple lattice phonons is referred to as a "trilinear coupling interaction" ($\sim PQ_{M_2^+}Q_{M_4^-}$), because of the total energy of the material can be further lowered by its presence. Consequently, a stable polar mode will condense in the ground state even if it is disfavored in the absence of the rotational modes. This fact makes the ferroelectricity "improper" (Benedek et al. 2015). Later, a similar phenomenon was discovered in $Ca_3Mn_2O_7$ by Benedek and Fennie (2011), in which the trilinear coupling interaction also induces the condensation of an essentially (weakly) stable mode (see Fig. 19a). In contrast to the $PbTiO_3/SrTiO_3$ superlattices, the two primary modes in the manganite are one in-phase oxygen octahedral-rational mode (Q_R) around the out-of-plane axis (e.g., $[001]_{pt}$ pseudo-tetragonal axis) and one out-of-phase oxygen octahedral-tilt mode (Q_T) around an in-plane axis (e.g., $[\bar{1}10]_{pt}$ pseudo-tetragonal axis) so that $F_{tri} \sim PQ_RQ_T$. Such materials containing these trilinear coupling interactions are referred to as *hybrid* improper ferroelectrics, and strategies toward realizing them by computational design have been described in several review articles (Benedek et al. 2012, 2015; Harris 2011; Mulder et al. 2013; Oh et al. 2015; Senn et al. 2015; Liu et al. 2015).

Another intriguing phenomenon reported for $Ca_3Mn_2O_7$ was a novel magneto-electric (ME) coupling mechanism. $Ca_3Mn_2O_7$ crystallizes in the $Cmc2_1$ phase at low temperature and adopts a G-type spins order with the spins along $[001]_{pt}$ and a weak ferromagnetic spin canting along $[110]_{pt}$ (see Fig. 19b). The spins pointing along the $[001]_{pt}$ direction arises from the spin anisotropy, whereas the weak ferromagnetism (wFM) results from the DM interaction in which DM interaction-related energy can be expressed as $E_{DM} = \mathbf{D} \cdot (\mathbf{L} \times \mathbf{M})$ where $\mathbf{D}$, $\mathbf{L}$, and $\mathbf{M}$ are the DM interaction vector, AFM vector $\mathbf{L} = \mathbf{S}_1\text{-}\mathbf{S}_2$, and wFM vector $\mathbf{M} = \mathbf{S}_1 + \mathbf{S}_2$, respectively. The $\mathbf{D}$ vector is also proportional to approximately $\boldsymbol{r_1} \times \boldsymbol{r_2}$, where $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are the distance vectors defining the Mn-O bonds in the Mn-O-Mn subsystem as shown in Fig. 19b. It is then readily found by the right-hand rule that the DM vector induced by Q_T can produce wFM according to E_{DM}, which is along the direction perpendicular to both $\mathbf{D}$ and $\mathbf{L}$ vectors (see Fig. 19b). In contrast, the DM vector induced by Q_R is in the same direction as $\mathbf{L}$, which leads to a zero E_{DM} and no weak ferromagnetism. This rotation-control mechanism leads to on/off-control of wFM. Therefore, if an external electric field switches the direction of the polarization, then either Q_R or Q_T will be reversed due to the trilinear coupling interaction. Whether the rotation or tilt mode switches is mainly determined according to which switching path has the lower energy barrier (Benedek and Fennie
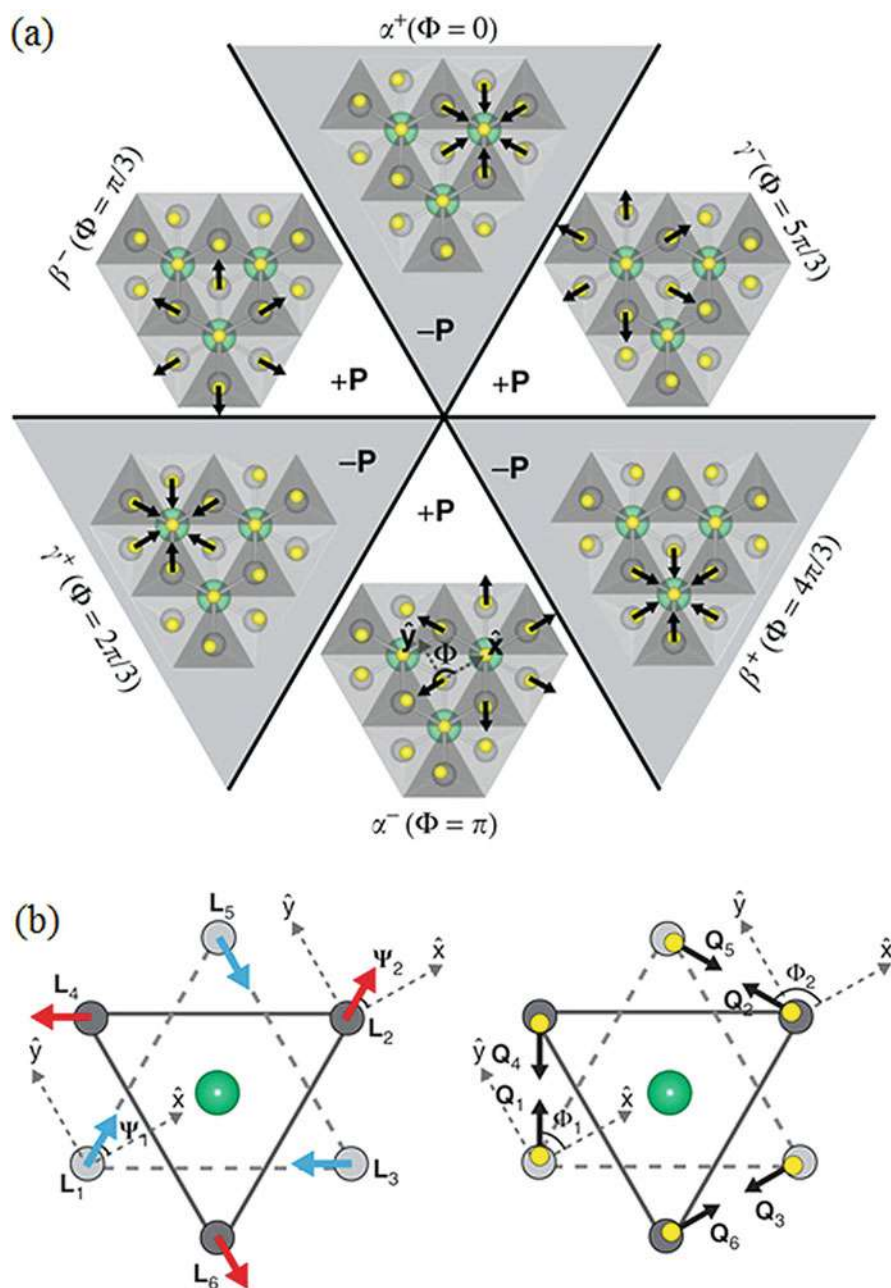
Fig. 17 (continued)

2011; Nowadnick and Fennie 2016). If the polarization switching favors switching of Q_T, the direction of the wFM will switch simultaneously because of the reversal of $\mathbf{D} \sim \mathbf{r_1} \times \mathbf{r_2}$, which corresponds to the octahedral tilt control route as shown in the bottom left panel of Fig. 19b.

Rondinelli and Fennie (2012) also reported hybrid improper ferroelectricity in 1:1 $ABO_3/A'BO_3$ superlattices based on the common *Pnma* structures found in ABO_3 and $A'BO_3$ components (Benedek and Fennie 2013). The basic idea here to produce hybrid improper ferroelectricity is to combine two components together and obtain the preferred criteria – coupling of two nonpolar lattice modes to a polar mode – for hybrid improper ferroelectricity. So, why are layered $(AA')B_2O_6$ perovskites and *Pnma* symmetry in each single perovskite component key features of a design strategy for hybrid improper ferroelectricity? To answer the question, the authors proposed a chemical criterion solely based on the symmetry consideration. As can be seen in the left panel of Fig. 20, in the ideal ABO_3 structure, there are inversion centers on both A and B sites. When there is A and A'-layered ordering, the inversion symmetry on the B sites is lifted, while the inversion symmetry on the A sites is kept. When there is B and B′-site-layered ordering, then the inversion symmetry on the A sites is lifted, while the inversion symmetry on B sites is kept. Furthermore, the oxygen octahedral-rotational modes can lift the inversion symmetry on the A site and preserve the inversion symmetry on B sites; therefore, one should anticipate that the A-site-layered cation ordering combined with an additional displacive mode could lift inversion symmetry without requiring a proper ferroelectric mechanism. Ultimately, this combination would lead to hybrid improper ferroelectricity driven by octahedral rotations, which led to a proposed energetic criterion based on lattice dynamical stability. For example, if both the $00a^+$ and a^-a^-0 (in Glazer notation) rotational patterns condense or are at least energetically competitive in each bulk component, then the prototype trilinear coupling interaction inducing the hybrid improper ferroelectricity should arise. The tilt pattern design strategy relies on these two rotational patterns, which combine to produce the common *Pnma* symmetry found in bulk ABO_3 perovskites. Using these guidelines, several new hybrid improper ferroelectrics and multiferroics were predicted.

Building on the work of Rondinelli and Fennie, a similar ME coupling mechanism was predicted in ultrashort period 1:1 $BiFeO_3/LaFeO_3$ superlattices (Zanolli

---

Fig. 17 (**a**) Schematic illustration of the six possible structural domains. Each domain is labelled by its phase Φ, defined to be the (counter)clockwise angle of the displacement directions of the apical oxygen of the trimer distortions relative to those in the α^+ domain. **P** indicates the polarization. (**b**) Left panel: The spin vectors $\mathbf{L_i}$ (i = 1–6) on the magnetic ions for the I (light grey ions) and II (dark grey ions) layers. Right panel: The local trimer distortion vectors $\mathbf{Q_i}$ ($i = 1$–6) on the magnetic ions. Ψ_i represents the angle of the reference $\mathbf{L_i}$ vector in the ith layer ($i = 1,2$), and the relationship among the spin vectors is determined because of the $120°$ spin arrangement; Φ_i represents the angle of the reference $\mathbf{Q_i}$ vector in the ith layer ($i = 1,2$), and the relationship among the $\mathbf{Q}$ vectors is determined according to the symmetry of the K_3 trimerization mode. (Figure reproduced from Das et al. (2014), Copyright Springer Nature, 2014)
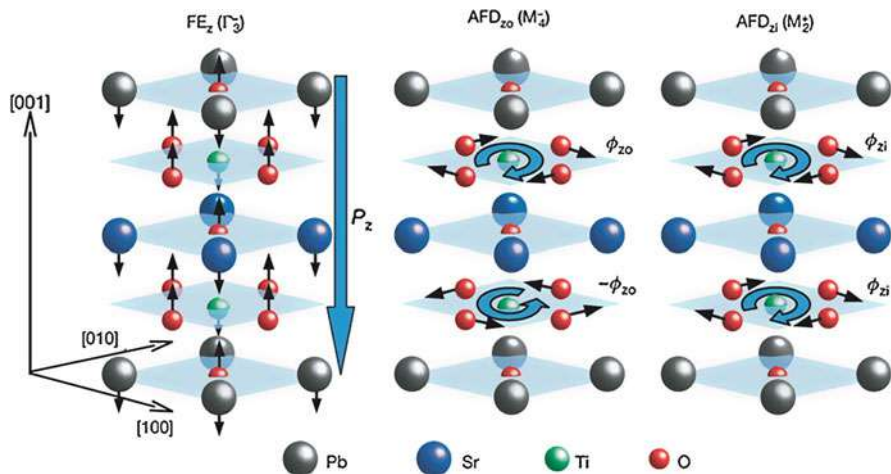
Fig. 18 A polar FE_z (Γ_3^- mode) producing a polarization P_z, AFD_{zo} (M_4^- mode) with out-of-phase oxygen octahedral rotation angle ϕ_{zo}, and AFD_{zi} (M_2^+ mode) with in-phase oxygen octahedral rotation angle ϕ_{zi}. (Figure reproduced from Bousquet et al. (2008), Copyright Springer Nature, 2008)

et al. 2013). Bulk $BiFeO_3$ crystallizes in the *R3c* structure, which competes with the *Pnma* phase (Dieguez et al. 2011). Bulk $LaFeO_3$ adopts the prescribed *Pnma* symmetry for its ground-state structure (Geller and Wood 1956). So, both compounds satisfy the proposed energetic criterion. Indeed, the 1:1 $BiFeO_3/LaFeO_3$ superlattice exhibits a polar *Pmc2₁* symmetry as predicted in Zanolli et al. (2013). Because both compounds have G-type magnetic structure in bulk, the superlattice was assumed to also exhibit G-type magnetic order. The hybrid improper ferroelectricity was confirmed by computing the energy as a function of the mode amplitudes upon freezing each mode separately in the high symmetry structure (i.e., *P4/mmm*) with density functional theory within the LDA. The spin alignment was predicted to be in plane and perpendicular to the polarization direction, which by symmetry allows for wFM controlled by Q_T along the out-of-plane direction as indicated by the form of E_{DM} and confirmed in the first-principles calculations. It is should be noted that there is also wFM controlled by Q_R lying in plane and along the polarization direction, but this kind of wFM cancels between layers and results in a null net weak ferromagnetism. Because the wFM in the 1:1 $BiFeO_3/LaFeO_3$ superlattice is controlled by the tilt mode, it is expected that if the switching path for the polarization reversal can be along a trajectory that changes Q_T, then the direction of the wFM can be simultaneously reversed. Furthermore, because both bulk compounds have high Néel temperatures (i.e., 643 K for $BiFeO_3$ and 750 K for $LaFeO_3$), the 1:1 $BiFeO_3/LaFeO_3$ superlattice was predicted to be room-temperature multiferroic.

A linear ME response was also predicted in 1:1 $LaFeO_3/LnFeO_3$ superlattices (Ln: lanthanide cation) (Ghosh et al. 2015) and they have the same multiferroic
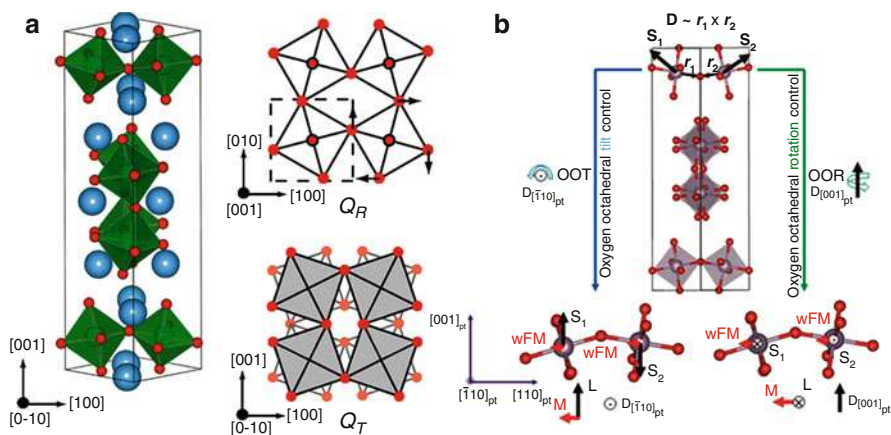
Fig. 19 (**a**) The $Cmc2_1$ ferroelectric ground-state structure of $Ca_3Mn_2O_7$. The large (blue) spheres are Ca ions. The schematic illustration of the atomic displacements corresponding to the X_2^+ (Q_R) rotation (top) in which the dashed square displays the unit cell of the $I4/mmm$ high-symmetry structure and the X_3^- (Q_T) tilt mode (bottom). The coordinate system is that of the $I4/mmm$ high-symmetry structure. Panel (**b**) shows the mechanisms for oxygen octahedral tilt (left, blue) and rotation (right, green) control of the direction of weak ferromagnetism. $\mathbf{S_i}$ (i = 1,2), $\mathbf{D_i}$ (i = 1,2), $\mathbf{r_i}$ (i = 1,2), $\mathbf{M}$, and $\mathbf{L}$ represent the local spin vectors, DM interaction vectors, unit vectors along the S_i-O bond direction, and the wFM and AFM vectors, respectively. The polar $Cmc2_1$ structure (center) is shown as reference without the A cations for clarity. (Figure (**a**) reproduced from Benedek and Fennie (2011), Copyright American Physical Society, 2011. Figure (**b**) reproduced from Lu and Rondinelli (2017), Copyright WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2017)

properties as those in the 1:1 $BiFeO_3/LaFeO_3$ superlattice, i.e., G-type magnetic order and hybrid improper ferroelectricity in $Pmc2_1$ structure. Because of the $m'2'm$ magnetic point group, there are α_{yz} and α_{zy} nonzero components in the magnetoelectric tensor. In addition, α_{zy} was proven to be one order-of-magnitude larger than α_{yz} with α_{zy} proportional to the magnitude of the electric polarization (see Fig. 21) but decreases with increasing m and n in $(LaFeO_3)_m/(LnFeO_3)_n$ period superlattices. Furthermore, the microscopic origin of the linear ME response on α_{zy} is due to changes of the DM interaction along the polarization direction induced by the applied electric field.

As clarified in Young et al. (2015), except the regular rotations such as the oxygen octahedral rotations coupled to the polar mode, there are also pseudo-rotations, which leads to the existence of many other types of improper ferroelectrics providing a new platform to study the multiferroics.

3.5   Asymmetric Multiferroics

To access the strong ME coupling in multiferroic materials, the concept of *asymmetric* multiferroics were proposed (Lu and Xiang 2014). Symmetric multiferroics are
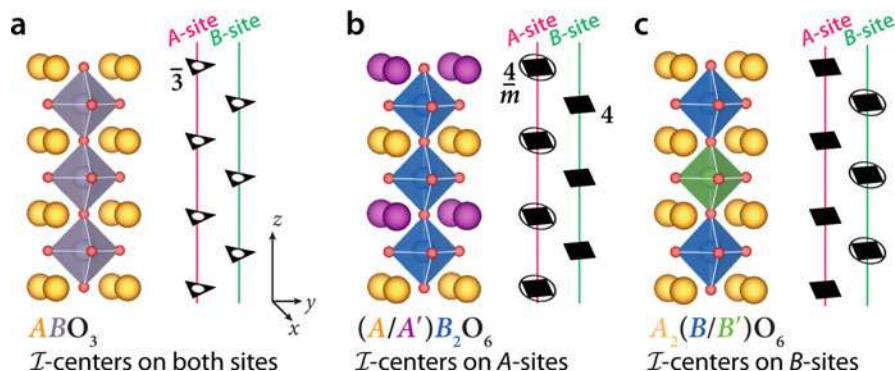
Fig. 20 (**a**) Bulk ABO_3 perovskites where inversion (I) centers are found on both A and B sites, and the highest site-symmetry operator is a threefold rotoinversion ($\bar{3}$). (**b**) A site cation ordering in layered perovskites, in which the inversion centers on the B site are broken leading to a fourfold rotation and remain on the A site through the $\frac{4}{m}$ operation. (**c**) B-site cation ordering in layered perovskites, in which the inversion centers on the A site are broken leading to a remaining fourfold rotation and $\frac{4}{m}$ operation on the B site. (Figure reproduced from Rondinelli and Fennie (2012), Copyright WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2012)
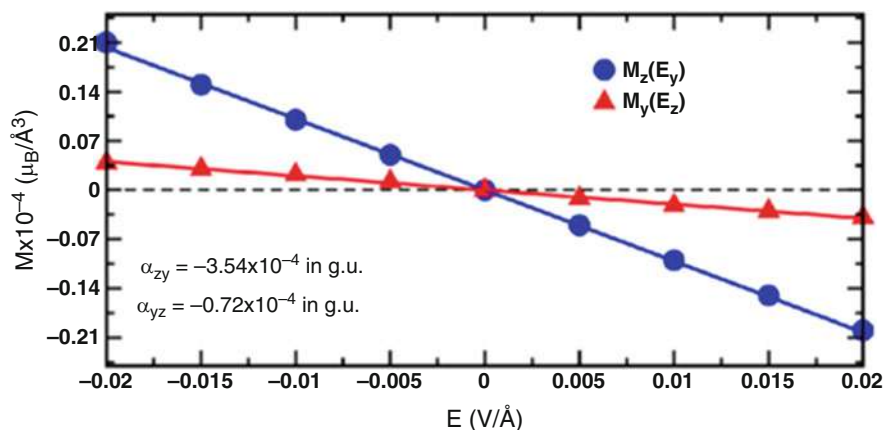
Fig. 21 Variation of magnetization with the electric field. Linear ME components α_{zy} and α_{yz} are computed from the slope of M_z vs. E_y and M_y vs. E_z curve, respectively. α_{zy} and α_{yz} are given in Gaussian units (g.u.). (Figure reproduced from Ghosh et al. (2015), Copyright American Physical Society, 2015)

those in which the two opposite electric polarization states are symmetry related, usually by an inversion center, and the magnetic states are also symmetrically equivalent (see Fig. 22). Multiferroics such as $BiFeO_3$ and $TbMnO_3$ fall within this classification. In contrast, asymmetric multiferroics exhibit two configurations with electric polarizations in opposite directions for each state but the polarization state of the phase configuration cannot be obtained from the other through the action
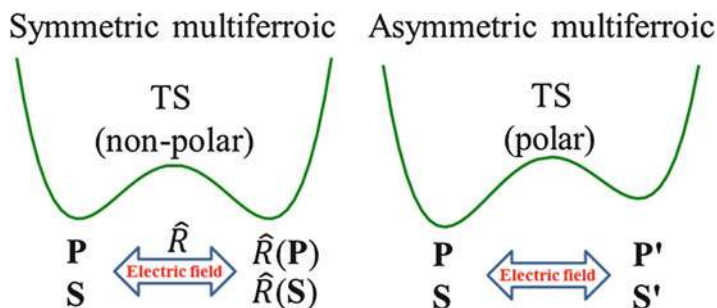
Fig. 22 Schematic illustration of symmetric multiferroic and asymmetric multiferroic. The transition state (TS) corresponds to a phase that has the highest energy in the switching path, and **P** and **S** represent the electric polarization and magnetic state, respectively. $\widehat{R}$ is the symmetry operation of point group. In a symmetric multiferroic, the two FE states and the associated magnetic properties are related by the symmetry operation of point group ($\widehat{R}$). (Figure reproduced from Lu and Xiang (2014), Copyright American Physical Society, 2014)

of any symmetry operation of the point group. In addition, the magnetic states can also differ (see Fig. 22). Furthermore, because the two polar state are inequivalent, the transition state between them is noncentrosymmetric, compared to the nonpolar transition state (TS) in symmetric multiferroics.

To explore the feasibility of such materials, computational studies based on DFT at the PBE level of theory with the $LiNbO_3$-type structure were performed. In this example, the high-temperature paraelectric structure adopts the $R\bar{3}c$ symmetry, and the low-temperature FE structure exhibits $R3c$ symmetry (see Fig. 23a). To obtain the two asymmetric polar states, the nominal paraelectric state ($R\bar{3}c$) was designed to be polar by substituting A with A′ or B with B′ to form an ordered $AA'BB'O_6$ compound. The **P** (P_{up}) state is inequivalent with **–P** (P_{dn}) state as shown in Fig. 23a. This strategy was further verified by performing DFT calculations on $AlA'FeCrO_6$ and $AlA'FeMoO_6$ series (A′ = Al,Ga,Sc,In,Y).

Here Fe^{3+}, Cr^{3+}, and Mo^{3+} have high-spin configurations and magnetic moments $5\mu_B$, $3\mu_B$, and $3\mu_B$, respectively. Thus because of the AFM spin-exchange interaction between nearest-neighboring Fe and Cr/Mo ions, the aforementioned compounds are ferrimagnets. As can be seen in Fig. 23b, when changing the states from **P** to **–P**, the J_1 spin-exchange interaction varies dramatically (see inset), leading to variable magnetic transition temperatures with a maximum difference as large as 300 K in $AlYFeMoO_6$. The large changes in the J_1 spin exchange interactions between the two states occur due to the large changes in the real-space exchange paths; for example, it is found that the Fe-O and Cr-O bond lengths are different for the **P** and **–P** states. This bond length difference influences the $\pi - \sigma$ couplings between the occupied Cr t_{2g} orbitals and unoccupied Fe e_g orbitals (see Fig. 23b). Finally, the large differences in the bond environment for the exchange paths leads to a much larger J_1 spin exchange interaction in the **–P** state. Therefore, if the **P** and **–P** states can be switched at a temperature that is lower than that of **–P**

Fig. 23 (continued)

state but higher than that of **P** state, then the reversal of the electric polarization will accompany the ferromagnetic-to-paramagnetic transition.

Although in this strategy the polarization and magnetism arise from different origins, where the electric polarization results from the A ions forming ionic or covalent bonds with the neighboring out-of-plane O ions and magnetism is derived from partially filled d-orbitals of the B-site cations, one may still obtain strong ME coupling. This behavior is different from other Type-I multiferroics with the polarization and magnetism also arising from different origins, such as $BiFeO_3$ and $BiMnO_3$. Lastly, the magnetic transition temperature in these asymmetric multiferroics can also be tuned by adjusting the B-site cations, which is well suited for a comprehensive electronic structure study.

3.6   Polymorphic Materials

Strain engineering is a useful tool to induce emergent phenomena, for example, strain-induced ferroelectricity (Schlom et al. 2014), which provides a route to design new multiferroics. Bulk $EuTiO_3$ is paraelectric antiferromagnetic. The magnetic transition temperature of the G-type AFM order arising from the Eu^{2+} ions occurs at 5.5 K (Katsufuji and Takagi 2001; Chien et al. 1974; McGuire et al. 1966). This material satisfies the design rules proposed by Fennie and Rabe (2006), which rely on an equilibrium material with an AFM and PE ground-state structure that hosts a low-frequency optical phonon mode (TO1 mode) that is exceeding sensitive to the magnetic order and applied epitaxial strain. Therefore, by applying compressive in-plane biaxial strain, it was found from DFT-PBE calculations that the frequency of the A_{2u} mode (split from the TO1 mode because of the in-plane strain) decreases for both AFM and FM orders. Specifically, when the strain reaches -0.92%, the frequency of the A_{2u} mode becomes zero for the FM state, while it remains finite for the AFM order. For strains at -1.2%, the frequency of the A_{2u} mode also becomes zero for the AFM spin configuration. This findings suggested that a large magnetocapacitive response could be found at -0.92% by applying an external magnetic field (see Fig. 24). Furthermore, a FM-to-FE phase should be realized at a strain value of -1.25% when the structure adopts a $P4mm$ symmetry with an

Fig. 23 (**a**) Top: the paraelectric $R\bar{3}c$ and ferroelectric $R3c$ structures. The arrow indicates the hexagonal c axis; Bottom: designing the asymmetric multiferroic by tuning the composition of the transition state (TS). The symmetry operations (C3, C2, i) are displayed. (**b**) Top: Cr-Fe π-σ coupling that is responsible for AFM coupling between a spin-down Cr^{3+} ion and a spin-up Fe^{3+} ion in $AlYFeCrO_6$. Inset: Comparison of the local structures and hopping parameters $t_{\pi\sigma}$ between two states with opposite polarizations (P_{dn} and P_{up}) for the spin interaction J_1 (see inset of the bottom figure). Bottom: Curie temperatures of the P_{dn} and P_{up} states in $AlA^{'}FeCrO_6$ and $AlA^{'}FeMoO_6$ systems (A$^{'}$ = Al, Ga, In, Sc, Y) from the MC simulations. Inset: Illustration of the superlattice and two different nearest-neighboring spin exchange interactions (J_1 and J_2). (Figure reproduced from Lu and Xiang (2014), Copyright American Physical Society, 2014)

Fig. 24 Compressive epitaxial strain (η) phase diagram for EuTiO$_3$ predicted from DFT. (Figure reproduced from Fennie and Rabe (2006), Copyright American Physical Society, 2006)
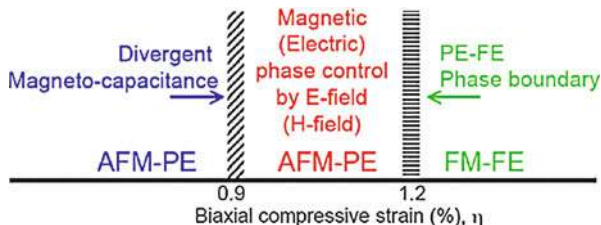
Table 2 Calculated lowest-frequency phonons, in cm^{-1}, of cubic SrMnO$_3$ with the calculated equilibrium lattice constants for G-AFM and FM orderings at the high-symmetry k points. Calculations performed at the DFT + U with $U = 2.7$ eV and $J = 1$ eV. Table reproduced from Lee and Rabe (2010), Copyright American Physical Society, 2010

	Γ	X	R	M
G-type AFM	121	116	84.5i	38.1i
FM	109i	113	119i	89.9i

electric polarization of ∼10μC/cm^2. Thus, a strong ME coupling could be expected at −1.25% strain (see Fig. 24), when the AFM-to-PE to FM-to-FE transitions occur by applying an external electric field along the polarization direction in FE phase (out-of-plane axis). Later, the FM-to-FE phase was confirmed by experiment with EuTiO$_3$ films grown under 1.1% tensile strain at 4.2 K (Lee et al. 2010).

SrMnO$_3$ is another interesting compound showing strong spin-phonon coupling. It is a PE and has a G-type magnetic order with the transition temperature of ∼233–260 K (Chmaissem et al. 2001; Takeda and Ohara 1974). There also exists a large frequency difference for the lowest-lying optical phonon mode between AFM and FM orders as shown in Table 2. Therefore, the FM-FE phase was predicted to appear at either compressive (−2.9%) (*I4cm* with a polarization along [001]$_{pt}$) or tensile strains (3.4%) (*Ima*2 with a polarization along [110]$_{pt}$) by using a GGA-PBE + U form in the DFT calculations (see Fig. 25) (Lee and Rabe 2010). The polarizations can be ∼54μC/cm^2 for both strain states, and the magnetic temperatures are both near 92 K. Moreover, a recent experiment indeed found a FE SrMnO$_3$ phase under tensile strain (1.7%) with the polarization along [110]$_{pt}$ but with G-type AFM order (Becher et al. 2015). It should be noted that such spin-phonon coupling as obtained from DFT calculations can be quite sensitive to the choice of the exchange-correlation functional (Hong et al. 2012); therefore, when making material predictions on new compounds relying on this interaction, it is important to test the robustness of the spin coupling for various functionals.

In our recent work (Lu and Rondinelli 2017), we showed an alternative promising route to realize electric field controllable magnetism at room temperature in multiferroics based on the family of A$_3$B$_2$O$_7$-layered perovskites. In this A$_3$B$_2$O$_7$ compound, as mentioned in Sect. 3.4.2, the polar mode (P), oxygen octahedral *rotation* (OOR) (Q_R), and oxygen octahedral *tilt* (OOT) (Q_T) modes couple together by a trilinear interaction, PQ_RQ_T, to stabilize the polar ground-state structure (i.e., *Cmc*2$_1$ symmetry) and lead to the functional electric polarization $P \propto Q_RQ_T$. Our previous work demonstrated that the polarization can be completely suppressed

Fig. 25 Total energies of various perovskite $SrMnO_3$ polymorphs obtained by freezing in the specified mode(s). The energies for structures with ferromagnetic (FM) spin order are shown in red, G-AFM in blue, C-AFM in green, and A-AFM in violet. Vertical black broken lines at 1.0%, 2.0%, 2.8%, and 3.4% strain indicate phase boundaries separating the *Imma* (G-AFM), *Ima*2 (G-AFM), *Ima*2 (C-AFM), *Ima*2 (A-AFM), and *Ima*2 (FM) states. (Figure reproduced from Lee and Rabe (2010), Copyright American Physical Society, 2010)
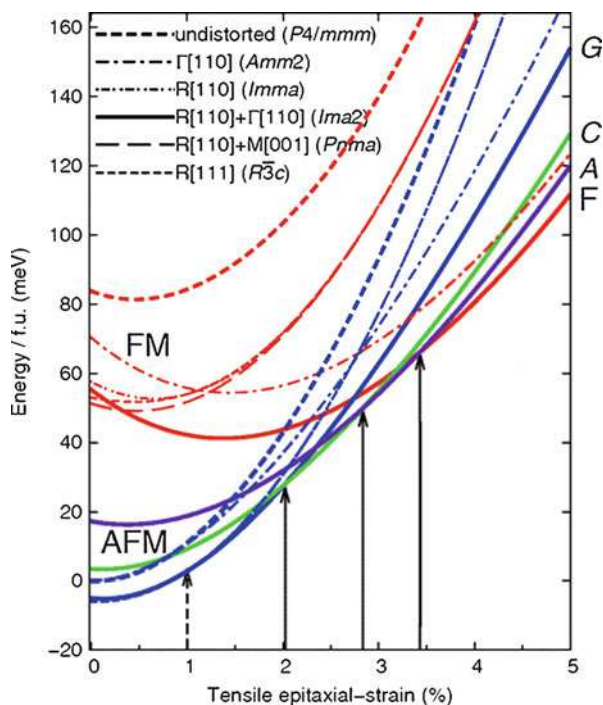
Table 3 Layered oxides exhibiting P-NP transitions in thin films. The cross indicates no NP phase appears in the relevant strain range. Table reproduced from Lu and Rondinelli (2017), Copyright WILEYVCH Verlag GmbH & Co. KGaA, Weinheim, 2017

	Ca_3Ti_2O7	$Sr_3Zr_2O_7$	$Sr_3Sn_2O_7$	$Ca_3Mn_2O_7$	$Ca_3Ge_2O_7$
Compressive NP	*Pbcn*	*Pbcn*	*Pbcn*	*Ccca*	*Ccca*
Tensile NP	*Pbcn*	$P4_2/mnm$	$P4_2/mnm$	×	×

by application of moderate compressive or tensile strains, i.e., a phase transition from a polar state to a nonpolar state (i.e., *Pbcn* symmetry) for $Ca_3Ti_2O_7$ (Lu and Rondinelli 2016). The strain-induced polar-to-nonpolar (P-NP) transition is a general phenomenon in the $A_3B_2O_7$ family (see Table 3). Remarkably, it arises because of the sensitivity of the trilinear coupling interaction that largely stabilizes the *polar* phase. We further showed that by interleaving two $n = 2\ A_3B_2O_7$ materials together, such that at least one exhibits a strain-induced P-NP transition, even more oxides can be made to exhibit the P-NP transition, specified as B-site cation-ordered $A_3BB'O_7$ films (see Fig. 26).

More interestingly, in the magnetic $A_3BB'O_7$ films, we demonstrated that the removal of the in-plane polarization across the tensile P-NP transition can be utilized to achieve electric field tunable weak ferromagnetism. This novel mechanism arises from a change in sense of the oxygen octahedral *rotation* across the tensile P-NP transition where the *in-phase* rotation in the polar phase changes to the *out-of-phase* rotation in the nonpolar phase (see Fig. 27). An addition requirement for
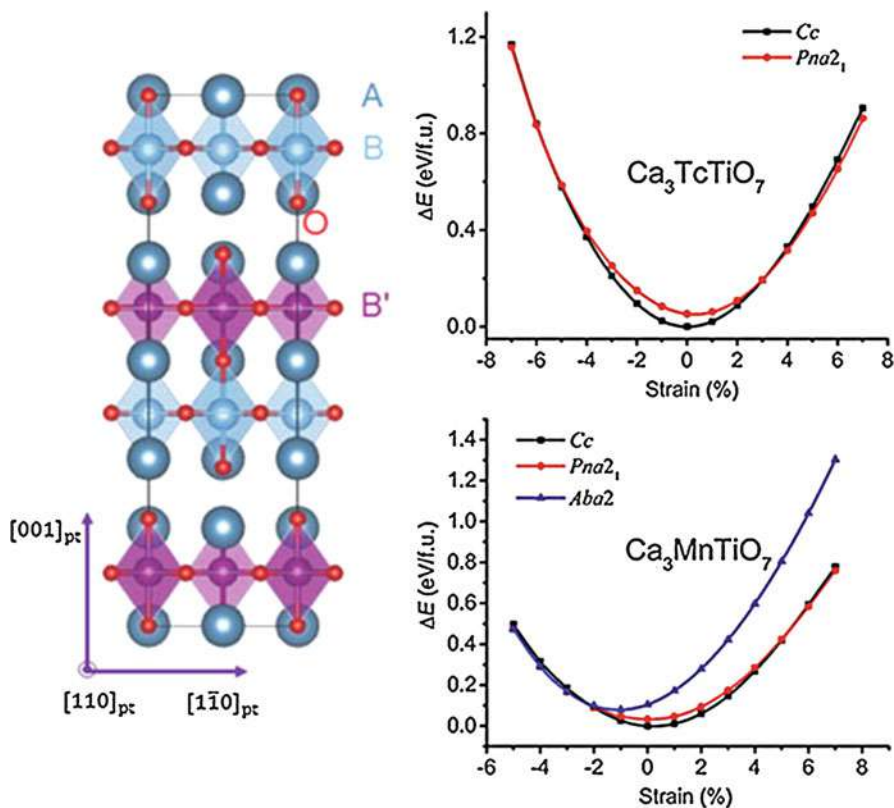
Fig. 26 Left: Illustration of the layered (001) B-cation order in the Ruddlesden-Popper $A_3BB'O_7$ compound. The high-symmetry structure, which adopts *I4/mmm* symmetry before substitution, undergoes a symmetry reduction to *I4mm* symmetry after substitution. The epitaxial plane is the (001) plane. Right-top and right-bottom show the calculated total energies as a function of biaxial strain for Ca_3TcTiO_7 and Ca_3MnTiO_7, respectively, where the energy is given relative to that of the *Cc* phase at 0% strain. (Figure reproduced from Lu and Rondinelli (2017), Copyright WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2017)

this effect is imposed by the rotation-tunable weak ferromagnetism direction, as shown in Fig. 19b. Either the direction arises as an intrinsic property of the compound or the direction is achieved by an external magnetic field through a spin-flop mechanism. We thus predicted with DFT-PBEsol calculations that the multiferroic cation-ordered Ca_3TcTiO_7 may be a viable candidate for switching-path-independent room-temperature electric field controllable magnetism.

3.7 Superlattices

An artificial superlattice consists of nanostructured materials, whereby each component in the nanostructure may exhibit physical properties different from its
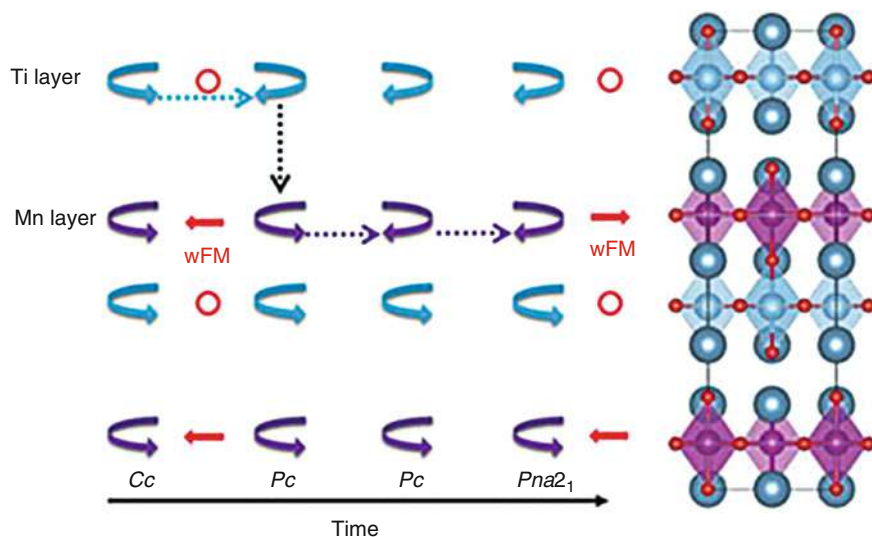
Fig. 27 Illustration of two-step transformation process by OOR at the (P-NP)$'$ phase transition boundary in Ca_3MnTiO_7. I and F indicate the Initial step and Final step in the switching process from the polar Cc (left, with both in- and out-of-plane polarization) phase to the polar $Pna2_1$ (right, with only out-of-plane polarization) phase through the polar Pc intermediate structure with different amplitude rotation angles. The blue and purple arrowed arcs indicate the BO_6 rotation sense in Ti and Mn perovskite layers, respectively. Red open circles and arrows indicate no wFM in Ti perovskite layers and the direction and magnitude of wFM in Mn perovskite layers, respectively. Dashed arrows with a time coordinate shown on the bottom track of the layers in which the sense of OOR changes during the process. (Figure reproduced from Lu and Rondinelli (2017), Copyright WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2017)

bulk analog. The superlattice may be considered as a new material with physical properties that are affected by interface effects (number density and proximity) such that emergent phenomena may appear (Lu et al. 2014). For example, as described in the Sect. 3.4.2, the nonpolar perovskite compounds with $Pnma$ structures can be used as "building blocks" to construct superlattices exhibiting hybrid improper ferroelectricity. However, the building blocks individually do not necessarily need to exhibit ferroelectricity. In this section, we discuss how one can use superlattices to design and discover new multiferroics as a general strategy, not only focusing on the hybrid improper ferroelectricity.

$LiOsO_3$ is a polar metal ($R3c$ symmetry) and nonmagnetic (Shi et al. 2013). $LiNbO_3$ is a ferroelectric insulator ($R3c$ symmetry) and nonmagnetic (Sim and Kim 2014; Giovannetti and Capone 2014). Recently, by carrying out DFT + U calculations (PBE exchange-correlation functional), Puggioni et al. (2015) predicted that a 1:1 $LiOsO_3/LiNbO_3$ superlattice transforms from a nonmagnetic metal into a Mott multiferroic (Pc symmetry) with a finite band gap through correlation effects. Both transition temperatures for the polar and G-type magnetic orders were predicted to be higher than room temperature (927 K and 379 K, respectively),

and the multiferroic phase was calculated to have an electric polarization of approximately $41.2\mu C/cm^2$ by the Berry phase method.

Hexagonal $LuFeO_3$ thin films are multiferroic with coexisting electric polarization (transition temperature, 1050 K) and G-type magnetic order (transition temperature, 440 K) at room temperature (Wang et al. 2013). The origin of the ferroelectricity is the same as that in $YMnO_3$ as discussed in Sect. 3.4. A related ferrate is $LuFe_2O_4$, which undergoes a three-dimensional charge ordering below 330 K, where the ratio of Fe^{2+} and Fe^{3+} in one triangular layer comprised of corner-sharing FeO_5 trigonal bipyramids is 1:2 (A layer), while it is 2:1 in the other triangular layer (B layer) (see Fig. 5b). Ferrimagnetic order sets in below 240 K.

A superlattice comprised of $LuFeO_3$ and $LuFe_2O_4$ has been synthesized recently by oxide molecular beam epitaxy with m periods as $(LuFeO_3)_m/(LuFe_2O_4)_1$ (Mundy et al. 2016). Experimental assessment of the structure and properties shows that the $m = 9$ $(LuFeO_3)_9/(LuFe_2O_4)_1$ superlattice exhibits room-temperature ferrimagnetism and an electric polarization ($\sim 6\mu C/cm^2$). Note that upon decreasing m from 9 to 2, the Curie temperature, magnetization, and polarization of the superlattice also decrease. The authors also found that both the magnetic moment per iron and the magnetic transition temperature in $LuFe_2O_4$ were enhanced relative to the bulk values. The increased transition temperature can be ascribed to the influence of the ferroelectric trimer distortion (Q) on a $LuFe_2O_4$ layer, which was demonstrated by considering a "model $LuFe_2O_4$" structure in which the Q is artificially added into the original $LuFe_2O_4$ structure and mimics the magnetic transition temperature by using a "Ising-like" model with different Q magnitudes in Monte Carlo simulations. In the electronic structure calculations within the GGA-PBE, two charge-ordered structures were considered to understand the effect of Q: first, COI which has the A and B layers along the c direction with the stacking sequence of ABBA (net magnetization, $0.5\mu_B/Fe$), and second, COII with ABAB ($1.2\mu_B/Fe$). It should be noted that the charge ordering also affects crystal symmetry: COI results in the structure having nonpolar $C2/c$ symmetry (ground state), and COII is compatible with Cm symmetry (see Fig. 28a). Only the COII structure is compatible with the Q trimer distortion. As Fig. 28a shows, this distortion is strongly coupled to the magnetic state, i.e., with increasing Q, the magnetic transition temperature in COII increases. Thus, it can be understood why the superlattice exhibits an enhanced magnetic ordering temperature. It results from the Q mode that arises when the superlattice transforms into the COII phase. Although the enhancement of the local magnetic moment remains unclear, two situations were presented: charge transfer from the $LuFe_2O_4$ layers to $LuFeO_3$ layers (see Fig. 28b) and/or charge ordering changes upon going from COI to COII. In any case, both situations are connected with the Q mode-induced structural transition.
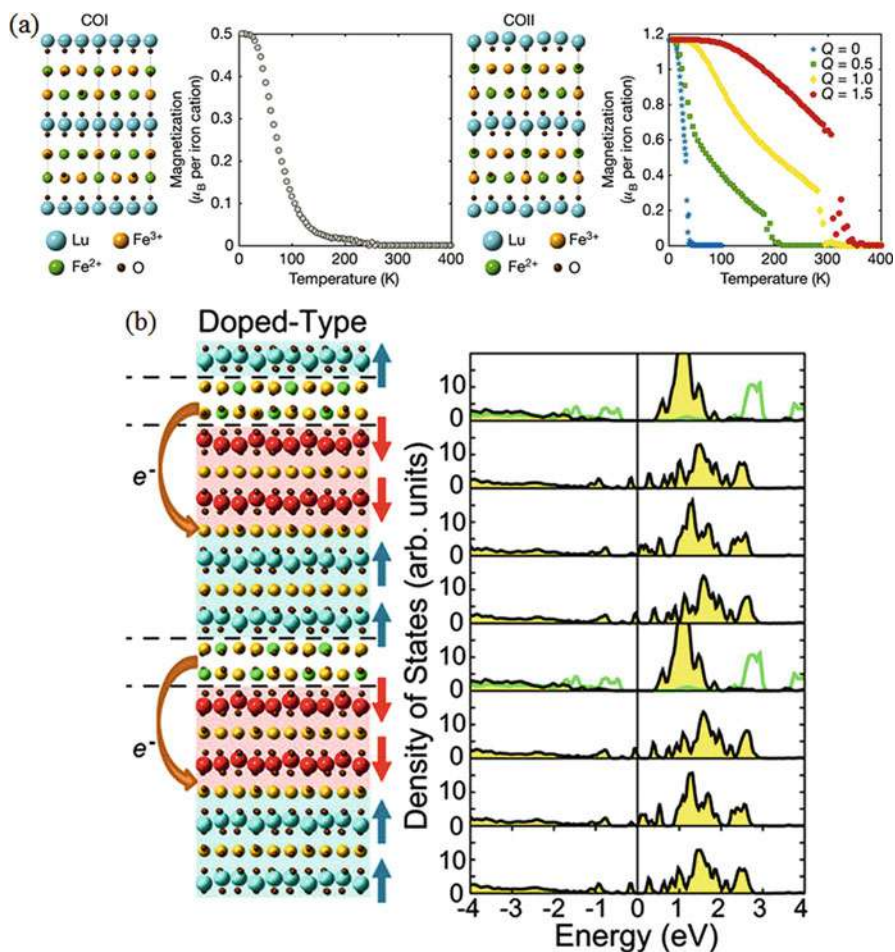
Fig. 28 (**a**) Monoclinic nonpolar $C2/m$ structures of bulk $LuFe_2O_4$ with an antiferroelectric charge-ordered (COI) state and the polar Cm structure with the ferroelectric charge-ordered (COII) state. The saturated magnetization per iron cation calculated as a function of temperature for each CO configuration is shown on the right. For the COII configuration, the temperature-dependent saturated magnetization per iron cation is also calculated as a function of Q, the amplitude of the trimer distortions with respect to the high-symmetry $R\overline{3}m$ structure. (**b**) Structure of the interface-doped $(LuFeO_3)_3/(LuFe_2O_4)_1$ superlattice. Electron transfer occurs from the $LuFe_2O_4$ layers to the $LuFeO_3$ layers in the structure (orange arrows). Lu^{3+} are shown in turquoise or red depending on the direction of the ferroelectric polarization, and Fe^{3+}, Fe^{2+}, and O^{2-} are represented by yellow, green, and brown colors, respectively. The DFT-projected density-of-states for Fe^{3+} and Fe^{2+} are plotted in yellow and green, respectively. (Figure reproduced from Mundy et al. (2016), Copyright Springer Nature, 2016)

3.8   Ferromagnetic-Multiferroic Heterostructure

$BiFeO_3$ as discussed has attracted enormous attention, because of its above-room-temperature ferroelectric and magnetic transition temperatures and large electric polarization ($\sim$100μC/cm^2 in thin film). By further investigating the FE and AFM domains in $BiFeO_3$ films (Zhao et al. 2006; Chu et al. 2008; Heron et al. 2014, 2011; Ederer and Spaldin 2005), several researchers have found that the polarization along the rhombohedral [111] direction can be switched by 180°, 109°, or 71°, upon which the direction of the AFM vector may also change by 90° (see Fig. 29a). For example, if the AFM vector is along [110], a 71° polarization switching from [1$\bar{1}$1] to [111] may lead to a 90° rotation of the AFM vector defining the easy plane to be along the [1$\bar{1}$0] direction as depicted in Fig. 29b. Then the wFM direction formerly along [$\bar{1}\bar{1}$2] when the AFM vector is along [110] will also rotate through the 71° FE domain switching. As in the multidomain structure, two domains linked by 71° FE domain wall have a net in-plane polarization, which can be reversed by applying an in-plane electric field. This geometry will result in the simultaneous 180° reversal of the net wFM because of the coupled FE and AFM domain structures (see Fig. 29b) (Heron et al. 2011).

Very recently, it was reported that an out-of-plane electric field also induces a 180° reversal of the wFM that includes a so-called two-step switching path in the FE domain transition process (Heron et al. 2014). Initially, the polarization, AFM vector, and wFM vector are along [1$\bar{1}$1] [110] and [1$\bar{1}\bar{2}$] directions, respectively. Switching occurs first through a 71° FE domain with the polarization, AFM vector, and wFM vector transformed along the [$\bar{1}\bar{1}$1], [1$\bar{1}$0], and [$\bar{1}\bar{1}$2] directions. Next, a 109° FE domain switching occurs, and the polarization, AFM vector, and wFM vector are aligned along [$\bar{1}\bar{1}\bar{1}$], [110], and [$\bar{1}$12] directions, respectively. After the 71° and 109° domain switching paths, the electric polarization and wFM vector will be reversed 180° (see Fig. 30a).

In the two cases mentioned above, the strong ferromagnet CoFe film is deposited on the $BiFeO_3$ film to amplify the microscopic switching process, i.e., to discern changes induced by applying in-plane (Heron et al. 2011) and out-of-plane electric fields (Heron et al. 2014), respectively. The enhancement arises from the exchange coupling between the local CoFe moments and wFM in $BiFeO_3$ (see Fig. 30b for the out-of-plane electric field configuration, where the net $Co_{0.9}Fe_{0.1}$ magnetization reverses after the voltage is applied).

3.9   Ferromagnetic-Ferroelectric Heterointerfaces

Interface engineering may be combined to design new multiferroics with ME coupling by combining two bulk components which separately exhibit the targeted ferroic properties – strong ferromagnetism and strong ferroelectricity. (100) Fe/$BaTiO_3$ heterostructures like those shown in Fig. 31a were proposed to fulfill this concept and exhibit ME coupling (Duan et al. 2006). Charge transfer at the interface leads to a net magnetism on the interfacial Ti ions, and when the thickness of $BaTiO_3$ is increased to 4 unit cells, the $BaTiO_3$ layers exhibit ferroelectricity,
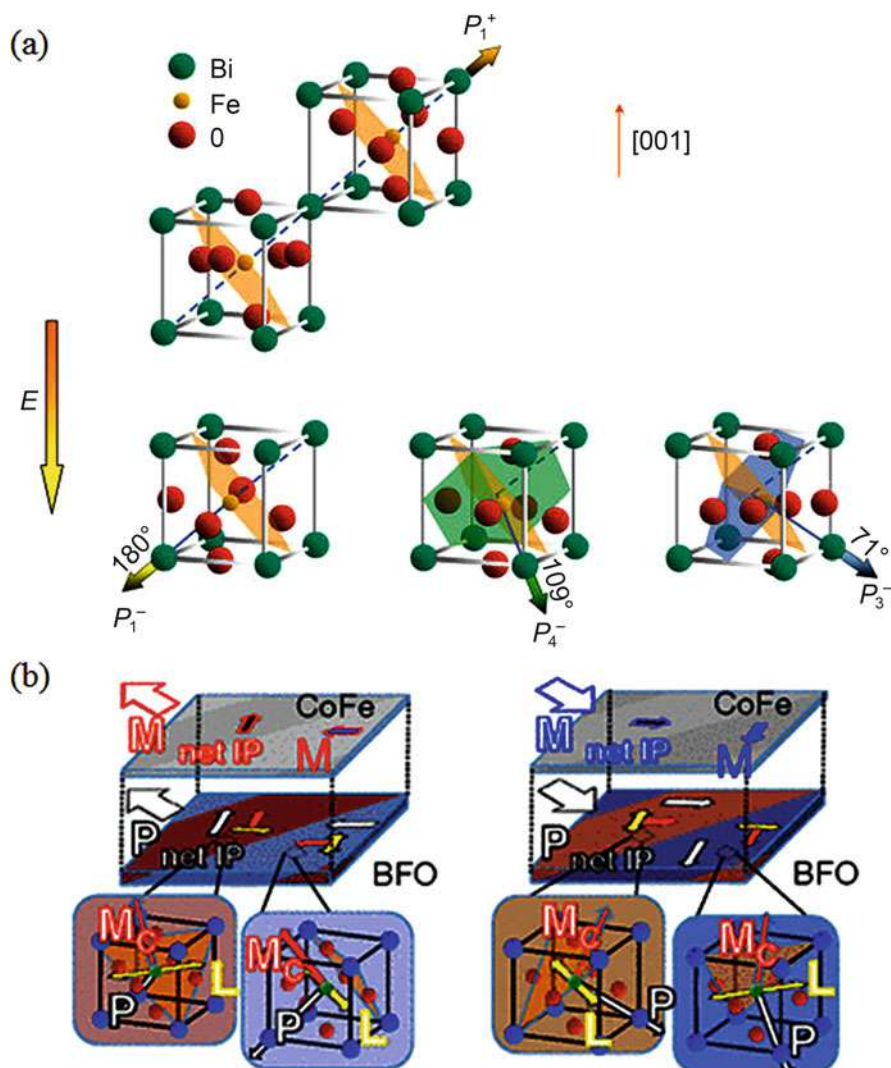
Fig. 29 (**a**) Possible 180°, 109°, and 71° polarization switching directions by applying an electric field. The antiferromagnetic plane changes from the orange plane to the green and blue planes upon 109° and 71° polarization switching, respectively. (**b**) Schematic description of the one-to-one magnetic interface coupling in the CoFe/BiFeO$_3$ heterostructure in the (left) as-grown state and (**b**) after the first electric pulse. (Figure (**a**) reproduced from Zhao et al. (2006), Copyright Springer Nature, 2006. Figure (**b**) reproduced from Heron et al. (2011), Copyright American Physical Society, 2011)

which results in an asymmetric distribution of the magnetic moments on both interfacial Fe and Ti ions. One thing to note is that the degree of charge transfer is dependent on the band alignment at the Schottky barrier formed between Fe (the metal) and BaTiO$_3$ (the ferroelectric). Owing to the underestimation of the
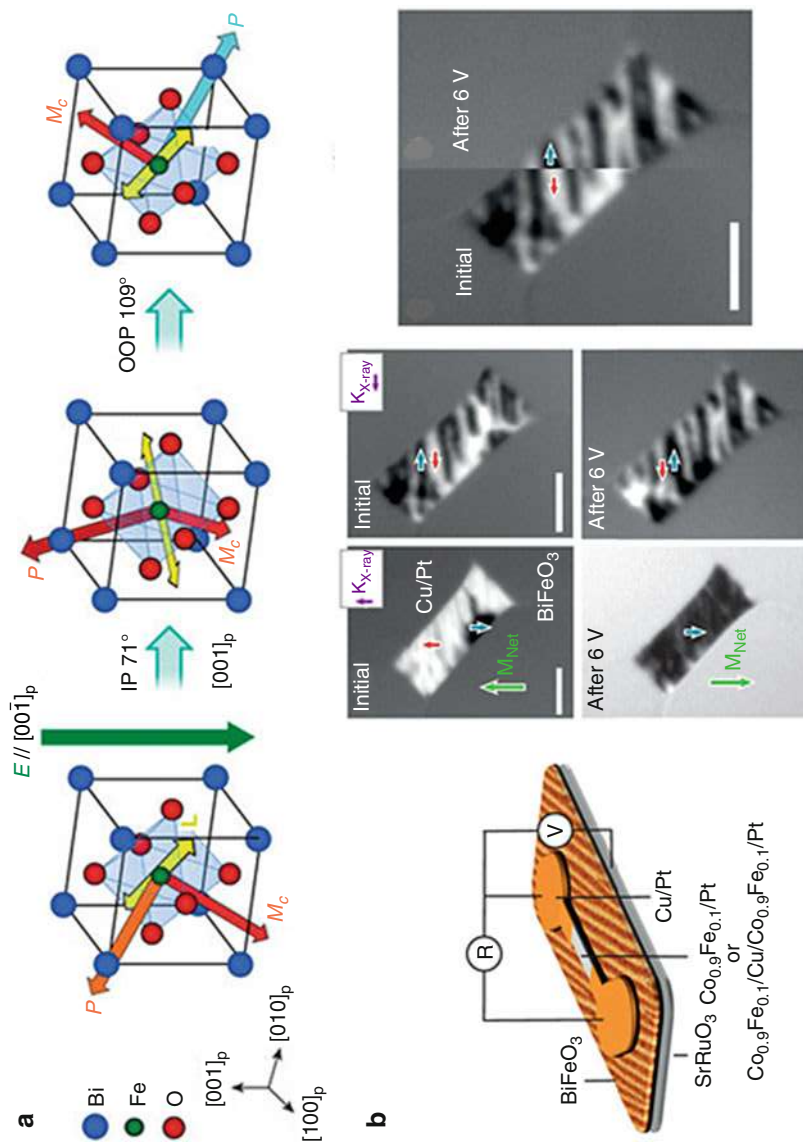
Fig. 30 (continued)

band gap from frequently used LDA and GGA density functionals in the study of multiferroic materials, spurious charge transfer can occur at multiferroic interfaces (Stengel 2011); therefore, when designing new heterostructures from DFT, one should carefully assess the functional-dependent band alignment.

The differences of the magnetic moments between the top and bottom interfaces per Fe and Ti ions are $0.06\mu_B$ and $0.22\mu_B$, respectively. The physical reason for this asymmetric charge distribution is due to the influence of the polar displacements in the $BaTiO_3$ layers on the minority-spin orbital hybridization between Fe 3d and Ti 3d. The polarization leads to shorter bond distances between Fe-O and Ti-O on the top interface than that on bottom interface, thus resulting in a strong orbital hybridization at the upper interface (see Fig. 31b). Therefore, it was proposed that an electric field could change the interfacial magnetic moments by coupling to the electric polarization, leading to a novel ME coupling mechanism.

In an experiment, it has now been found that there is an oxidized Fe layer at the (100) Fe/$BaTiO_3$ heterointerface (Radaelli et al. 2014), and the magnetization can be switched on/off upon reversing the electric field (see Fig. 31c). By carrying out first-principles calculations with the PBE exchange-correlation functional, the origin of this behavior has been attributed to change in the magnetization of the FeO layer in the model system (see Fig. 31d). It can transform from FM to AFM depending upon the direction of the polarization in the $BaTiO_3$ layer when the electronic correlations are sufficiently strong (see Fig. 31d). This is because in the polarization-down state, the Fe-O bond length shortens and the Fe-O-Fe bond angle tends to 180°, which favors the AFM spin alignment. Therefore, the on/off magnetization switching with the reversal of the electric field found at the room temperature in experiment is due to an electric polarization-mediated spin-exchange interaction within the oxidized Fe layers.

3.10   Domains and Domain Walls

As discussed for $BiFeO_3$, the FE and AFM domains can couple to each other and enable strong ME coupling. In this section, we focus on the properties of the domains and domain walls in $YMnO_3$ (Fiebig et al. 2002; Choi et al. 2010), which

---

Fig. 30 (**a**) Schematic depiction of the switching path for an applied electric field along the out-of-plane direction (i.e., $\left[00\bar{1}\right]_p$). (**b**) Left: Schematic description of a magnetoelectric device consisting of either $Co_{0.9}Fe_{0.1}$ or a $Co_{0.9}Fe_{0.1}/Cu/Co_{0.9}Fe_{0.1}$ spin valve on $BiFeO_3$. Middle: Initial and final (after applied 6 V) directions of the in-plane $Co_{0.9}Fe_{0.1}$ moments with components viewed perpendicular (vertical $K_{X\text{-ray}}$, where $K_{X\text{-ray}}$ defines the in-plane component of the incident X-ray beam) and parallel to the stripe domains (horizontal $K_{X\text{-ray}}$). The directions of the magnetizations in each domain are highlighted with blue and red arrows, which correspond to the local moment direction being perpendicular or parallel to $K_{X\text{-ray}}$. The net $Co_{0.9}Fe_{0.1}$ magnetization (green arrows) reverses after the voltage is applied. Right: $K_{X\text{-ray}}$ images of the initial and final states merged near the center of the $Co_{0.9}Fe_{0.1}$ reveal the magnetization reversal at each domain. (Figure reproduced from Heron et al. (2014), Copyright Springer Nature, 2014)
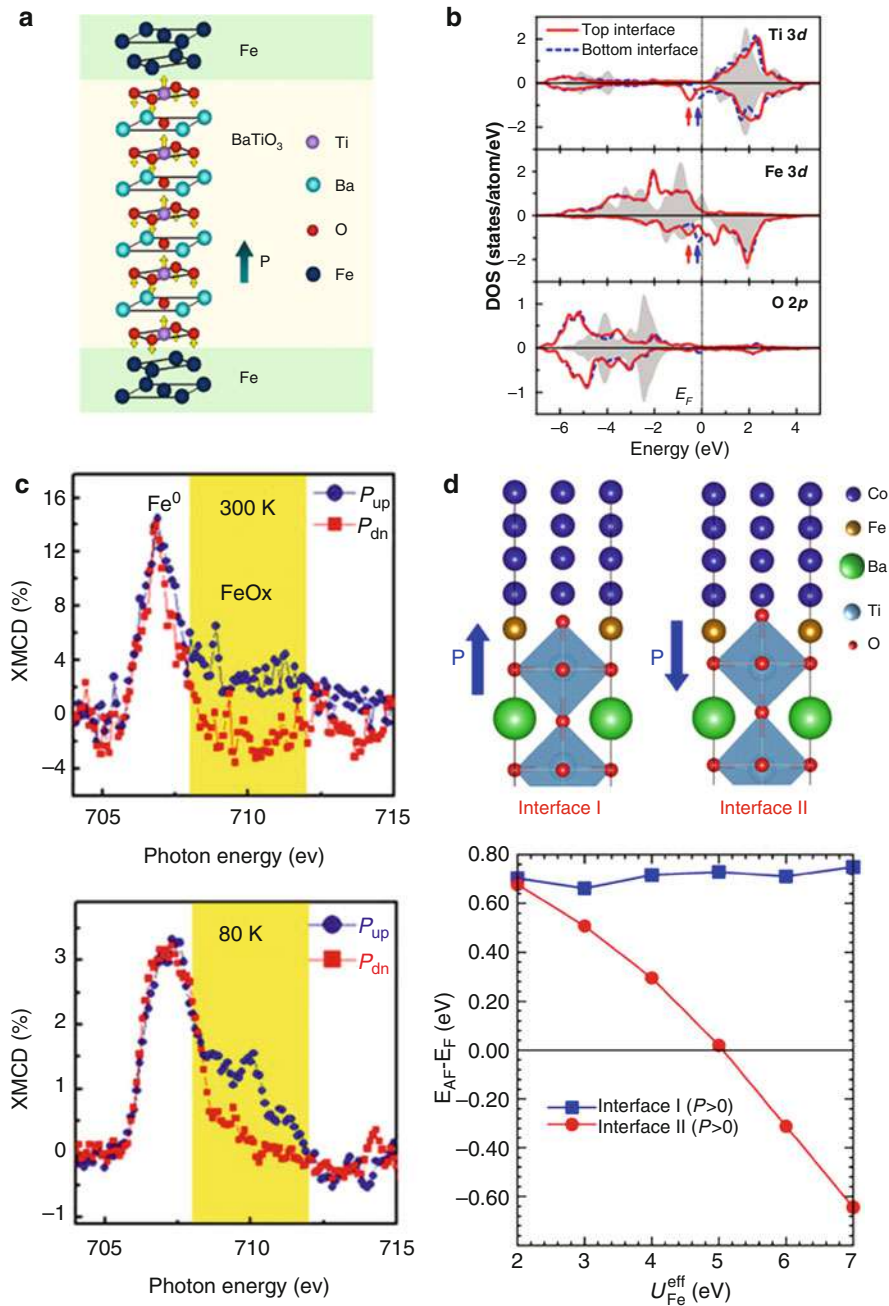
Fig. 31 (continued)

is another well-studied multiferroic material. The types of computational studies performed for this material can be applied to other multiferroics with coexisting FE and magnetic domain structures.

YMnO$_3$ is the improper ferroelectric that the ferroelectricity is caused by the dimerization K$_3$ mode through a coupling interaction $P_z Q_{K_3}^3 \cos(3\varphi)$, where φ is the azimuthal angle describing the direction of the tilt of the K$_3$ mode and describes the tilt distortion (i.e., K$_3$ mode) in each domain. Experimentally, the six domains are observed to be separated by $\sim 60^\circ$ in φ (Choi et al. 2010), resulting in the formation of antiphase (AP) DWs (see Fig. 32a). It was also demonstrated that these domains separated by $\sim 60^\circ$ in φ are also FE DWs, which indicate that the polarization in the two domains exhibit opposite polarization directions. Therefore, the experiments observed those DWs that are both ferroelectric and antiphase, which were also confirmed to be the lowest-energy DWs by first-principles calculations combined with a phenomenological Landau model analysis (Artyukhin et al. 2014).

Additional computational studies showed that there are three basic types of DWs in hexagonal perovskites (Kumagai and Spaldin 2013), which are the {210}, {120}, and {110} DWs as specified in the $P6_3cm$ unit cell (see Fig. 32a). The computed energies of the three types of DWs indicated that the energy of the {210} < {120} < {110} (see Fig. 32b) and the {110} DW can be regarded as the combination of the {210} and {120} DWs (see Fig. 32c). Furthermore, all the DWs are atomically sharp with effectively zero width, which contrasts that found in PbTiO$_3$, where the widths of the 180° DWs are larger than one unit cell (Meyer and Vanderbilt 2002; Lubk et al. 2009; Jia et al. 2008). The authors also showed that any DW with an arbitrary direction can be obtained by using a combination of the {210}, {120}, and {110} DWs through the relationship shown in Fig. 32d. The reason that only the AP + FE DWs were observed in experiment can be explained by noting that the FE-only and AP-only DWs are much higher in energy than the combined AP + FE DWs – they are almost equal to three and two times the average energy of the AP + FE DWs, respectively. This is because the FE-only and AP-only

Fig. 31 (**a**) The structure of Fe/(BaTiO$_3$)$_m$ multilayer for $m = 4$ (m is the number of the BaTiO$_3$ layers). Arrows indicate the displacements of Ti and O atoms in BaTiO$_3$ with the net polarization pointing up. (**b**) Orbital-projected DOS for the interfacial atoms of the structure shown in (**a**). The solid and dashed curves correspond to the DOS of atoms at the top and bottom interfaces, respectively. (**c**) XMCD signal in the Fe–L$_3$ energy region for BaTiO$_3$ polarization P$_{up}$ (blue circles) or P$_{dn}$ (red squares) taken on capacitors at 300 K and 80 K, respectively. The yellow shading indicates the region of the spectra corresponding to the oxidized Fe layer in contact with BaTiO$_3$. (**d**) Top: Schematic illustration of the interface regions in the calculations. From top to bottom: Co layer (blue spheres), FeO layer (light brown and red spheres for Fe and O, respectively), and BaTiO$_3$ layer (green, light blue, and red spheres indicate Ba, Ti, and O, respectively). The left (right) panel shows interface I (II), in which the polarization P$_{up}$ (P$_{dn}$) is pointing away from (towards) the BaTiO$_3$ layer. Bottom: Energy difference between FM and AFM spin configurations of the interfacial Fe atoms as a function of the Hubbard U parameter within the DFT + U method. (Figures (**a**) and (**b**) reproduced from Duan et al. (2006), Copyright American Physical Society, 2006. Figures (**c**) and (**d**) reproduced from Radaelli et al. (2014), Copyright Springer Nature, 2014)
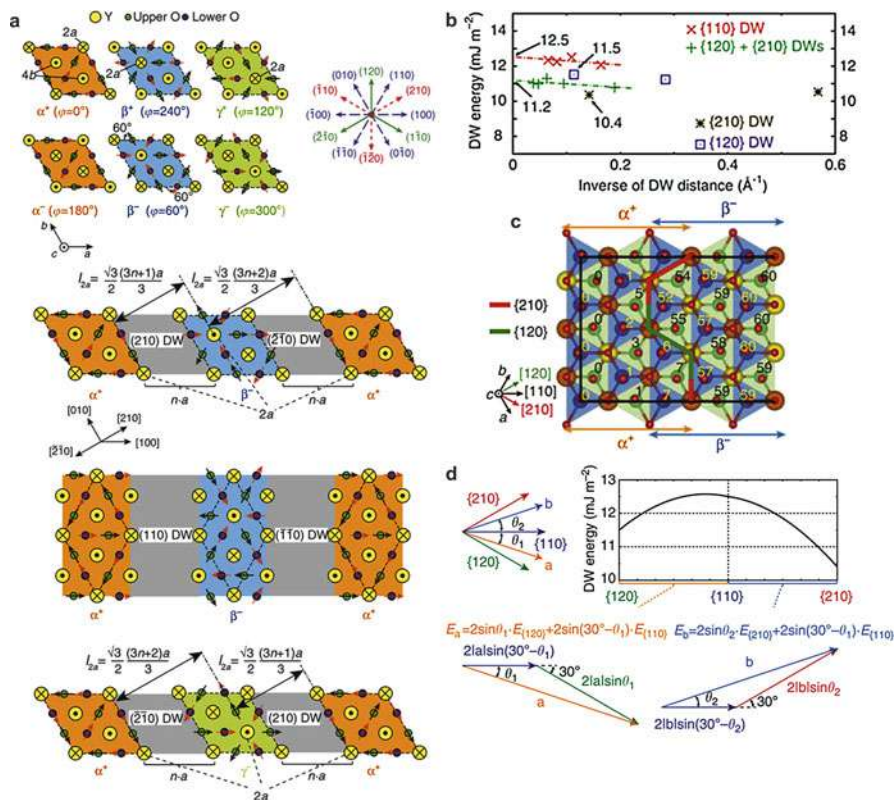
Fig. 32 (**a**) Schematic description of the six possible structural domains in h-YMnO$_3$ viewed down the c axis, parallel to the direction of ferroelectric polarization. The large yellow circles show the Y cation positions with crosses or dots indicating displacement into or out of the plane. The small circles are the adjacent oxygen ions at the apical vertices of the trigonal bipyramids with arrows indicating the direction of their displacements from the high-symmetry $P6_3/mmc$ phase. Each domain is labelled by its phase φ, defined to be the (counter)clockwise angle of the displacement directions of upper (lower) oxygen atoms relative to those in the α^+ domain. The (210) and $(\overline{2}10)$ domain walls (DWs) between α^+ and β^- domains are inequivalent as seen by comparing the spacing between the $2a$ Wyckoff positions, ℓ_{2a}. In contrast, the (110) and $(\overline{1}10)$ DWs are equivalent by symmetry. DWs between α^+ and γ^- domains are reversed relative to those between α^+ and β^- domains. (**b**) Energies for {110}, {120}, and {210} DWs calculated with different supercell sizes at the DFT + U with $U = 8$ eV and $J = 0.88$ eV. The dash-dotted lines are linear extrapolations to infinite spacing (zero inverse distance). The plus signs (+) indicate the averages of the {120} and {210} DW energies calculated with the supercells containing both DWs. (**c**) Calculated {110} DW structure for the sharp model. It can be seen that the {110} DW can be regarded as a combination of {120} and {210} DWs. (**d**) DW energy as a function of the DW direction, which is geometrically interpolated from the calculated {120}, {210}, and {110} DW energies, $E_{\{120\}}$, $E_{\{210\}}$, and $E_{\{110\}}$. (Figure reproduced from Kumagai and Spaldin (2013), Copyright Springer Nature, 2013)
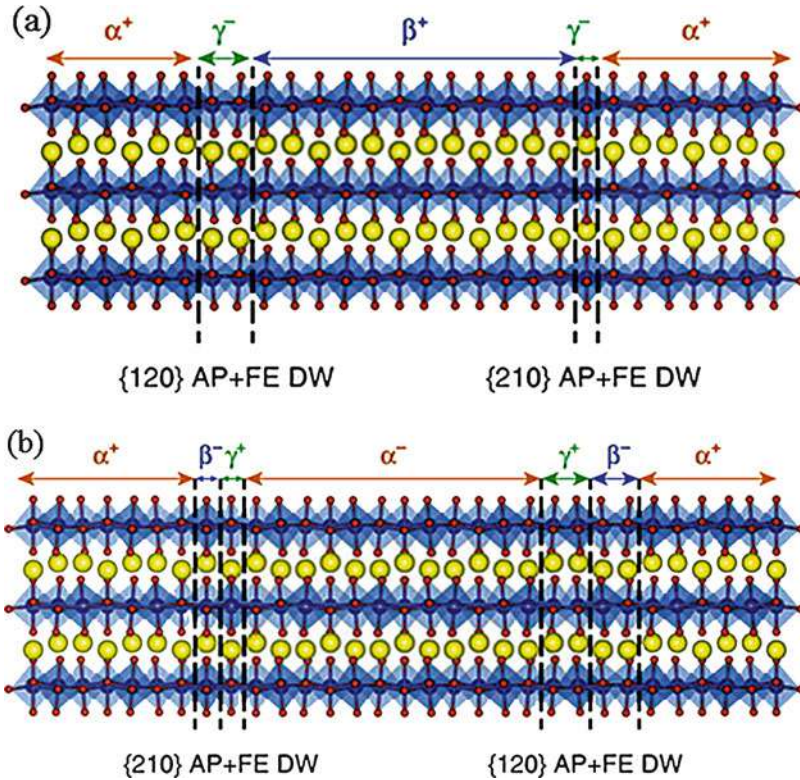
Fig. 33 Calculated (**a**) anti-phase (AP) and (**b**) ferroelectric (FE) DWs perpendicular to the [210] axis with 270- and 300-atom supercells, respectively. The relaxed structures locally preserve the up-up-down or down-down-up displacements of the Y ions across the DWs, resulting in the intermediate domains. (Figure reproduced from Kumagai and Spaldin (2013), Copyright Springer Nature, 2013)

DWs are the composite of the three and two AP + FE DWs, respectively, as can be seen in Fig. 33a and b.

By carrying out a model analysis based on an effective spin Hamiltonian shown in Equation (1), Das et al. (2014) obtained a trilinear coupling term $F_{tri} \propto (\boldsymbol{Q}_I \cdot \boldsymbol{L}_I + \boldsymbol{Q}_{II} \cdot \boldsymbol{L}_{II})M_z$, where $\boldsymbol{Q}_i$ and $\boldsymbol{L}_i$ ($i = I,II$) are order parameters for the K_3 mode having the direction of Φ and AFM vector in the ith layer (see Fig. 34a). According to F_{tri}, $\boldsymbol{L}_i$ will change as $\boldsymbol{Q}_i$ evolves across the DWs so that $\boldsymbol{Q}_i \cdot \boldsymbol{L}_i = 1$ or -1. For example, if $\boldsymbol{Q}_i$ is changed by $\sim 60°$ from α^+ to that in β^-, this produces $\boldsymbol{Q}_i \cdot \boldsymbol{L}_i = -1$ when $\boldsymbol{L}_i$ changes by $\left|\frac{\pi}{3}\right|$ (see Fig. 34b) and $\boldsymbol{Q}_i \cdot \boldsymbol{L}_i = 1$ when $\boldsymbol{L}_i$ changes by $\left|\frac{2\pi}{3}\right|$ (see Fig. 34c). In the first scenario, the weak ferromagnetism along the z direction, M_z, will not change because $\boldsymbol{Q}_i \cdot \boldsymbol{L}_i$ does not change, while in the second scenario, M_z reverses, thus leading to a simultaneous reversal of the electric polarization and wFM.
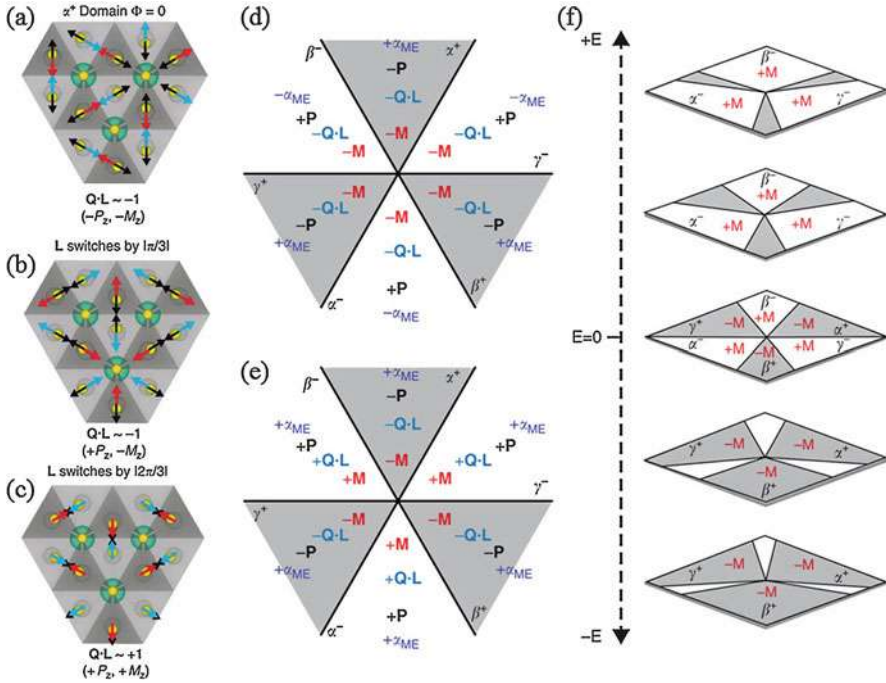
Fig. 34 (**a**) Initial equilibrium α^+ and **L** domains. (**b**) β^- domain; **L** rotates by $60°$, therefore $-M_z$ maps onto $-M_z$. (**c**) β^- domain; **L** rotates by $120°$, therefore $-M_z$ maps onto $+M_z$. (**d**) The change of the domain by $\frac{\pi}{3}$ and with a rotation of the AFM spin vectors, $\mathbf{L_I}$ and $\mathbf{L_{II}}$, by either $\frac{\pi}{3}$ or $-\frac{\pi}{3}$. (**e**) The change of the domain by $\frac{\pi}{3}$ and with a rotation of the AFM spin vectors, $\mathbf{L_I}$ and $\mathbf{L_{II}}$, by either $-\frac{2\pi}{3}$ or $\frac{2\pi}{3}$. (**f**) The schematic description of the electric field switching of the weak ferromagnetic domains. (Figure reproduced from Das et al. (2014), Copyright Springer Nature, 2014)

Upon considering the changes of $\mathbf{L}_i$ for all six domains arranged in the sequence of $\alpha^+ ->\beta^- ->\gamma^+ ->\alpha^- ->\beta^+ ->\gamma^-$, two scenarios were discussed: (1) $\mathbf{L}_i$ changes but $\mathbf{Q}_i \cdot \mathbf{L}_i$ remains unchanged (see Fig. 34d) and (2) $\mathbf{L}_i$ changes and $\mathbf{Q}_i \cdot \mathbf{L}_i$ reverses (see Fig. 34e). In the first case, there exists a net magnetization owing to the microstructure, which may be not favored. The second case is more likely to occur, because it avoids the formation of the free AFM DWs due to the existence of a net magnetization across the vortex shown in Fig. 34f. An electric field may induce a positive wFM or negative wFM according to the direction of the applied field because the polarization is coupled to the electric field (see Fig. 34f).

4 Perspectives on Multiferroic Materials Design

In this chapter, we summarized the most studied mechanisms active in multiferroic materials and tried to highlight how given a particular mechanism the ferroic properties can be enhanced. In many cases, electronic structure calculations based

on density functional calculations were pivotal in understanding the mechanism and in providing the basis for model formulation. In other cases, such computational methods enabled the discovery of new routes to multiferroism. In addition, there are still many other routes to provide multiferroicity, such as by considering organic substitution in the ABX_3 compounds in which A and X are organic ligands and B is a transition metal (Stroppa et al. 2011, 2013), and different concepts of multiferroicity (Van Aken et al. 2007; Zimmermann et al. 2014; Tolédano et al. 2015; Boström et al. 2018): ferrotoroidicity that has the form of $\mathbf{T} = \sum_i \mathbf{r}_i \times \mathbf{S}_i$ in which $\mathbf{T}$ and $\mathbf{r}_i$ are the toroidal moment and the position of the spin $\mathbf{S}_i$, respectively.

In metal-organic framework (MOF) compounds (Stroppa et al. 2013), ferro-electricity can also arise from a hybrid improper ferroelectric mechanism with the two primary modes being a Jahn-Teller distortion ($Q_{X_4^+}$) and the other one a rotational mode ($Q_{X_1^-}$) such that the polarization $P \propto Q_{X_4^+} Q_{X_1^-}$. In addition for magnetic MOF compounds, both $Q_{X_4^+}$ and $Q_{X_1^-}$ determine the amplitude of the wFM, whereas the direction of the wFM only depends on $Q_{X_4^+}$. In this way, strong ME coupling can be achieved because the polarization and wFM are coupled through the trilinear term $P Q_{X_4^+} Q_{X_1^-}$. It should be emphasized that in all the hybrid improper ferroelectrics, the ME coupling is dependent of the transition path, that is, how the polarization transforms from the $+P$ to $-P$ states. Only when the path includes unambiguously the reversal of the primary mode controlling the direction of the wFM is ME coupling be realized. Generally, this prerequisite is not easily determined by theory and required detailed experimentation. Recently a path-independent ME coupling mechanism was proposed in hybrid improper ferroelectric $A_3B_2O_7$ and $A_3BB'O_7$ compounds through an unusual polar-to-nonpolar phase transition arising from a change in the sense of the out-of-plane oxygen-rotational modes between the two phases. The latter modes are what control the direction of the wFM in each magnetic layer (Lu and Rondinelli 2017); this promising route, however, remains to be validated in experiment.

In ferrotoroidic materials, it is hard to detect the existence of the ferrotoroidicity, because its conjugate field, which is given by the cross product of the electric ($\mathbf{E}$) and magnetic ($\mathbf{H}$) field ($\sim \mathbf{E} \times \mathbf{H}$), can also contribute a linear ME coupling by each field separately. Therefore, such reports on this type of multiferroicity are rare (Van Aken et al. 2007; Zimmermann et al. 2014; Tolédano et al. 2015).

Although there has been much progress, finding or designing multiferroics with high-ordering temperatures, high polarizations, and strong magnetism with robust ME coupling at room temperature is still challenging. Often the magnetic ordering temperature is lower than room temperature, because of the small symmetric spin-exchange interaction and/or the spin frustration. Therefore, many studies have focused on $BiFeO_3$, because it is a room-temperature multiferroic, although other compounds hosting different mechanisms have been predicted (Zanolli et al. 2013; Lu and Xiang 2014; Lu and Rondinelli 2017; Zhao et al. 2014; Wang et al. 2015). Chemical A-O bonding (e.g., ionic and covalent bonding types) usually favor large polarization, but because A-site ferroelectricity relies on a different cation sublattice than the magnetic order (often B-site cation sublattice), these materials always tend to display weak ME coupling (Khomskii 2009). Although

some computational studies propose strong ME coupling in materials based on the $LiNbO_3$-type structure (Wang et al. 2015; Fennie 2008; Ederer and Fennie 2008; Varga et al. 2009; Giovannetti et al. 2016), the magnetoelectric response is based on changing the sign of the DM interaction vector; this sign change is difficult to access in experiment. In addition, a first-principles study proposed that the $LiNbO_3$-type compound Zn_2FeOsO_6 can be a multiferroic with large polarization and ferrimagnetism (Wang et al. 2015), but the strong ME coupling in the system is based on the same coupling behavior found between the AFM domain and FE domains in $BiFeO_3$. This prediction remains to be confirmed.

Despite the aforementioned challenges, the prospects for utilizing multiferroic compounds in several fields remains very much attractive, especially for photo-voltaic devices and in four-states memory devices (Lu et al. 2015a; Fiebig et al. 2016). In $BiFeO_3$, photo-induced voltages are relevant to the $71°$ DWs and were shown to increase by increasing the thickness of the samples, which effectively increase the total DWs (Yang et al. 2010). The enhanced response was attributed to the built-in electric field across the DWs that intrinsically separate the photo-generated electron-hole pairs. For four-state memory applications, $La_{0.1}Bi_{0.9}MnO_3$ films on the $SrTiO_3$ (001) substrates serves as an excellent example (Gajek et al. 2007). It is a FM-FE thin film, and experiment has demonstrated that magnetic and FE tunnel junctions can coexist even when the sample size is miniaturized down to 2 nm, which supports the coexistence of FM and FE orders. Based on the spin-filter effects between $La_{0.1}Bi_{0.9}MnO_3$ and $La_{2/3}Sr_{1/3}MnO_3$ electrodes separated by a $SrTiO_3$ spacer, a four-state cycle can be obtained by applying the magnetic and electric field in the magnetic junction owing to the charge screening at the electrode-barrier interfaces and its different spatial extension in the two electrodes comprising the FE junction.

In summary, we reviewed several mechanisms that may be utilized to design new multiferroic materials and described how density functional calculations can be used to assess the coexisting ferroic responses and phase stability. By considering both the scientific contributions that materials in this field have made to our understanding of how to harmonize (and couple) multiple ordered states and the continued technological importance of multiferroics, we are excited to see which materials will be designed next.

References

Aguado-Puente P, García-Fernández P, Junquera J (2011) Interplay of couplings between anti-ferrodistortive, ferroelectric, and strain degrees of freedom in monodomain $PbTiO_3/SrTiO_3$ superlattices. Phys Rev Lett 107:217601

Anisimov VI, Zaanen J, Andersen OK (1991) Band theory and Mott insulators: Hubbard U instead of Stoner I. Phys Rev B 44:943

Anisimov VI, Aryasetiawan F, Lichtenstein AI (1997) First-principles calculations of the electronic structure and spectra of strongly correlated systems: The LDA + U method. J Phys Condens Matter 9:767

Aoyama T et al (2014) Giant spin-driven ferroelectric polarization in $TbMnO_3$ under high pressure. Nat Commun 5:4927

Artyukhin S et al (2014) Landau theory of topological defects in multiferroic hexagonal manganites. Nat Mater 13:42–49

Ascher E, Rieder H, Schmid H, Stoessel H (1966) Some properties of ferromagnetoelectric Nickel-Iodine boracite, $Ni_3B_7O_{13}I$. J Appl Phys 37:1404

Astrov DN (1960) The magnetoelectric effect in antiferromagnetics. Sov Phys JETP 11:708–709

Baettig P, Seshadri R, Spaldin NA (2007) Anti-polarity in ideal $BiMnO_3$. J Am Chem Soc 129:9854

Becher C et al (2015) Strain-induced coupling of electrical polarization and structural defects in $SrMnO_3$ films. Nat Nanotechnol 10:661–665

Belik A, Takayama-Muromachi E (2006) Magnetic properties of $BiMnO_3$ studied with Dc and Ac magnetization and specific heat. Inorg Chem 45:10224

Belik A et al (2007) Origin of the monoclinic-to-monoclinic phase transition and evidence for the centrosymmetric crystal structure of $BiMnO_3$. J Am Chem Soc 129:971

Belik AA et al (2016) Low-temperature structural modulations in $CdMn_7O_{12}$, $CaMn_7O_{12}$, $SrMn_7O_{12}$, and $PbMn_7O_{12}$ Perovskites Studied by Synchrotron X-ray Powder Diffraction and Mössbauer Spectroscopy. J Phys Chem C 120:8278–8288

Benedek NA, Fennie C (2011) Hybrid improper ferroelectricity: A mechanism for controllable polarization-magnetization coupling. J Phys Rev Lett 106:107204

Benedek NA, Fennie CJ (2013) Why are there so few perovskite ferroelectrics? J Phys Chem C 117:13339

Benedek NA, Mulder AT, Fennie CJ (2012) Polar octahedral rotations: a path to new multifunctional materials. J Solid State Chem 195:11–20

Benedek NA et al (2015) Understanding ferroelectricity in layered perovskites: new ideas and insights from theory and experiments. Dalton Trans 44:10543–10558

Boström HLB, Senn MS, Goodwin AL (2018) Recipes for improper ferroelectricity in molecular perovskites. Nat Commun 9:2380

Bousquet E et al (2008) Improper ferroelectricity in perovskite oxide artificial superlattices. Nature 452:732–736

Cao K, Guo G-C, Vanderbilt D, He L (2009) First-principles modeling of multiferroic RMn_2O_5. Phys Rev Lett 103:257201

Ceperley DM, Alder B (1980) Ground state of the electron gas by a stochastic method. J Phys Rev Lett 45:566

Chapon LC et al (2004) Structural anomalies and multiferroic behavior in magnetically frustrated $TbMn_2O_5$. Phys Rev Lett 93:177402

Cheong S-W, Mostovoy M (2007) Multiferroics: A magnetic twist for ferroelectricity. Nat Mater 6:13–20

Chien C-L, DeBenedetti S, Barros F, De S (1974) Magnetic properties of $EuTiO_3$, Eu_2TiO_4, and $Eu_3Ti_2O_7$. Phys Rev B 10:3913

Chmaissem O et al (2001) Relationship between structural parameters and the Néel temperature in $Sr_{1-x}Ca_xMnO_3$ (0<~x<~1) and $Sr_{1-y}Ba_yMnO_3$ (y<~0.2). Phys Rev B 64:134412

Choi YJ et al (2008) Ferroelectricity in an Ising chain magnet. Phys Rev Lett 100:047601

Choi T et al (2010) Insulating interlocked ferroelectric and structural antiphase domain walls in multiferroic $YMnO_3$. Nat Mater 9:253–258

Chu YH et al (2008) Electric-field control of local ferromagnetism using a magnetoelectric multiferroic. Nat Mater 7:478–482

Das H et al (2014) Bulk magnetoelectricity in the hexagonal manganites and ferrites. Nat Commun 5:2998

de Groot J et al (2012) Charge order in $LuFe_2O_4$: an unlikely route to ferroelectricity. Phys Rev Lett 108:187601

Deslippe J et al (2012) Berkeley GW: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures. Comput Phys Commun 183:1269–1289

Di Domenico M, Eibschutz M, Guggenheim HJ, Camlibel I (1969) Dielectric behavior of ferroelectric BaMF$_4$ above room temperature. Solid State Commun 7:1119–1122

Diéguez O, Iñiguez J (2015) First-principles predictions of low-energy phases of multiferroic BiFeO$_3$. Phys Rev B 91:184113

Dieguez O, Gonzalez-Vazquez OE, Wojdeł JC, Iñiguez J (2011) First-principles predictions of low-energy phases of multiferroic BiFeO$_3$. Phys Rev B 83:094105

Duan C-G, Jaswal SS, Tsymbal EY (2006) Predicted magnetoelectric effect in Fe/BaTiO$_3$ multilayers: Ferroelectric control of magnetism. Phys Rev Lett 97:047201

Dzyaloshinskii IE (1960) On the magneto-electrical effect in antiferromagnets. Sov Phys JETP 10:628–629

Ederer C, Fennie CJ (2008) Electric-field switchable magnetization via the Dzyaloshinskii-Moriya interaction: FeTiO$_3$ versus BiFeO$_3$. J Phys Condens Matter 20:434219

Ederer C, Spaldin NA (2005) Weak ferromagnetism and magnetoelectric coupling in bismuth ferrite. Phys Rev B 71:060401(R)

Erenstein W, Mathur ND, Scott JF (2006) Scott: Multiferroic and magnetoelectric materials. Nature 442:759–765

Efremov DV, Van den Brink J, Khomskii DI (2004) Bond-versus site-centred ordering and possible ferroelectricity in manganites. Nat Mater 3:853–856

Ergönenc Z, Kim B, Liu P, Kresse G, Franchini C (2018) Converged GW quasiparticle energies for transition metal oxide perovskites. Phys Rev Mater 2:024601

Faqir H, Chiba H, Kikuchi M, Syono Y, Mansori M, Satre P, Sebaoun A (1999) High-temperature XRD and DTA studies of BiMnO$_3$ perovskite. J Solid State Chem 142:113–119

Fennie CJ (2008) Ferroelectrically induced weak ferromagnetism by design. J Phys Rev Lett 100:167203

Fennie CJ, Rabe KM (2005) Ferroelectric transition in YMnO$_3$ from first principles. Phys Rev B 72:100103(R)

Fennie CJ, Rabe KM (2006) Magnetic and electric phase control in epitaxial EuTiO$_3$ from first principles. Phys Rev Lett 97:267602

Fiebig M et al (2000) Determination of the magnetic symmetry of hexagonal manganites by second harmonic generation. Phys Rev Lett 84:5620

Fiebig M et al (2002) Observation of coupled magnetic and electric domains. Nature 419:818–820

Fiebig M, Lottermoser T, Meier D, Trassin M (2016) The evolution of multiferroics. Nat Rev Mater 1:16046

Fox DL, Tilley DR, Scott JF, Guggenheim H (1980) Magnetoelectric phenomena in BaMnF$_4$ and BaMn$_{0.99}$Co$_{0.01}$F$_4$. J Phys Rev B 21:2926

Gajek M et al (2007) Tunnel junctions with multiferroic barriers. Nat Mater 6:296–302

Geller S, Wood EA (1956) Crystallographic studies of perovskite-like compounds. I. Rare earth orthoferrites and YFeO$_3$, YCrO$_3$, YAlO$_3$. Acta Crystallogr 9:563–568

Ghosh S, Das H, Fennie CJ (2015) Linear magnetoelectricity at room temperature in perovskite superlattices by design. Phys Rev B 92:184112

Giovannetti G, Capone M (2014) Dual nature of the ferroelectric and metallic state in LiOsO$_3$. Phys Rev B 90:195113

Giovannetti G et al (2011) Dielectric properties and magnetostriction of the collinear multiferroic spinel CdV$_2$O$_4$. Phys Rev B 83:060402(R)

Giovannetti G et al (2016) Magnetoelectric coupling in the type-I multiferroic ScFeO$_3$. Phys Rev B 94:195116

Goian V et al (2012) Absence of ferroelectricity in BiMnO$_3$ ceramics. J Appl Phys 112:074112

Goodenough JB, Longo JM (1970) Landolt-Börnstein, numerical data and functional relationships in science and technology, New series, vol III.4. Springer, Berlin, p 126

Han TC, Chao HH (2010) Observation of large electric polarization in orthorhombic TmMnO$_3$ thin films. Appl Phys Lett 97:232902

Harris AB (2011) Symmetry analysis for the Ruddlesden-Popper systems Ca$_3$Mn$_2$O$_7$ and Ca$_3$Ti$_2$O$_7$. Phys Rev B 84:064116

Hemberger J et al (2007) Multiferroic phases of Eu$_{1-x}$Y$_x$MnO$_3$. Phys Rev B 75:035118

Heron JT et al (2011) Electric-field-induced magnetization reversal in a ferromagnet-multiferroic heterostructure. Phys Rev Lett 107:217202

Heron JT et al (2014) Deterministic switching of ferromagnetism at room temperature using an electric field. Nature 516:370–373

Higashiyama D, Miyasaka S, Tokura Y (2005) Magnetic-field-induced polarization and depolarization in $HoMn_2O_5$ and $ErMn_2O_5$. Phys Rev B 72:064421

Hill NA (2000) Why are there so few magnetic ferroelectrics? J Phys Chem B 104:6694–6709

Hill NA, Rabe KM (1999) First-principles investigation of ferromagnetism and ferroelectricity in bismuth manganite. Phys Rev B 59:8759

Hong J et al (2012) Spin-phonon coupling effects in transition-metal perovskites: A DFT+U and hybrid-functional study. Phys Rev B 85:054417

Hubbard J (1963) Electron Correlations in Narrow Energy Bands. Proc R Soc A Math Phys Eng Sci 276:238–257

Hur N, Park S, Sharma PA, Ahn JS, Guha S, Cheong S-W (2004a) Electric polarization reversal in a multiferroic material induced by magnetic fields. Nature 429:392–395

Hur N et al (2004b) Colossal magnetodielectric effects in $DyMn_2O_5$. Phys Rev Lett 93:107207

Hybertsen MS, Louie SG (1986) Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. Phys Rev B 34:5390

Ikeda N et al (2000) Charge frustration and dielectric dispersion in $LuFe_2O_4$. J Phys Soc Jpn 69:1526–1532

Ikeda N et al (2005) Ferroelectricity from iron valence ordering in the charge-frustrated system $LuFe_2O_4$. Nature 436:1136–1138

Ismailza IG, Kizhaev SA (1965) Determination of the curie point of the ferroelectrics $ymno_3$ and $ybmno_3$ High temperature X-ray examination of yttrium manganate and ytterbium. Sov Phys Solid State 7:236

Jardon C et al (1999) Experimental study of charge ordering transition in $Pr_{0.67}Ca_{0.33}MnO_3$. J Magn Magn Mater 196:475–476

Jia C-L et al (2008) Atomic-scale study of electric dipoles near charged and uncharged domain walls in ferroelectric films. Nat Mater 7:57–61

Johnson RD, Radaelli PG (2014) Diffraction studies of multiferroics. Annu Rev Mater Res 44:269–298

Johnson RD et al (2012) Giant improper ferroelectricity in the ferroaxial magnet $CaMn_7O_{12}$. Phys Rev Lett 108:067201

Johnson RD et al (2013) X-Ray imaging and multiferroic coupling of cycloidal magnetic domains in ferroelectric monodomain $BiFeO_3$. Phys Rev Lett 110:217206

Kagomiya I et al (2003) Lattice distortion at ferroelectric transition of YMn_2O_5. Ferroelectrics 286:167–174

Kato K, Iida S, Yanai K, Mizushima K (1983) Ferrimagnetic ferroelectricity of Fe_3O_4. J Magn Magn Mater 783:31–34

Katsufuji T, Takagi H (2001) Coupling between magnetism and dielectric properties in quantum paraelectric $EuTiO_3$. Phys Rev B 64:054415

Katsura H, Nagaosa N, Balatsky AV (2005) Spin current and magnetoelectric effect in noncollinear magnets. Phys Rev Lett 95:057205

Kenzelmann M et al (2005) Magnetic inversion symmetry breaking and ferroelectricity in $TbMnO_3$. Phys Rev Lett 95:087206

Khomskii DI (2006) Multiferroics: Different ways to combine magnetism and ferroelectricity. J Magn Magn Mater 306:1–8

Khomskii DI (2009) Classifying multiferroics: Mechanisms and effects. Physics (College Park, Md). 2:20

Kimura T et al (2003a) Magnetic control of ferroelectric polarization. Nature 426:55–58

Kimura T, Kawamoto S, Yamada I, Azuma M, Takano M, Tokura Y (2003b) Magnetocapacitance effect in multiferroic $BiMnO_3$. Phys Rev B 67:180401(R)

Kimura T, Lawes G, Goto T, Tokura Y, Ramirez AP (2005) Magnetoelectric phase diagrams of orthorhombic $RMnO_3$ (R=Gd, Tb, and Dy). Phys Rev B 71:224425

Kimura T, Lashley JC, Ramirez AP (2006) Inversion-symmetry breaking in the noncollinear magnetic phase of the triangular-lattice antiferromagnet $CuFeO_2$. Phys Rev B 73:220401(R)

King-Smith RD, Vanderbilt D (1993) Theory of polarization of crystalline solids. Phys Rev B 47:1651

Kiselev SV, Ozerov RP, Zhdanov GS (1963) Detection of magnetic order in ferroelectric $BiFeO_3$ by neutron diffraction. Sov Phys Dokl 7:742

Krukau AV, Vydrov OA, Izmaylov AF, Scuseria GE (2006) Influence of the exchange screening parameter on the performance of screened hybrid functionals. J Chem Phys 125:224106

Kumagai Y, Spaldin N (2013) Structural domain walls in polar hexagonal manganites. Nat Commun 4:1540

Kurumaji T et al (2011) Magnetic-field induced competition of two multiferroic orders in a triangular-lattice helimagnet MnI_2. Phys Rev Lett 106:167206

Landau LD, Lifshitz EML (1959) Electrodynamics of continuous media. Fizmatgiz, Moscow

Lee JH, Rabe KM (2010) Epitaxial-strain-induced multiferroicity in $SrMnO_3$ from first principles. Phys Rev Lett 104:207204

Lee JH et al (2010) A strong ferroelectric ferromagnet created by means of spin-lattice coupling. Nature 466:954–958

Liu XQ et al (2015) Hybrid improper ferroelectricity in Ruddlesden-Popper $Ca_3(Ti,Mn)_2O_7$ ceramics. Appl Phys Lett 106:202903

Lorenz B, Wang Y-Q, Chu C-W (2007) Ferroelectricity in perovskite $HoMnO_3$ and $YMnO_3$. Phys Rev B 76:104405

Lottermoser T et al (2004) Magnetic phase control by an electric field. Nature 430:541–544

Lu C, Hu W, Tian Y, Wu T (2015a) Multiferroic oxide thin films and heterostructures. Appl Phys Rev 2:021304

Lu X-Z, Rondinelli JM (2016) Epitaxial-strain-induced polar-to-nonpolar transitions in layered oxides. Nat Mater 15:951–955

Lu X-Z, Rondinelli JM (2017) Room temperature electric-field control of magnetism in layered oxides with cation order. Adv Funct Mater 27:1604312

Lu XZ, Xiang H (2014) Designing asymmetric multiferroics with strong magnetoelectric coupling. J Phys Rev B 90:104409

Lu XZ, Whangbo M-H, Dong S, Gong XG, Xiang H (2012) Giant ferroelectric polarization of $CaMn_7O_{12}$ induced by a combined effect of Dzyaloshinskii-Moriya interaction and exchange striction. J Phys Rev Lett 108:187204

Lu XZ, Gong XG, Xiang HJ (2014) Polarization enhancement in perovskite superlattices by oxygen octahedral tilts. Comput Mater Sci 91:310–314

Lu XZ, Wu X, Xiang HJ (2015b) General microscopic model of magnetoelastic coupling from first principles. Phys Rev B 91:100405(R)

Lubk A, Gemming S, Spaldin NA (2009) First-principles study of ferroelectric domain walls in multiferroic bismuth ferrite. Phys Rev B 80:104110

Malashevich A, Vanderbilt D (2008) First principles study of improper ferroelectricity in $TbMnO_3$. Phys Rev Lett 101:037210

McGuire TR et al (1966) Structure of $EuTiO_3$. J Appl Phys 37:981–982

Mercone S et al (2004) Anomaly in the dielectric response at the charge-orbital-ordering transition of $Pr_{0.67}Ca_{0.33}MnO_3$. Phys Rev B 69:174433

Meyer B, Vanderbilt (2002) Ab initio study of ferroelectric domain walls in $PbTiO_3$. Phys Rev B 65:104111

Mitsui T et al (1982) Landolt-Börnstein, numerical data and functional relationships in science and technology. New series. Ferroelectrics and Related Substancies. Non-Oxides. vol 16. Springer, Berlin

Montanari E, Righi L, Calestani G, Migliori A, Gilioli E, Bolzoni F (2005) Room temperature polymorphism in metastable $BiMnO_3$ prepared by high-pressure synthesis. Chem Mater 17:1765–1773

Montanari E et al (2007) Structural anomalies at the magnetic transition in centrosymmetric $BiMnO_3$. Phys Rev B 75:220101(R)

Moreira dos Santos A, Parashar S, Raju AR, Zhao YS, Cheetham AK, Rao CNR (2002a) Evidence for the likely occurrence of magnetoferroelectricity in the simple perovskite, $BiMnO_3$. Solid State Commun 122:49–52

Moreira dos Santos A, Cheetham AK, Atou T, Syono Y, Yamaguchi Y, Ohoyama K, Chiba H, Rao CNR (2002b) Orbital ordering as the determinant for ferromagnetism in biferroic $BiMnO_3$. Phys Rev B 66:064425

Moriya T (1960a) Anisotropic superexchange interaction and weak ferromagnetism. Phys Rev 120:91

Moriya T (1960b) New mechanism of anisotropic superexchange interaction. Phys Rev Lett 4:228

Mostovoy M (2006) Ferroelectricity in spiral magnets. Phys Rev Lett 96:067601

Mulder AT, Benedek NA, Rondinelli JM, Fennie CJ (2013) Turning ABO_3 antiferroelectrics into ferroelectrics: Design rules for practical rotation-driven ferroelectricity in double perovskites and $A_3B_2O_7$ Ruddlesden-popper compounds. Adv Funct Mater 23:4810–4820

Mundy JA et al (2016) Atomically engineered ferroic layers yield a room-temperature magneto-electric multiferroic. Nature 537:523–527

Nakajima T et al (2008) Electric polarization induced by a proper helical magnetic ordering in a delafossite multiferroic $CuFe_{1-x}Al_xO_2$. Phys Rev B - Condens Matter Mater Phys 77: 052401

Nakamura M, Tokunaga Y, Kawasaki M, Tokura Y (2011) Multiferroicity in an orthorhombic $YMnO_3$ single-crystal film. Appl Phys Lett 98:082902

Neaton JB, Ederer C, Waghmare UV, Spaldin NA, Rabe KM (2005) First-principles study of spontaneous polarization in multiferroic $BiFeO_3$. Phys Rev B 71:014113

Noda K et al (2006) Magnetic-field-induced switching between ferroelectric phases in orthorhombic-distortion-controlled $RMnO_3$. J Appl Phys 99:08S905

Nowadnick E, Fennie C (2016) Domains and ferroelectric switching pathways in $Ca_3Ti_2O_7$ from first principles. J Physical Review B 94:104105

Oh YS, Luo X, Huang F-T, Wang Y, Cheong S-W (2015) Experimental demonstration of hybrid improper ferroelectricity and the presence of abundant charged walls in $(Ca,Sr)_3Ti_2O_7$ crystals. Nat Mater 14:407–413

Paiera J, Marsman M, Hummer K, Kresse G (2006) Screened hybrid density functionals applied to solids. J Chem Phys 124:154709

Pearson RG (1975) Concerning Jahn-Teller effects. PNAS 72:2104–2106

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865

Perdew JP et al (2008) Generalized gradient approximation for solids and their surfaces. Phys Rev Lett 100:136406

Picozzi S, Stroppa A (2012) Advances in ab-initio theory of multiferroics materials and mechanisms: Modelling and understanding. Eur Phys J B 85:240

Picozzi S, Yamauchi K, Sanyal B, Sergienko IA, Dagotto E (2007) Dual nature of improper ferroelectricity in a magnetoelectric multiferroic. Phys Rev Lett 99:227201

Picozzi S, Yamauchi K, Sergienko IA, Sen C, Sanyal B, Dagotto E (2008) Microscopic mechanisms for improper ferroelectricity in multiferroic perovskites: A theoretical review. J Phys Condens Matter 20:434208

Puggioni D, Giovannetti G, Capone M, Rondinelli JM (2015) Design of a mott multiferroic from a nonmagnetic polar metal. Phys Rev Lett 115:087202

Radaelli PG et al (2009) Incommensurate magnetic structure of YMn_2O_5: A stringent test of the multiferroic mechanism. Phys Rev B 79:020404(R)

Radaelli G et al (2014) Electric control of magnetism at the $Fe/BaTiO_3$ interface. Nat Commun 5:3404

Ramesh R, Spaldin NA (2007) Multiferroics: Progress and prospects in thin films. Nat Mater 6:21–29

Resta R (1992) Theory of the electric polarization in crystals. Ferroelectrtics 136:51–55

Resta R (1994) Macroscopic polarization in crystalline dielectrics: The geometric phase approach. Rev Mod Phys 66:899–915

Rohlfing M, Louie SG (2000) Electron-hole excitations and optical spectra from first principles. Phys Rev B - Condens Matter Mater Phys 62:4927

Rondinelli JM, Fennie CJ (2012) Octahedral rotation-induced ferroelectricity in cation ordered perovskites. Adv Mater 24:1961–1968

Rondinelli JM, Poeppelmeier KR, Zunger A (2015) Research update: Towards designed functionalities in oxide-based electronic materials. APL Mater 3:080702

Schlom DG et al (2014) Elastic strain engineering of ferroic oxides. MRS Bull 39:118–130

Schmid H (1994) Multi-ferroic magnetoelectrics. Ferroelectrics 162:317–338

Seixas L, Rodin AS, Carvalho A, Castro Neto AH (2016) Multiferroic two-dimensional materials. Phys Rev Lett 116:206803

Seki S et al (2007) Impurity-doping-induced ferroelectricity in the frustrated antiferromagnet $CuFeO_2$. Phys Rev B - Condens Matter Mater Phys 75:100403(R)

Seki S, Onose Y, Tokura Y (2008) Spin-driven ferroelectricity in triangular lattice antiferromagnets $ACrO_2$ (A=Cu, Ag, Li, or Na). Phys Rev Lett 101:067204

Senn MS et al (2015) Negative thermal expansion in hybrid improper ferroelectric Ruddlesden-popper perovskites by symmetry trapping. Phys Rev Lett 114:035701

Sergienko IA, Dagotto E (2006) Role of the Dzyaloshinskii-Moriya interaction in multiferroic perovskites. Phys Rev B - Condens Matter Mater Phys 73:094434

Sergienko IA, Sen C, Dagotto E (2006) Ferroelectricity in the magnetic E-phase of orthorhombic perovskites. Phys Rev Lett 97:227204

Seshadri R, Hill NA (2001) Visualizing the role of Bi 6s 'lone pairs' in the off-center distortion in ferromagnetic $BiMnO_3$. Chem Mater 13:2892–2899

Shi Y et al (2013) A ferroelectric-like structural transition in a metal. Nat Mater 12:1024–1027

Sim H, Kim BG (2014) First-principles study of octahedral tilting and ferroelectric-like transition in metallic $LiOsO_3$. Phys Rev B 89:201107(R)

Solovyev IV, Dederichs PH, Anisimov VI (1994) Corrected atomic limit in the local-density approximation and the electronic structure of d impurities in Rb. Phys Rev B 50:16861

Sosnowska I, Neumaier TP, Steichele E (1982) Spiral magnetic ordering in bismuth ferrite. J Phys C Solid State Phys 15:4835–4846

Spaldin NA, Cheong S-W, Ramesh R (2010) Multiferroics: Past, present, and future. Phys Today 63:38–43

Stengel M (2011) Band alignment at metal/ferroelectric interfaces: Insights and artifacts from first principles. Phys Rev B 83:235112

Stroppa A, Picozzi S (2010) Hybrid functional study of proper and improper multiferroics. Phys Chem Chem Phys 12:5405–5416

Stroppa A, Marsman M, Kresse G, Picozzi S (2010) The multiferroic phase of $DyFeO_3$: An ab initio study. New J Phys 12:093026

Stroppa A et al (2011) Electric control of magnetization and interplay between orbital ordering and ferroelectricity in a multiferroic metal-organic framework. Angew Chem Int Ed 50:5847–5850

Stroppa A, Barone P, Jain P, Perez-Mato JM, Picozzi S (2013) Hybrid improper ferroelectricity in a multiferroic and magnetoelectric metal-organic framework. Adv Mater 25:2284–2290

Sun J, Xiao B, Ruzsinszky A (2012) Communication: Effect of the orbital-overlap dependence in the meta generalized gradient approximation. J Chem Phys 137:051101

Sun J et al (2013) Semilocal and hybrid meta-generalized gradient approximations based on the understanding of the kinetic-energy-density dependence. J Chem Phys 138:044113

Sun J, Ruzsinszky A, Perdew JP (2015) Strongly constrained and appropriately normed semilocal density functional. Phys Rev Lett 115:036402

Sun J et al (2016) Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. Nat Chem 8:831–836

Takeda T, Ohara SJ (1974) Magnetic structure of the cubic perovskite type $SrMnO_3$. Phys Soc Jpn 37:275

Teague JR, Gerson R, James WJ (1970) Dielectric hysteresis in single crystal $BiFeO_3$. Solid State Commun 8:1073–1074

Tolédano P et al (2015) Primary ferrotoroidicity in antiferromagnets. Phys Rev B 92:094431

Van Aken BB, Palstra TTM, Filippetti A, Spaldin NA (2004) The origin of ferroelectricity in magnetoelectric $YMnO_3$. Nat Mater 3:164–170

Van Aken BB, Rivera J-P, Schmid H, Fiebig M (2007) Observation of ferrotoroidic domains. Nature 449:702–705

Van den Brink J, Khomskii DI (2008) Multiferroicity due to charge ordering. J Phys Condens Matter 20:434217

Vanderbilt D, King-Smith RD (1993) Electric polarization as a bulk quantity and its relation to surface charge. Phys Rev B 48:4442

Varga T et al (2009) Coexistence of weak ferromagnetism and ferroelectricity in the high pressure $LiNbO_3$-type phase of $FeTiO_3$. Phys Rev Lett 103:047601

Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Can J Phys 58:1200–1211

Wang C, Guo G-C, He L (2007) Ferroelectricity driven by the noncentrosymmetric magnetic ordering in multiferroic $TbMn_2O_5$: A first-principles study. Phys Rev Lett 99:177202

Wang C, Guo G-C, He L (2008) First-principles study of the lattice and electronic structure of $TbMn_2O_5$. Phys Rev B 77:134113

Wang KF, Liu J-M, Ren ZF (2009) Multiferroicity: The coupling between magnetic and polarization orders. Adv Phys 58:321–448

Wang J et al (2003) Epitaxial $BiFeO_3$ multiferroic thin film heterostructures. Science 299:1719–1722

Wang PS, Ren W, Bellaiche L, Xiang H (2015) Predicting a Ferrimagnetic Phase of Zn_2FeOsO_6 with Strong Magnetoelectric Coupling. J Phys Rev Lett 114:147204

Wang PS, Lu XZ, Gong XG, Xiang HJ (2016) Microscopic mechanism of spin-order induced improper ferroelectric polarization. Comput Mater Sci 112:448–458

Wang W et al (2013) Room-temperature multiferroic hexagonal $LuFeO_3$ films. Phys Rev Lett 110:237601

Xiang HJ, Whangbo M-H (2007) Charge order and the origin of giant magnetocapacitance in $LuFe_2O_4$. Phys Rev Lett 98:246403

Xiang HJ, Wei S-H, Whangbo M-H, Da Silva JLF (2008) Spin-orbit coupling and ion displacements in multiferroic $TbMnO_3$. Phys Rev Lett 101:037209

Xiang HJ, Kan EJ, Wei S-H, Whangbo M-H, Gong XG (2011a) Predicting the spin-lattice order of frustrated systems from first principles. Phys Rev B 84:224429

Xiang HJ et al (2011b) General theory for the ferroelectric polarization induced by spin-spiral order. Phys Rev Lett 107:157202

Xiang H, Lee C, Koo H-J, Gong X, Whangbo M-H (2013a) Magnetic properties and energy-mapping analysis. Dalton Trans 42:823–853

Xiang HJ, Wang PS, Whangbo M-H, Whangbo, Gong XG (2013b) Unified model of ferroelectricity induced by spin order. Phys Rev B 88:054404

Yang SY et al (2010) Above-bandgap voltages from ferroelectric photovoltaic devices. Nat Nanotechnol 5:143–147

Young J, Stroppa A, Picozzi S, Rondinelli JM (2015) Anharmonic lattice interactions in improper ferroelectrics for multiferroic design. J Phys Condens Matter 27:283202

Zanolli Z, Wojdeł JC, Iñiguez J, Ghosez P (2013) Electric control of the magnetization in $BiFeO_3/LaFeO_3$ superlattices. Phys Rev B 88:060102(R)

Zhang G et al (2011) Multiferroic properties of $CaMn_7O_{12}$. Phys Rev B 84:174413

Zhang H-G, Ma X-C, Xie L (2015) The structural and magnetic properties of Sr-doped multiferroic $CaMn_7O_{12}$. Int J Mod Phys B 29:1550221

Zhao T et al (2006) Electrical control of antiferromagnetic domains in multiferroic $BiFeO_3$ films at room temperature. Nat Mater 5:823–829

Zhao HJ et al (2014) Near roomerature multiferroic materials with tunable ferromagnetic and electrical properties. Nat Commun 5:4021

Zhong W, Vanderbilt D, Rabe KM (1994) Phase transitions in $BaTiO_3$ from first principles. Phys Rev Lett 73:1861

Zhong W, Vanderbilt D, Rabe KM (1995) First-principles theory of ferroelectric phase transitions for perovskites: The case of $BaTiO_3$. Phys Rev B 52:6301

Zimmermann AS, Meier D, Fiebig M (2014) Ferroic nature of magnetic toroidal order. Nat Commun 5:4796

Strain Control of Domain Structures in Ferroelectric Thin Films: Applications of Phase-Field Method

49

Jian-Jun Wang and Long-Qing Chen

Contents

Abstract

Ferroelectric thin films have potential applications in many devices such as memories, microwaves, transduction sensors, actuators, photovoltaics, etc. The mesoscale domain structures and thus properties of ferroelectric thin films depend crucially on the amount of strain imposed upon by the underlying substrates. Phase-field method has been extensively applied to understanding the underlying physics of the experimentally observed domain structures and predicting their responses to external electrical, mechanical, thermal, and chemical stimuli. In this chapter, the fundamentals of the thin-film phase-field method and its applications in predicting the effects of strains on the phase transitions,

J.-J. Wang · L.-Q. Chen (✉)
Department of Materials Science and Engineering, The Pennsylvania State University, Pennsylvania, PA, USA
e-mail: wjj8384@gmail.com; lqc3@psu.edu

domain structures, and the domain switching are reviewed. The prospect of using phase-field method in microstructure design and property optimization for ferroelectric thin films is discussed.

1   Introduction

The advances in film-growth methods including the pulsed laser deposition (PLD) (Ramesh et al. 1990; Hubler 1992; Frey et al. 1994; Koster et al. 1999) and molecular beam epitaxy (MBE) (Chao and Arthur 1975; Spah et al. 1988; Schlom et al. 1988; Kwo et al. 1988; Eckstein and Bozovic 1995; Theis et al. 1998) have allowed the growth of high-quality epitaxial ferroelectric thin films. An in-plane biaxial strain is typically generated and imposed on the film via the epitaxial lattice mismatch. Compared with the ferroelectric bulk counterparts which would normally fracture under moderate strains such as $\sim\pm0.1\%$, epitaxial thin films can withstand much larger in-plane biaxial strains up to $\sim\pm3\%$ (Schlom et al. 2007, 2008, 2014; Martin and Rappe 2017). Owing to the strong coupling between the polarization and the lattice degrees of freedom in ferroelectrics, novel phases, physics, and properties may emerge in heavily strained thin films. For example, a stress-free $SrTiO_3$ crystal is paraelectric down to 0 K. However, when a (001)-oriented $SrTiO_3$ was epitaxially grown on a (110)-oriented $DyScO_3$ substrate which can impose a tensile biaxial strain of $\sim0.8\%$ on $SrTiO_3$ film, a room-temperature ferroelectricity was observed in the $SrTiO_3$ thin film (Haeni et al. 2004). The similar effect was observed in strained $BaTiO_3$ thin films which show enhanced room-temperature remnant polarization and ferroelectricity at temperatures much higher than the bulk Curie point (Choi et al. 2004). While a stress-free $BiFeO_3$ bulk crystal has a stable rhombohedral symmetry below the Curie temperature of 830 °C (Catalan and Scott 2009), a $BiFeO_3$ thin film on a $LaAlO_3$ substrate is subjected to a huge compressive strain ($\sim-4.5\%$) and was shown to exhibit a highly distorted tetragonal symmetry. As thickness of a $BiFeO_3$ film increases, it gradually loses the coherency strains, leading to the formation of a rhombohedral-tetragonal alternating "stripe-like" mixed-phase structure (Zeches et al. 2009; Li et al. 2015).

The thermodynamic analysis was normally performed to understand the strain effect on the phase transition of ferroelectric thin films (Pertsev et al. 1998; Roytburd 1998; Sheng et al. 2008; Shirokov et al. 2009; Qiu et al. 2010). It starts from the Landau phenomenological theory of a ferroelectric bulk crystal, wherein the Gibbs energy density is used to describe the total free energy of the bulk system with temperature, electric field, and the stress as independent variables. For studying the effect of strain on transition temperatures, it is actually more convenient to use strain rather than stress as the independent variable. Therefore, a Legendre transformation can be performed to obtain the Helmholtz free energy density, in which strain is the independent variable in the mechanical energy term. The stress components can be solved from the combination of thermodynamic property equations (first-order derivatives of the Gibbs free energy) and the mechanical boundary conditions of the epitaxial thin film. More specifically, the first-order derivative of the Gibbs free

energy with respect to a stress component is equal to the corresponding negative strain component. The in-plane strain components of the thin film are fixed by the film/substrate lattice mismatch due to interfacial coherency, and the out-of-plane stress components of the thin film are zero due to the traction-free film surface. Eventually, the equilibrium state and the thermodynamic properties of the thin film can be established by minimizing the total thermodynamic potential of the thin film with respect to the thermodynamic variables.

The thermodynamic analysis has been extensively employed to predict the strain-temperature phase diagrams of various ferroelectric thin films including $PbTiO_3$ (Pertsev et al. 1998; Zembilgotov et al. 2005), $PbZr_{1-x}Ti_xO_3$ (Pertsev et al. 2003; Li et al. 2003a; Qiu et al. 2010), $BaTiO_3$ (Emelyanov et al. 2002; Shirokov et al. 2007; Sheng et al. 2008), $BiFeO_3$ (Zhang et al. 2007; Ma et al. 2008; Karpinsky et al. 2017), etc. One advantage of the thermodynamic analysis is the easy implementation of the thermodynamic potential into MATHEMATICA or MATLAB programs. Using the "minimization function" built in the software, the thermodynamic potential can be conveniently minimized to obtain the equilibrium state and the polarization-related properties such as the polarization, spontaneous strain, dielectric permittivity, and piezoelectric coefficients. However, when considering polydomain structures, the thermodynamic analysis becomes much more complicated (Kouhar et al. 2001; Bratkovsky and Levanyuk 2002; Xu et al. 2015). Moreover, there is no direct access or even impossible to get the kinetic behavior of the polarization and properties on the domain walls through the thermodynamic analysis. In order to achieve this, the interfacial energy or domain wall energy and the time-evolving behavior of the polarization need to be considered.

2   Phase–Field Method

Phase-field method combines the material thermodynamics and kinetics. It has become an important tool for studying microstructure evolution (Chen 2002; Boettinger et al. 2002; Gránásy et al. 2006; Chen 2008; Steinbach 2009). The phase-field method for ferroelectric epitaxial thin films was firstly developed by Li et al., and it has been employed to study the strain-temperature phase diagrams, domain structures, domain switching, and other domain-related properties for various ferroelectric thin films (Li et al. 2001, 2002a, 2002b, and 2003b, 2006; Li and Chen 2006; Zhang et al. 2008; Kontsos and Landis 2010; Winchester et al. 2011; Britson et al. 2014; Hong et al. 2014; Chen et al. 2014; Wang et al. 2016). In the phase-field method of ferroelectric thin films, the polarization vector is selected as the order parameter. The spatial distribution of the polarization represents the domain structures of a thin film. The evolution of polarization can be described by the time-dependent Landau-Ginzburg (TDGL) equations, i.e.:

$$\frac{\partial \mathbf{P}(\mathbf{x}, t)}{\partial t} = -L \frac{\delta F}{\delta \mathbf{P}(\mathbf{x}, t)}, \tag{1}$$

where $\mathbf{P}$, $\mathbf{x}$, t, F, and L represent the polarization vector, spatial coordinate, time, the total free energy of the thin-film system, and the kinetic coefficient that is related to the domain wall mobility, respectively.

In the phase-field method of ferroelectric thin films (Li et al. 2001; Chen 2008), only polarization-related energy contributions are considered in the total free energy, which normally includes the Landau energy (bulk chemical energy), elastic energy, electric energy, and gradient energy, i.e.:

$$F = \iiint_V \left[f_{\text{bulk}}\left(\mathbf{P}, T\right) + f_{\text{grad}}\left(\nabla \mathbf{P}\right) + f_{\text{electric}}\left(\mathbf{P}, \mathbf{E}\right) + f_{\text{elastic}}\left(\mathbf{P}, \boldsymbol{\varepsilon}\right) \right] dV,$$

(2)

where f_{bulk}, f_{grad}, f_{electric}, f_{elastic}, $\boldsymbol{\varepsilon}$, $\mathbf{E}$, and $\nabla \mathbf{P}$ represent the bulk, gradient, electric, elastic energy densities, strain tensor, electric field vector, and polarization gradient, respectively. The bulk energy density can be expressed by a Landau polynomial as function of polarization, i.e.:

$$\begin{aligned} f_{\text{bulk}} = {} & \alpha_1 \left(P_1^2 + P_2^2 + P_3^2\right) + \alpha_{11} \left(P_1^4 + P_2^4 + P_3^4\right) \\ & + \alpha_{12} \left(P_1^2 P_2^2 + P_1^2 P_3^2 + P_2^2 P_3^2\right) + \alpha_{111} \left(P_1^6 + P_2^6 + P_3^6\right) \\ & + \alpha_{112} \left[P_1^4 \left(P_2^2 + P_3^2\right) + P_2^4 \left(P_1^2 + P_3^2\right) + P_3^4 \left(P_1^2 + P_2^2\right)\right] \\ & + \alpha_{123} P_1^2 P_2^2 P_3^2 + \alpha_{1111} \left(P_1^8 + P_2^8 + P_3^8\right) \\ & + \alpha_{1122} \left(P_1^4 P_2^4 + P_1^4 P_3^4 + P_2^4 P_3^4\right) \\ & + \alpha_{1112} \left[P_1^6 \left(P_2^2 + P_3^2\right) + P_2^6 \left(P_1^2 + P_3^2\right) + P_3^6 \left(P_1^2 + P_2^2\right)\right] \\ & + \alpha_{1123} \left(P_1^4 P_2^2 P_3^2 + P_1^2 P_2^4 P_3^2 + P_1^2 P_2^2 P_3^4\right), \end{aligned}$$

(3)

where α_i, α_{ij}, α_{ijk}, and α_{ijkl} are called Landau coefficients relating to the different-order dielectric stiffness. They can be fitted to the bulk properties at zero stress (Haun et al. 1989a; 1989b; 1989c; Li et al. 2005; Wang et al. 2007; Liang et al. 2009; Pohlmann et al. 2017).

The gradient energy density is defined as

$$f_{\text{grad}} = \frac{1}{2} \gamma_{ijkl} P_{i,j} P_{k,l},$$

(4)

where γ_{ijkl} are the components of the gradient energy coefficient tensor and $P_{i,j} = \partial P_i / \partial x_j$.

The electric energy density is given by:

$$f_{\text{electric}} = -P_i\left(\mathbf{x}\right) E_i\left(\mathbf{x}\right) - \frac{1}{2} \varepsilon_0 \kappa_{ij}^{\text{b}} E_i\left(\mathbf{x}\right) E_j\left(\mathbf{x}\right),$$

(5)

where κ_{ij}^{b} are components of the background dielectric constant tensor (Rupprecht and Bell 1964; Tagantsev 2008). The electric field distribution can be solved from

the electrostatic equilibrium equation, which can be dramatically affected by the electric boundary condition. For example, under open-circuit boundary conditions, if the bound charges at the film surfaces are not compensated, a depolarization field will be generated and may significantly suppress the polarization. This effect will be dominant when the thickness of the thin film is down to several unit cells, leading to the disappearance of the ferroelectricity, which is also called the size effect of the ferroelectric thin films (Scott 1988, 1991; Junquera and Ghosez 2003).

The elastic energy density can be calculated according to the Hooke's law:

$$f_{\text{elastic}} = \frac{1}{2} c_{ijkl} e_{ij}(\mathbf{x}) e_{kl}(\mathbf{x}) = \frac{1}{2} c_{ijkl} \left[\varepsilon_{ij}(\mathbf{x}) - \varepsilon_{ij}^0(\mathbf{x}) \right] \left[\varepsilon_{kl}(\mathbf{x}) - \varepsilon_{kl}^0(\mathbf{x}) \right], \tag{6}$$

where c_{ijkl} are components of the elastic stiffness tensor and $e_{ij}(\mathbf{x})$, $\varepsilon_{ij}(\mathbf{x})$, and $\varepsilon_{ij}^0(\mathbf{x})$ are components of the elastic strain tensor, total strain tensor, and eigenstrain tensor, respectively. The eigenstrain tensor $\varepsilon_{ij}^0(\mathbf{x})$, or the stress-free strain tensor, arises from the coupling between the polarization and the lattice degrees of freedom in ferroelectrics. This coupling is also named the electrostrictive effect (Cross et al. 1980; Cross 1996; Sundar and Newnham 1992). It is a universal effect in insulators, and in ferroelectrics the quadratic coupling can be especially strong. The strength of the quadratic coupling between polarization and the lattice degrees of freedom is described by the electrostrictive coefficient Q_{ijkl}, which is a fourth-order tensor. Its values can be obtained by experimental measurements (Yamada 1972) or first-principle calculations (Wang et al. 2010). The eigenstrain $\varepsilon_{ij}^0(\mathbf{x})$ can be calculated from $\varepsilon_{ij}^0(\mathbf{x}) = Q_{ijkl} P_k(\mathbf{x}) P_l(\mathbf{x})$. The total strain $\varepsilon_{ij}(\mathbf{x})$ can be solved from the mechanical equilibrium equation using Khachaturyan's microelasticity theory (Khachaturyan 1983), in combination with the thin-film mechanical boundary conditions (Li et al. 2002).

In the phase-field model of ferroelectric thin film, materials parameters needed include the Landau coefficients, gradient energy coefficients, background dielectric constant tensor, elastic compliance tensor, and electrostrictive coefficient tensor. With these materials parameters available, the domain structures under different temperatures and in-plane biaxial strain conditions can be obtained starting from a random distribution of polarization or from an existing domain structure.

3   Phase Transitions and Domain Structures

In the phase-field simulation of ferroelectric thin films, the equilibrium domain structures can be predicted by quenching an annealed paraelectric state with a fixed substrate constraint to lower temperatures. The main advantage of the phase-field method is the fact that one does not have to make a priori assumptions on the possible ferroelectric phases and domain structures that might appear under a given temperature and an epitaxial strain.

3.1   Misfit Strain–Temperature Phase Diagrams

The phase-field method has been employed to calculate the strain-temperature phase diagrams for several ferroelectric thin-film oxides such as $PbTiO_3$ (Li et al. 2001), $BaTiO_3$ (Choi et al. 2004), $SrTiO_3$ (Haeni et al. 2004), and $PbZr_{1-x}Ti_xO_3$ (Choudhury et al. 2005) (Fig. 1a–d). A common feature of these phase diagrams is the enhanced Curie temperature by the epitaxial strain. This effect can be dramatic as the Curie temperature can be increased by several hundred degrees by only 1% compressive or tensile strain. This effect originates from the strong coupling between the polarization and the lattice degrees of freedom. Another common feature among these phase diagrams is the existence of transition regions along the ferroelectric/paraelectric phase boundaries. For the stress-free bulk ferroelectrics including $PbTiO_3$, $BaTiO_3$, $SrTiO_3$, and $PbZr_{1-x}Ti_xO_3$, the phase transition is normally first-order at which the polarization changes sharply at the transition temperature. However, under thin-film mechanical boundary conditions with con-
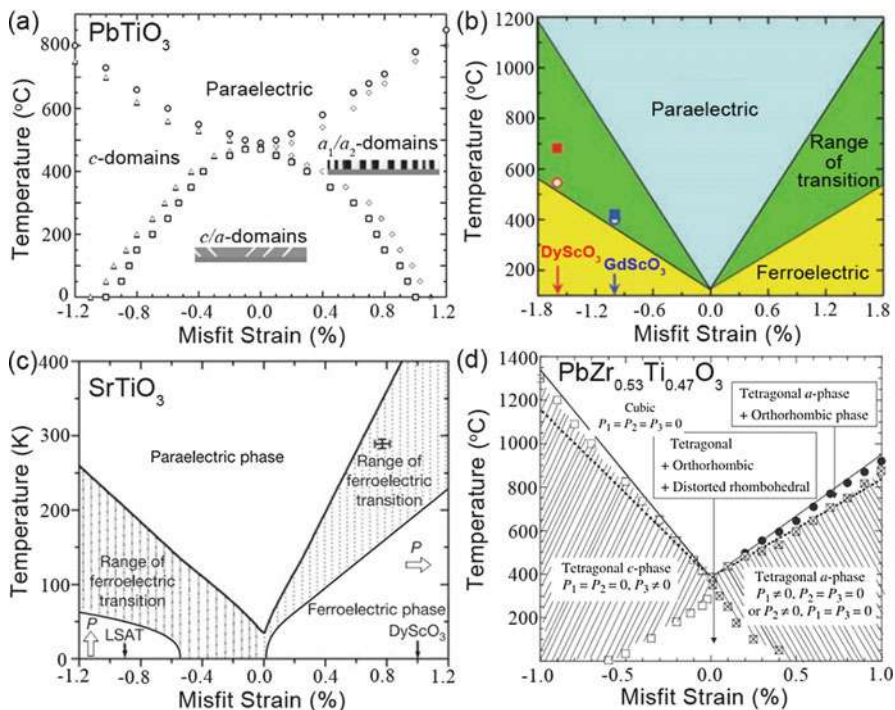
Fig. 1 Misfit strain-temperature phase diagrams of (**a**) $PbTiO_3$ (adapted from Li et al. 2001), (**b**) $BaTiO_3$ (adapted from Choi et al. 2004), (**c**) $SrTiO_3$ (adapted from Haeni et al. 2004), and (**d**) $PbZr_{0.53}Ti_{0.47}O_3$ (adapted from Choudhury et al. 2005) thin films predicted using the phase-field method. (Copyright 2001, American Institute of Physics, Copyright 2004 American Association for the Advancement of Science, Copyright 2004, Springer Nature, Copyright 2001, American Institute of Physics)

straint strains in-plane and stress-free out-of-plane, the paraelectric to ferroelectric phase transition may become second-order (Pertsev et al. 1999). This is caused by the renormalization of the fourth-order Landau coefficients due to the coupling between the polarization and the strain. On the strain-temperature phase diagram of PbTiO$_3$ thin film (Fig. 1a), there is a large region wherein the in-plane a domain and out-of-plane c domain can coexist, which has been verified by numerous experimental observations (Seifert et al. 1995; Kiguchi et al. 2011; Tang et al. 2014). For PbZr$_{0.53}$Ti$_{0.47}$O$_3$ thin film shown in Fig. 1d, there are also some mixed-phase regions. For example, under moderate strains and low temperatures, the tetragonal phase, orthorhombic phase, and distorted rhombohedral phase may coexist. These mixed-phase regions are normally not predicted from thermodynamic analysis without considering the multi-domain structures.

3.2   Domain Structures in Single–Layered Thin Films

A common feature of ferroelectric materials is the formation of domain structures when the material is cooled down from a paraelectric phase to a ferroelectric phase. A ferroelectric domain is defined by a region with uniform electrical polarization. The domain size may range from several nanometers to hundreds of nanometers depending on many factors such as the film thickness, the epitaxial strain, and the electric boundary conditions. The type of domain structures is determined by the total free energy, which can be altered by external stimuli including the mechanical, electric, thermal, and chemical stimuli. One advantage of phase-field modeling of the domain structures in ferroelectric thin films is the theoretically feasible incorporation of arbitrary boundary conditions, although only some of the possible boundary conditions can be realized in experimental configurations.

By incorporating materials parameters and incorporating boundary conditions of the experimental thin films into the phase-field model, the equilibrium domain structures can be obtained and compared with available experimental observations. For example, Fig. 2a shows the a/c domain structures predicted in a (001)-oriented PbTiO$_3$ thin film under a moderate compressive strain, while Fig. 2b gives the domain structures observed in a PbTiO$_3$ thin film epitaxially grown on a (001)-oriented SrTiO$_3$ substrate which can impose a compressive strain on the PbTiO$_3$ thin film. The a domain represents the domains with the polarization along either the x_1 or x_2 axis of the pseudocubic lattice, and the c domain represents the domains with the polarization along the x_3 axis of the pseudocubic lattice. The predicted domain structures (Li et al. 2002) show amazing similarity to the domain structures observed (Seifert et al. 1995) using the transmission electron microscopy (TEM). The formation of the a/c domain structures can be understood from the competition between the elastic energy, the electrostatic energy, and the gradient energy. For the moderately compressed thin films under short-circuit electric boundary conditions, a single c domain can minimize the electrostatic energy and the gradient energy. In this case however the macroscopic deformation along the out-of-plane direction will be maximized, leading to a high elastic energy. Therefore, some a domains
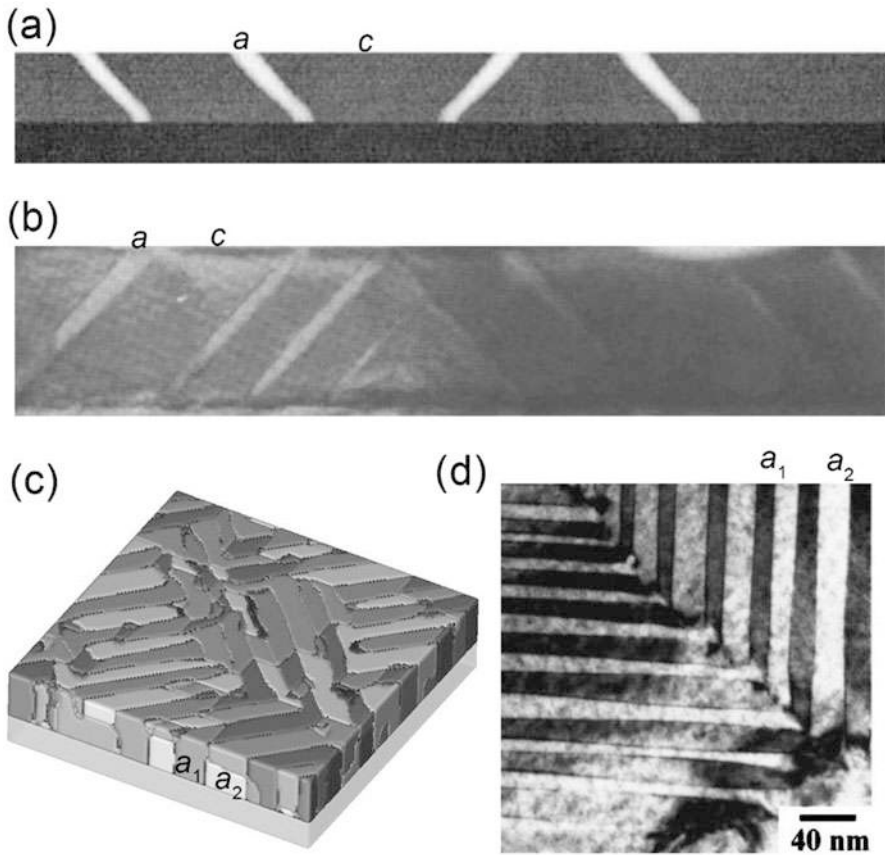
Fig. 2 (**a**) Predicted *a*/*c* domain structures in a (001)-oriented PbTiO$_3$ thin film under moderate compressive strains (adapted from Li et al. 2002), compared with (**b**) the experimental observation in a PbTiO$_3$ thin film epitaxially grown on a (001)-oriented SrTiO$_3$ substrate (adapted from Seifert et al. 1995). (**c**) Predicted a_1/a_2 twinning domain structures in a PbZr$_{0.2}$Ti$_{0.8}$O$_3$ thin film under equally biaxial tensile strains (adapted from Wang et al. 2016), compared with (**d**) the experimental observation in a PbTiO$_3$ thin film epitaxially grown on a (001)-oriented KTaO$_3$ substrate. (Adapted from Lee et al. 2001. Copyright 2002, 2016, Elsevier, Copyright 2001, American Institute of Physics)

are generated to decrease the macroscopic deformation and the elastic energy but comprise to slightly increase the electrostatic energy and the gradient energy.

3.3 Domain Structures in Bilayer Thin Films

The phase-field method was also employed by Artemev et al. to study the domain structures in ferroelectric bilayer thin films (Artemev et al. 2008). They found that a poled state can be achieved by a low electric field in the bilayer film with one layer

in a single-domain state and the other layer in a polydomain state. In contrast to the bilayer film with both layers in single-domain states, the presence of a polydomain layer in the bilayer film is responsible for a higher dielectric constant and lower coercive field. The increase in the applied electric field can lead to a transition to single-domain states in both layers of the bilayer film, and the critical field depends on the thermodynamic potentials of the two layers (Artemev et al. 2008). Using the thin-film phase-field method, Xue et al. investigated the effect of interfacial coherency between the two layers in a $Pb(Zr_{0.3}Ti_{0.7})O_3/Pb(Zr_{0.7}Ti_{0.3})O_3$ bilayer thin films on the domain structures (Xue et al. 2013). If the lattice parameters of the two layers end up with the same average lattice parameter, the two layers can be regarded as coherent. For example, the cubic lattice parameters for $Pb(Zr_{0.3}Ti_{0.7})O_3$ and $Pb(Zr_{0.7}Ti_{0.3})O_3$ are 4.0185 Å and 4.1032 Å, respectively, giving rise to an average lattice parameter of 4.0609 Å when the two layers have the same thickness. With a coherent interface, the strained bilayer film has a lattice parameter of 4.0609 Å. Consequently, the $Pb(Zr_{0.3}Ti_{0.7})O_3$ layer tolerates a tensile strain of $(4.0609-4.0185)/4.0185 = 1.05\%$, and the $Pb(Zr_{0.7}Ti_{0.3})O_3$ layer tolerates a compressive strain of $(4.0609-4.1032)/4.1032 = -1.03\%$. For a totally incoherent bilayer film, each layer can be regarded to have similar lattice parameters to their bulk counterparts, unless the clamping effect from the substrate is considered.

The interfacial coherency has significant effects on the domain structures in each layer (Fig. 3a and b). For $Pb(Zr_{0.3}Ti_{0.7})O_3/Pb(Zr_{0.7}Ti_{0.3})O_3$ bilayer thin films with coherent interface, tetragonal a_1/a_2 twinning domain structures (all in-plane polarizations) are predicted in the $Pb(Zr_{0.3}Ti_{0.7})O_3$ layer (Fig. 3a), while in its single-layer thin film, the domain structures are tetragonal a/c types (Fig. 3c, mixed in-plane and out-of-plane polarizations). The $Pb(Zr_{0.7}Ti_{0.3})O_3$ layer in the bilayer film shows tetragonal c domains (all out-of-plane polarizations), compared with rhombohedral domains in its single-layer thin-film counterpart (Fig. 3d). For the bilayer film with incoherent interface, the domain structures in each layer are similar to their single-layer thin-film counterparts, i.e., a/c domains in $Pb(Zr_{0.3}Ti_{0.7})O_3$ layer and rhombohedral domains in $Pb(Zr_{0.7}Ti_{0.3})O_3$ layer (Fig. 3b). The interfacial coherency effects on the domain morphologies can be understood from the strain configuration in the bilayer film. In $Pb(Zr_{0.3}Ti_{0.7})O_3$ layer, the tensile strain 1.05% tends to align all polarizations in-plane, leading to the formation of a_1/a_2 twinning domain structures under the synergistic effect from the bulk chemical energy and elastic energy. In $Pb(Zr_{0.7}Ti_{0.3})O_3$ layer, the compressive strain -1.03% tends to switch all polarizations out-of-plane, leading to the formation of c domains (Xue et al. 2013; Liu et al. 2016).

3.4 Strain Phase Separation

The thin-film phase-field model was also used to understand the formation of mixed phases and domains in strained thin films from the perspective of strain phase separation. In a constrained system, domains or phases can form and coexist with each other to minimize the overall elastic energy, given that each domain or phase
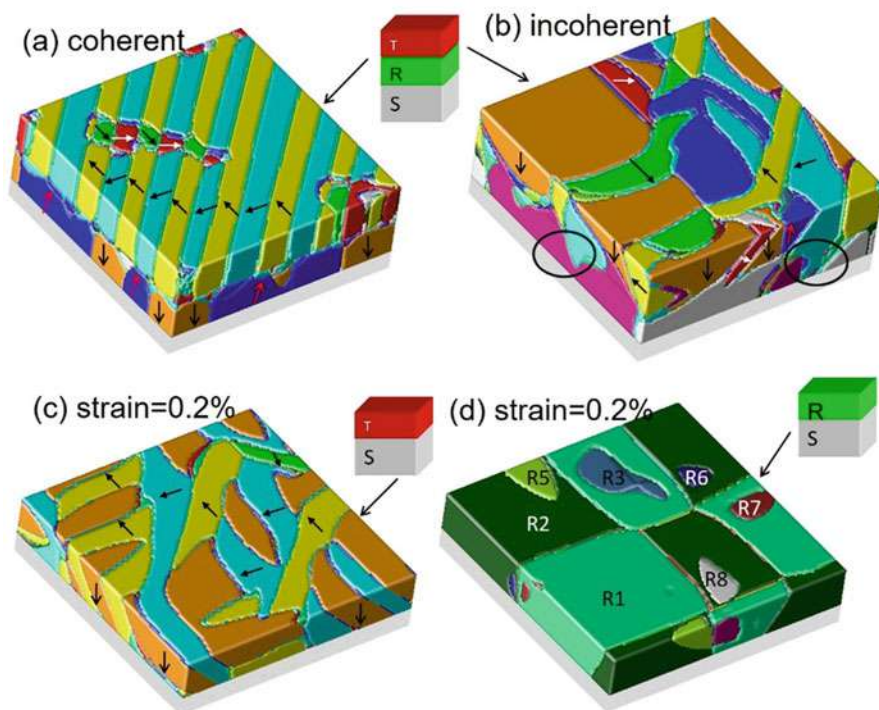
Fig. 3 Effects of interfacial coherency on the domain structures in a $Pb(Zr_{0.3}Ti_{0.7})O_3/Pb(Zr_{0.7}Ti_{0.3})O_3$ bilayer thin film (Xue et al. 2013). Domain structures of the bilayer film when the interface between the two layers is (**a**) coherent and (**b**) incoherent, compared with the domain structures in single-layer (**c**) $Pb(Zr_{0.3}Ti_{0.7})O_3$ thin film and (**d**) $Pb(Zr_{0.7}Ti_{0.3})O_3$ thin film under a moderate strain of 0.2% due to thermal expansion. (Copyright 2013, Elsevier)

has different local strains. This phenomenon is similar to the phase decomposition process from a chemically homogeneous phase to a two-phase mixture wherein each phase has a different chemical composition. The volume fraction of each phase with a specific composition can be obtained from the lever rule. This idea was used to understand the observed rhombohedral-tetragonal mixed phases in severely compressed $BiFeO_3$ thin films (Xue et al. 2016, 2017). $BiFeO_3$ has a pseudocubic lattice parameter of $\sim$4.0 Å. When it is epitaxially grown on a $LaAlO_3$ substrate with a lattice parameter of $\sim$3.79 Å, a large compressive strain up to $\sim$−5.25% will be imposed on the thin film, leading to a transition from the rhombohedral phase to a tetragonal-like monoclinic phase in $BiFeO_3$ thin film (Zeches et al. 2009; Li et al. 2015; Zhang et al. 2012). With the film thickness increasing to be larger than 100 nm, the large epitaxial strain is partially relaxed, and the rhombohedral phase emerges, forming a tetragonal-like/rhombohedral mixed domain structure (Fig. 4a
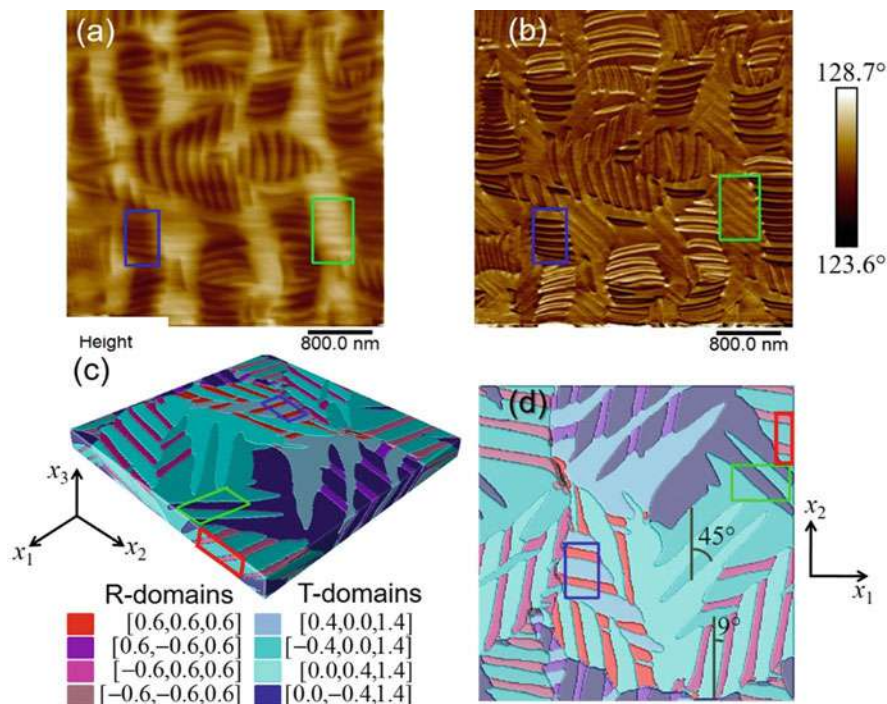
Fig. 4 Experimentally observed (**a**) topography and (**b**) the corresponding in-plane piezoresponse force microscopy result of a mixed-phase $BiFeO_3$ thin film grown on $LaAlO_3$ substrate, compared with the simulated (**a**) three-dimensional domain structures and (**b**) in-plane domain morphology from thin-film phase-field method. (Adapted from Xue et al. 2016. Copyright 2016, American Physics Society)

and b). The phase-field model was employed to predict domain structures (Fig. 4c and d) which agree well with the experimental observations and give insights to the underlying physics. With the relaxation of the mismatch strain, the overall strain will be smaller than the critical strain, under which the transition from rhombohedral phase to tetragonal-like phase occurs. Therefore, the phase separation happens to decrease the total elastic energy, analogous to the chemical decomposition process (Xue et al. 2016, 2017).

3.5   Understanding the Polarization Switching in Ferroelectric Thin Films

One advantage of the phase-field method over the thermodynamic analysis is the capability of studying the polarization switching behavior in ferroelectric thin films. The polarization switching driven by externally applied electric fields was simulated

by Li et al. for epitaxial BaTiO$_3$ thin films (Li et al. 2008). They found that the coercive field and saturated polarization in the hysteresis loop strongly depend on the type and density of the interfacial dislocation. For example, the coercive field can be decreased by increasing the density of interfacial misfit dislocations on a (100) slip plane with a Burger vector along [$-$100] direction. With an optimal combination of different misfit dislocations, the remnant polarization can be greatly enhanced and the coercive field can be greatly reduced. The polarization versus electric field hysteresis loops can be distorted to be asymmetric by threading dislocations. Wu et al. investigated the effects of background dielectric constant and interfacial coherency on the hysteresis loops in BaTiO$_3$/SrTiO$_3$ bilayer films (Wu et al. 2012, 2015). It was found that the hysteresis loops can be constricted when the background dielectric constant is decreased. The decrease in the background dielectric constant has a similar effect to the decrease in the space charge density. The phase-field simulation predicted the domain structure evolution during the switching process in the ferroelectric/paraelectric bilayer film, revealing a dipole-dipole inter-action similar to the exchange-spring interaction in magnetic multilayer systems (Wu et al. 2012). The interfacial coherency also has a remarkable impact on the polarization switching behavior. In the bilayer film with fully coherent interfaces, the hysteresis loops are square-like, exhibiting a large coercive field and a large remnant polarization. For the bilayer films with partially relaxed heterointerfaces, the remnant polarization is relatively high but the coercive field is dramatically reduced. For the bilayer film with fully relaxed heterointerfaces, the hysteresis loops are very slim, showing both small remnant polarization and coercive field.

The polarization switching is normally achieved by the externally applied electric field. However, it can also be achieved by a mechanical loading via the flexoelectric effects in epitaxial thin films. The flexoelectric effect is defined as the coupling between the polarization and the strain gradient. This effect is much stronger in ferroelectric thin films than in their bulk counterparts. It has been implemented into the thin-film phase-field model by several research groups (Chen et al. 2014; Ahluwalia et al. 2014; Gu et al. 2014; Chen et al. 2015). In these works, the flexoelectric effects on the domain structures and switching were studied. For example, Chen et al. studied the stability and switching patterns of domain structures in ferroelectric thin films subjected to mechanical loads and related flexoelectric fields (Chen et al. 2015). As shown in Fig. 5a and b, a mixed a/c domain structure can be switched to a single-domain structure by cylindrical bending with a large surface flexoelectric field. For the upward bending case shown in Fig. 5a, the out-of-plane c^- domains are gradually switched to in-plane at first and then are switched to the out-of-plane c^+, while for the downward bending case shown in Fig. 5b, all in-plane domains are switched to the out-of-plane c^- domain. These two switching patterns can be understood from the different flexoelectric fields due to the different bending directions. For the film with an upward bending, the induced flexoelectric field is upward, leading to the switching of polarization to c^+. For the film with a downward bending, the induced flexoelectric field is downward, leading to the switching of polarization to c^-.
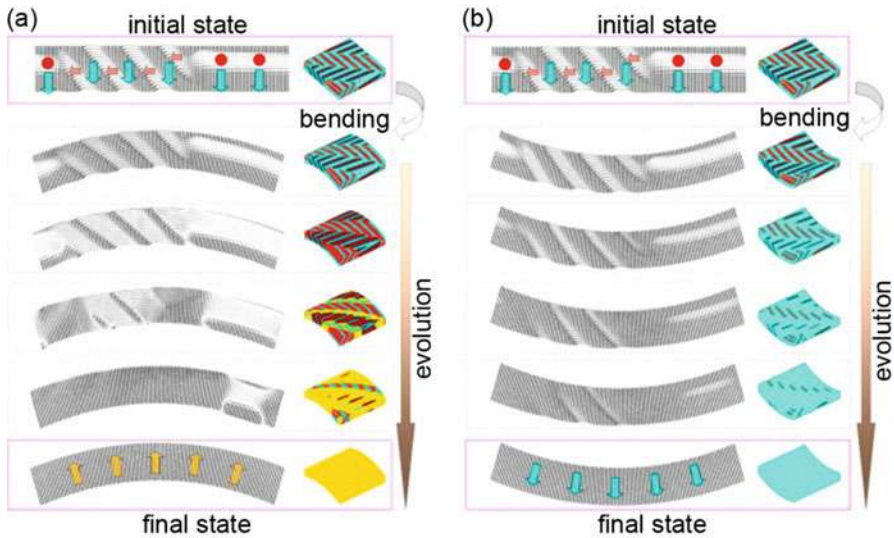
Fig. 5 Switching behavior of the domain structure from a/c mixed structures to (**a**) single out-of-plane c^+ domain under upward bending and to (**b**) single out-of-plane c^- domain under downward bending. (Adapted from Chen et al. 2015. Copyright 2015, Elsevier)

4 Toward a Multiscale Simulation

Although the thin-film phase-field method has gained great success in understanding the experimental domain structures and phase transitions, and in predicting the domain structures and switching in thin films under conditions beyond the experiment, it has only been applied to limited material systems due to the lack of the thermodynamic parameters including the Landau coefficients, electrostrictive coefficients, elastic constants, and gradient energy coefficients. For example, up to now, there are very limited materials that have available Landau potentials such as $BaTiO_3$, $PbTiO_3$, $PbZr_{1-x}Ti_xO_3$, $BiFeO_3$, $CaTiO_3$, $SrTiO_3$, $KNa_{1-x}Nb_xO_3$, $LiNbO_3$, $LiTaO_3$, $SrBi_2Nb_2O_9$, and $Sr_{0.8}Bi_{2.2}Ta_2O_9$. However, the ferroelectrics that have been discovered are far more beyond these. In order to obtain the Landau coefficients, the polarization and dielectric responses including both the first-order and higher-order dielectric stiffness as function of temperature have to be measured, which is experimentally challenging especially when the required temperature range is broad. An alternative approach to obtaining the Landau coefficients is the first-principle calculation, which has been employed to obtain the Landau coefficients for $BaTiO_3$ (Íñiguez et al. 2001; Geneste 2009), $CaTiO_3$ (Eklund 2010; Gu et al. 2012), and $BiFeO_3$ (Marton et al. 2017). In addition to the Landau coefficients, the first-principle method has also been employed to calculate the phase transitions (Zhong et al. 1994), elastic constants (Wu et al. 2005), electrostrictive coefficients (Wang

et al. 2010), and polarization domain wall energies (Wang et al. 2013; Diéguez et al. 2013; Li et al. 2014). For the domain wall motion speed which is related to the kinetic coefficient L in the TDGL equation, it can be studied using the atomistic molecular dynamics and coarse-grained Monte Carlo simulations (Shin et al. 2007). All these efforts point toward the multiscale simulation-assisted phase-field modeling for ferroelectric thin films (Völker et al. 2011a, 2011b), which can provide principal inputs for future high-throughput calculation (Shen et al. 2017) and optimization of properties and microstructure design for ferroelectric thin films using machine learning (Li et al. 2017).

References

Ahluwalia R, Tagantsev AK, Yudin P, Setter N, Ng N, Srolovitz DJ (2014) Influence of flexoelectric coupling on domain patterns in ferroelectrics. Phys Rev B 89:174105

Artemev A, Geddes B, Slutsker J, Roytburd A (2008) Thermodynamic analysis and phase field modeling of domain structures in bilayer ferroelectric thin films. J Appl Phys 103:074104

Boettinger WJ, Warren JA, Beckermann C, Karma A (2002) Phase field simulation of solidification. Ann Rev Mater Res 32:163

Bratkovsky AM, Levanyuk AP (2002) Elastic domain structure and the transition between polydomain and monodomain states in thin epitaxial films. Phys Rev B 65:094102

Britson J, Nelson C, Pan XQ, Chen LQ (2014) First-order morphological transition of ferroelastic domains in ferroelectric thin films. Acta Mater 75:188–197

Catalan G, Scott JF (2009) Physics and applications of bismuth ferrite. Adv Mater 21:2463–2485

Chao AY, Arthur AY (1975) Molecular beam epitaxy. Prog Solid State Chem 10:157–192

Chen LQ (2002) Phase-field models for microstructure evolution. Annu Rev Mater Res 32:113

Chen LQ (2008) Phase-field method of phase transitions/domain structures in ferroelectric thin film s: a review. J Am Ceram Soc 91:1835–1844

Chen HT, Soh AK, Ni Y (2014) Phase field modeling of flexoelectric effects in ferroelectric epitaxial thin films. Acta Mech 225:1323–1333

Chen WJ, Zheng Y, Feng X, Wang B (2015) Utilizing mechanical loads and flexoelectricity to induce and control complicated evolution of domain patterns in ferroelectric nanofilms. J Mech Phys Solids 79:108–113

Choi KJ, Biegalski M, Li YL, Sharan A, Schubert J, Uecker R, Reiche P, Chen YB, Pan XQ, Gopalan V, Chen LQ, Schlom DG, Eom CB (2004) Enhancement of ferroelectricity in strained BaTiO$_3$ thin films. Science 306:1005

Choudhury S, Li YL, Chen LQ (2005) A phase diagram for epitaxial PbZr$_{1-x}$Ti$_x$O$_3$ thin films at the bulk morphotropic boundary composition. J Am Ceram Soc 88:1669–1672

Cross LE (1996) Ferroelectric materials for electromechanical transducer applications. Mater Chem Phys 43:108–115

Cross LE, Jang SJ, Newnham RE, Nomura S, Uchino K (1980) Large electrostrictive effects in relaxor ferroelectrics. Ferroelectrics 23:187–191

Diéguez O, Aguado-Puente P, Junquera J, Íñiguez J (2013) Domain walls in a perovskite oxide with two primary structural order parameters: first-principles study of BiFeO$_3$. Phys Rev B 89:059901

Eckstein J, Bozovic I (1995) High-temperature superconducting multilayers and heterostructures grown by atomic layer-by-layer molecular beam epitaxy. Annu Rev Mater Sci 25:679–709

Eklund CJ (2010) Ph.D. thesis, Rutgers University, New Brunswick

Emelyanov AY, Pertsev NA, Kholkin AL (2002) Effect of external stress on ferroelectricity in epitaxial thin films. Phys Rev B 66:214108

Frey T, Chi CC, Tsuei CC, Shaw T, Bozso F (1994) Effect of atomic oxygen on the initial growth mode in thin epitaxial cuprate films. Phys Rev B 49:3483–3491

Geneste G (2009) Landau free energy of ferroelectric crystals by thermodynamic integration. Phys Rev B 79:064101

Gránásy L et al (2006) Phase field theory of crystal nucleation and polycrystalline growth: a review. J Mater Res 21:309

Gu YJ, Rabe K, Bousquet E, Gopalan V, Chen LQ (2012) Phenomenological thermodynamic potential for CaTiO$_3$ single crystals. Phys Rev B 85:064117

Gu YJ, Li ML, Morozovska AN, Wang Y, Eliseev EA, Gopalan V, Chen LQ (2014) Flexoelectricity and ferroelectric domain wall structures: phase-field modeling and DFT calculations. Phys Rev B 89:174111

Haeni JH, Irvin P, Chang W, Uecker R, Reiche P, Li YL, Choudhury S, Tian W, Hawley ME, Craigo B, Tagantsev AK, Pan XQ, Streiffer SK, Chen LQ, Kirchoefer SW, Levy J, Schlom DG (2004) Room-temperature ferroelectricity in strained SrTiO$_3$. Nature 430:758

Haun MJ, Furman E, Jang SJ, Cross LE (1989a) Thermodynamic theory of the lead zirconate-titanate solid solution system, part I: phenomenology. Ferroelectrics 99:13–25

Haun MJ, Furman E, McKinstry HA, Cross LE (1989b) Thermodynamic theory of the lead zirconate-titanate solid solution system, part II: Tricritical behavior. Ferroelectrics 99:27–44

Haun MJ, Zhuang ZQ, Furman E, Jang SJ, Cross LE (1989c) Thermodynamic theory of the lead zirconate-titanate solid solution system, part III: curie constant and sixth-order polarization interaction dielectric stiffness coefficients. Ferroelectrics 99:45–54

Hong ZJ, Britson J, Hu JM, Chen LQ (2014) Local 90 degrees switching in Pb(Zr$_{0.2}$Ti$_{0.8}$)O$_3$ thin film: phase-field modeling. Acta Mater 73:75–82

Hubler GK (1992) Pulsed laser deposition. MRS Bull 17:26–29

Íñiguez J, Ivantchev S, Perez-Mato JM, García A (2001) Devonshire-landau free energy of BaTiO$_3$ from first principles. Phys Rev B 63:144103

Junquera J, Ghosez P (2003) Critical thickness for ferroelectricity in perovskite ultrathin films. Nature 422:506

Karpinsky DV, Eliseev EA, Xue F, Silibin MV, Franz A, Glinchuk MD, Troyanchuk IO, Gavrilov SA, Gopalan V, Chen LQ, Morozovska AN (2017) Thermodynamic potential and phase diagram for multiferroic bismuth ferrite (BiFeO3). npj Comput Mater 3:20

Khachaturyan AG (1983) Theory of structural transformations in solid. Wiley, New York

Kiguchi T, Aoyagi K, Ehara Y, Funakubo H, Yamada T, Usami N, Konno TJ (2011) Configuration and local elastic interaction of ferroelectric domains and misfit dislocation in PbTiO$_3$/SrTiO$_3$ epitaxial thin films. Sci Technol Adv Mater 12:034413

Kontsos A, Landis CM (2010) Phase-field modeling of domain structure energetics and evolution in ferroelectric thin films. J Appl Mech 77:041014

Koster G, Rijnders GJHM, Blank DHA, Rogalla H (1999) Imposed layer-by-layer growth by pulsed laser interval deposition. Appl Phys Lett 74:3729

Kouhar VG, Pertsev NA, Waser R (2001) Thermodynamic theory of epitaxial ferroelectric thin films with dense domain structures. Phys Rev B 64:214103

Kwo J, Hong M, Trevor DJ, Fleming RM, White AE, Farrow RC, Kortan AR, Short KT (1988) In situ epitaxial growth of Y$_1$Ba$_2$Cu$_3$O$_{7-x}$ films by molecular beam epitaxy with an activated oxygen source. Appl Phys Lett 53:2683

Lee KS, Choi JH, Lee JY, Baik S (2001) Domain formation in epitaxial Pb(Zr,Ti)O3 thin films. J Appl Phys 90:4095

Li ML, Gu YJ, Wang Y, Chen LQ, Duan WH (2014) First-principles study of 180° domain walls in BaTiO$_3$: mixed Bloch-Neel-Ising character. Phys Rev B 90:054106

Li Q, Nelson CT, Hsu SL, Damodaran AR, Li LL, Yadav AK, McCarter M, Martin LW, Ramesh R, Kalinin SV (2017) Quantification of flexoelectricity in PbTiO$_3$/SrTiO$_3$ superlattice polar vortices using machine learning and phase-field modeling. Nat Commun 8:1468

Li YJ, Wang JJ, Ye JC, Ke XX, Gou GY, Wei Y, Xue F, Wang J, Wang CS, Peng RC, Deng XL, Yang Y, Ren XB, Chen LQ, Nan CW, Zhang JX (2015) Mechanical switching of nanoscale multiferroic phase boundaries. Adv Funct Mater 25:3405–3413

Li YL, Chen LQ (2006) Temperature-strain phase diagram for BaTiO$_3$ thin films. Appl Phys Lett 88:072905

Li YL, Hu SY, Liu ZK, Chen LQ (2001) Phase-field model of domain structures in ferroelectric thin films. Appl Phys Lett 78:3878–3880

Li YL, Hu SY, Liu ZK, Chen LQ (2002a) Effect of substrate constraint on the stability and evolution of ferroelectric domain structures in thin films. Acta Mater 50:395–411

Li YL, Hu SY, Liu ZK, Chen LQ (2002b) Effect of electrical boundary conditions on ferroelectric domain structures in thin films. Appl Phys Lett 81:427–429

Li YL, Choudhury S, Liu ZK, Chen LQ (2003a) Effect of external mechanical constraints on the phase diagram of epitaxial PbZr$_{1-x}$Ti$_x$O$_3$PbZr$_{1-x}$Ti$_x$O$_3$ thin films – thermodynamic calculations and phase-field simulations. J Appl Phys 83:1608

Li YL, Hu SY, Chen LQ (2003b) Ferroelectric domain morphologies of (001)PbZr$_{1-x}$Ti$_x$O$_3$ epitaxial thin films. J Appl Phys 97:034112

Li YL, Cross LE, Chen LQ (2005) A phenomenological thermodynamic potential for BaTiO$_3$ single crystals. J Appl Phys 98:064101

Li YL, Choudhury S, Haeni JH, Biegalski MD, Vasudevarao A, Sharan A, Ma HZ, Levy J, Gopalan V, Trolier-McKinstry S, Schlom DG, Jia QX, Chen LQ (2006) Phase transitions and domain structures in strained pseudocubic (100) SrTiO$_3$ thin films. Phys Rev B 73: 184112

Li YL, Hu SY, Choudhury S, Baskes MI, Saxena A, Lookman T, Jia QX, Schlom DG, Chen LQ (2008) Influence of interfacial dislocations on hysteresis loops of ferroelectric films. J Appl Phys 104:104110

Liang LY, Li YL, Chen LQ, Hu SY, Lu GH (2009) A thermodynamic free energy function for potassium niobate. Appl Phys Lett 94:072904

Liu GQ, Zhang Q, Huang HH, Munroe P, Nagarajan V, Simons H, Hong ZJ, Chen LQ (2016) Reversible polarization rotation in epitaxial ferroelectric bilayers. Adv Mater Interfaces 3:1600444

Ma H, Chen L, Wang JL, Ma J, Boey F (2008) Strain effects and thickness dependence of ferroelectric properties in epitaxial BiFeO$_3$ thin films. Appl Phys Lett 92:182902

Martin LW, Rappe AM (2017) Thin-film ferroelectric materials and their applications. Nat. Rev. Mater. 2:16087

Marton P, Klíč A, Paściak M, Hlinka J (2017) First-principles-based landau-Devonshire potential for BiFeO$_3$. Phys Rev B 96:174110

Pertsev NA, Zembilgotov AG, Tagantsev AK (1998) Effect of mechanical boundary conditions on phase diagrams of epitaxial ferroelectric thin films. Phys Rev Lett 80:1988

Pertsev NA, Zembilgotov AG, Tagantsev AK (1999) Equilibrium states and phase transitions in epitaxial ferroelectric thin films. Ferroelectrics 223:79–90

Pertsev NA, Kukhar VG, Kohlstedt H, Waser R (2003) Phase diagrams and physical properties of single-domain epitaxial Pb(Zr$_{1-x}$Ti$_x$)O$_3$ thin films. Phys Rev B 67:054107

Pohlmann H, Wang JJ, Wang B, Chen LQ (2017) A thermodynamic potential and the temperature-composition phase diagram for single-crystalline K$_{1-x}$Na$_x$NbO$_3$ ($0 \leq x \leq 0.5$). Appl Phys Lett 110:102906

Qiu QY, Mahjoub R, Alpay SP, Nagarajan V (2010) Misfit strain-film thickness phase diagrams and related electromechanical properties of epitaxial ultra-thin lead zirconate titanate films. Acta Mater. 58:823–835

Ramesh R, Luther K, Wilkens B, Hart DL, Wang E, Tarascon JM, Inam A, Wu XD, Venkatesan T (1990) Epitaxial growth of ferroelectric bismuth titanate thin films by pulsed laser deposition. Appl Phys Lett 57:1505

Roytburd AL (1998) Thermodynamics of polydomain heterostructures. I. Effect of macrostresses. J. Appl. Phys. 83:228–238

Rupprecht G, Bell RO (1964) Dielectric constant in paraelectric perovskites. Phys Rev 135:A748

Schlom DG, Eckstein JN, Hellman ES, Streiffer SK, Harris JS, Beasley MR, Bravman JC, Geballe TH, Webb C, Dessonneck KE, Turner F (1988) Molecular beam epitaxy of layered Dy–Ba–Cu–O com-pounds. Appl Phys Lett 53:1660

Schlom DG, Chen, LQ, Eom CB, Rabe KM, Streiffer SK, Triscone JM (2007) Strain tuning of ferroelectric thin films. Ann. Rev. Mater. Res. 37:589–626

Schlom DG, Chen LQ, Pan XQ, Schmehl A, Zurbuchen MA (2008) A thin film approach to engineering functionality into oxides. J. Am. Ceram. Soc. 91:2429–2454

Schlom DG, Chen LQ, Fennie CJ, Gopalan V, Muller DA, Pan XQ, Ramesh R, Uecker R (2014) Elastic strain engineering of ferroic oxides. MRS Bulletin 39:118–130

Scott JF (1988) Properties of ceramic KNO_3 thin film memories. Physica B 150:160–167

Scott JF (1991) Phase transitions in ferroelectric thin films. Phase Trans 30:107–110

Seifert A, Lange FF, Speck JS (1995) Epitaxial growth of $PbTiO_3$ thin films on (001) $SrTiO_3$ from solution precursors. J Mater Res 10:680–691

Shen ZH, Wang JJ, Lin YH, Nan CW, Chen LQ, Shen Y (2017) High-throughput phase-field design of high-energy-density polymer nanocomposites. Adv Mater 30:1704380

Sheng G, Hu JM, Zhang JX, Li YL, Choudhury S, Jia QX, Liu ZK, Chen LQ (2008) Misfit strain-misfit strain diagram of epitaxial $BaTiO_3$ thin films: Thermodynamic calculations and phase-field simulations. Appl. Phys. Lett. 93:232904

Shin YH, Grinberg I, Chen IW, Rappe AM (2007) Nucleation and growth mechanism of ferroelectric domain-wall motion. Nature 449:881–884

Shirokov VB, Yuzyuk YI, Dkhil B, Lemanov VV (2007) Phenomenological theory of phase transitions in epitaxial $BaTiO_3$ thin films. Phys Rev B 75:224116

Shirokov VB, Yuzyuk YI, Dkhil B, Lemanov VV (2009) Phenomenological theory of phase transitions in epitaxial $Ba_xSr_{1-x}TiO_3$ thin films. Appl. Phys. Lett. 79:144118

Spah RJ, Hess HF, Stormer HL, White AE, Short KT (1988) Parameters for *in situ* growth of high Tc superconducting thin films using an oxygen plasma source. Appl Phys Lett 53:441

Steinbach I (2009) Phase-field models in materials science. Model Simul Mater Sci Eng 17: 073001

Sundar V, Newnham RE (1992) Electrostriction and polarization. Ferroelectrics 135:431–446

Tagantsev AK (2008) Landau expansion for ferroelectrics: which variable to use? Ferroelectrics 375:19

Tang YL, Zhu YL, Wang YJ, Wang WY, Xu YB, Ren WJ, Zhang ZD, Ma XL (2014) Atomic-scale mapping of dipole frustration at 90° charged domain walls in ferroelectric $PbTiO_3$ films. Sci Rep 4:4115

Theis CD, Yeh J, Schlom DG, Hawley ME, Brown GW (1998) Adsorption-controlled growth of $PbTiO_3$ by reactive molecular beam epitaxy. Thin Solid Films 325:107–114

Völker B, Marton P, Elsässer C, Kamlah M (2011a) Multiscale modeling for ferroelectric materials: a transition from the atomic level to phase-field modeling. Contin Mech Thermodyn 23: 435–451

Völker B, Landis CM, Kamlah M (2011b) Multiscale modeling for ferroelectric materials: multiscale modeling for ferroelectric materials: identification of the phase-field model's free energy for PZT from atomistic simulations. Smart Mater Struc 21:035025

Wang JJ, Meng FY, Ma XQ, Xu MX, Chen LQ (2010) Lattice, elastic, polarization, and electrostrictive properties of $BaTiO_3$ from first-principles. J Appl Phys 108:034107

Wang JJ, Wang Y, Ihlefeld JF, Hopkins PE, Chen LQ (2016) Tunable thermal conductivity via domain structure engineering in ferroelectric thin films: a phase-field simulation. Acta Mater 111:220–231

Wang Y, Nelson C, Melville A, Winchester B, Shang SL, Liu ZK, Schlom DG, Pan XQ, Chen LQ (2013) BiFeO$_3$ domain wall energies and structures: a combined experimental and density functional theory plus U study. Phys Rev Lett 110:267601

Wang YL, Tagantsev AK, Damjanovic D, Setter N (2007) Landau thermodynamic potential for BaTiO$_3$. J Appl Phys 101:104115

Winchester B, Wu PP, Chen LQ (2011) Phase-field simulation of domain structures in epitaxial BiFeO$_3$ films on vicinal substrates. Appl Phys Lett 99:052903

Wu XF, Vanderbilt D, Hamann DR (2005) Systematic treatment of displacements, strains, and electric fields in density-functional perturbation theory. Phys Rev B 72:035105

Wu PP, Ma XQ, Li YL, Gopalan V, Chen LQ (2012) Dipole spring ferroelectrics in superlattice SrTiO$_3$/BaTiO$_3$ thin films exhibiting constricted hysteresis loops. Appl Phys Lett 100:092905

Wu PP, Ma XQ, Li YL, Eom CB, Schlom DG, Gopalan V, Chen LQ (2015) Influence of interfacial coherency on ferroelectric switching of superlattice BaTiO$_3$/SrTiO$_3$. Appl Phys Lett 107:122906

Xu RJ, Zhang JL, Chen ZH, Martin LW (2015) Orientation-dependent structural phase diagrams and dielectric properties of PbZr$_{1-x}$Ti$_x$O$_3$ polydomain thin films. Phys Rev B 91:144106

Xue F, Wang JJ, Sheng G, Huang E, Cao Y, Huang HH, Munroe P, Mahjoub R, Li YL, Nagarajan V, Chen LQ (2013) Phase field simulations of ferroelectrics domain structures in PbZr$_x$Ti$_{1-x}$O$_3$ bilayers. Acta Mater 61:2909–2918

Xue F, Li YJ, Gu YJ, Zhang JX, Chen LQ (2016) Strain phase separation: formation of ferroelastic domain structures. Phys Rev B 94:220101

Xue F, Ji YZ, Chen LQ (2017) Theory of strain phase separation and strain spinodal: applications to ferroelastic and ferroelectric systems. Acta Mater 133:147–159

Yamada T (1972) Electromechanical properties of oxygen-octahedra ferroelectric crystals. J Appl Phys 43:328

Zeches RJ, Rossell MD, Zhang JX, Hatt AJ, He Q, Yang CH, Kumar A, Wang CH, Melville A, Adamo C, Sheng G, Chu YH, Ihlefeld JF, Erni R, Ederer C, Gopalan V, Chen LQ, Schlom DG, Spaldin NA, Martin LW, Ramesh R (2009) A strain-driven morphotropic phase boundary in BiFeO$_3$. Science 326:977–980

Zembilgotov AG, Pertsev NA, Böttger U, Waser R (2005) Effect of anisotropic in-plane strains on phase states and dielectric properties of epitaxial ferroelectric thin films. Appl Phys Lett 86:052903

Zhang JX, Li YL, Wang Y, Liu ZK, Chen LQ, Chu YH, Zavaliche F, Ramesh R (2007) Effect of substrate-induced strains on the spontaneous polarization of epitaxial BiFeO$_3$ thin films. J Appl Phys 101:114105

Zhang JX, Li YL, Choudhury S, Chen LQ, Chu YH, Zavaliche F, Cruz MP, Ramesh R, Jia QX (2008) Computer simulation of ferroelectric domain structures in epitaxial BiFeO$_3$ thin films. J Appl Phys 103:094111

Zhang W, Ouyang J, Roytburd AL (2012) Effect of thickness-mediated misfit strain on the heterophase polydomain structure of epitaxial BiFeO$_3$ films. Script Mater 66:499–502

Zhong WL, Vanderbilt D, Rabe KM (1994) Phase transitions in BaTiO$_3$ from first principles. Phys Rev Lett 73:1861

Battery Electrodes, Electrolytes, and Their Interfaces

50

Iek-Heng Chu, Minghao Zhang, Shyue Ping Ong, and Ying Shirley Meng

Contents

Abstract

In recent years, first-principles modeling techniques have made tremendous advances. This allows researchers to estimate the various properties of materials and provide invaluable insights into the physical processes from a microscopic perspective, which cannot directly be assessable by experiments. With the continuing increasing computation powers, first-principles methods are expected to play a more important role in materials design. This chapter aims to serve as a battery-related computation handbook for general readers who may be

The authors Iek-Heng Chu and Minghao Zhang contributed equally to this chapter

I.-H. Chu · M. Zhang · S. P. Ong (✉) · Y. S. Meng (✉)
Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA
e-mail: ihchu@eng.ucsd.edu; johnson1228@gmail.com; miz016@eng.ucsd.edu; ongsp@eng.ucsd.edu; shirleymeng@ucsd.edu

new to first-principles calculations. Specifically, this chapter will introduce the well-established ab initio modeling methods widely used in battery-related studies from both the thermodynamic and kinetics aspects. The thermodynamic approaches that will be discussed include the computations of equilibrium voltage and voltage profiles, electronic structure, and stability analyses. The kinetics approaches will cover common methods for ionic diffusion studies: transition state theory and nudged elastic band method and ab initio molecular dynamics. This is followed by the conclusion and outlook.

Keywords
Alkali-ion battery · Electrodes · Electrolytes · Interfaces · First-principles modeling · Voltage profile · Electronic structure · Phase stability · Surface · Defects · Ab initio molecular dynamics · Ion diffusion · Nudged elastic band · Transition state theory

1   Introduction

Global warming coupled with energy demand has made the development of renewable energy technologies (solar, geothermal, tidal, wind, etc.) a global interest and major concern. However, most renewable energy technologies suffer from high variability and mismatch between demand and supply. Reliable and cost-effective electrical energy storage is therefore a crucial enabler to renewables, as well as nonpolluting means of transportation (Lewis 2007). Among the multitude of energy storage technologies (e.g., zinc-manganese battery, nickel-cadmium battery, nickel-hydrogen battery, lead-acid battery, alkali-ion battery, fuel cell, redox flow battery, etc.), the rechargeable alkali-ion battery has emerged as one of the most attractive candidates for grid and vehicular applications and the dominant one for mobile consumer applications, due to its high specific energy, high operation voltage, wide working temperature, and long cycle life (Tarascon and Armand 2001).

A rechargeable battery comprises two electrodes – the cathode and the anode – separated by an electrolyte (Fig. 1). Alkali ions shuttle between the two electrodes, with the electrolyte acting as an alkali-ion conductor and electrical insulator. During discharge, alkali ions (A^+ in Fig. 1) are extracted from anode and inserted to the cathode, passing through the electrolyte in the process. Meanwhile, electrons pass through the external circuit to perform work. The reverse process occurs during charge.

The energy contained in any battery is the integral of the voltage multiplied by the charge capacity. To achieve high-energy and high-power density for long cycling life in alkali-ion battery, the electrode should have high specific capacity (charge stored per unit mass or volume), high operating voltage, reasonable electron and ionic conductivity, and good phase and electrochemical stability. Suitable electrolyte selection for a pair of electrodes in alkali-ion battery is another important aspect. To prevent the reduction and oxidation of the electrolyte, it is necessary that Fermi
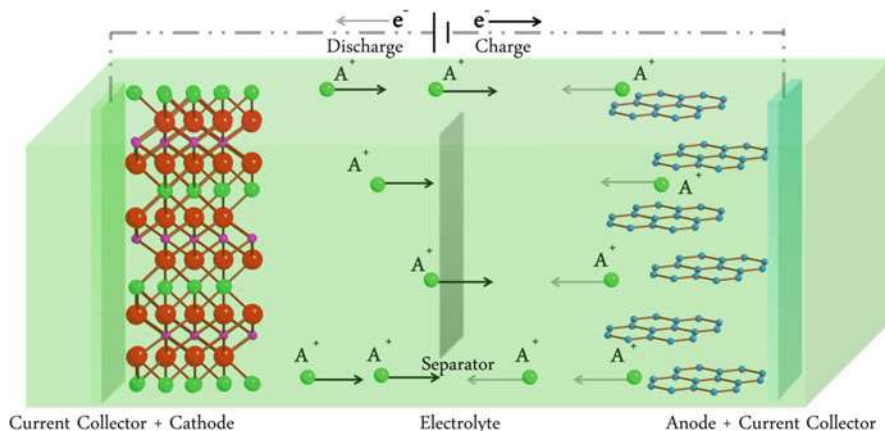
Fig. 1 Schematic of a rechargeable alkali-ion battery with transition metal oxide as cathode (green, alkali-ion; purple, TM; red, oxygen ion), carbon-based material as anode

level of anode should be lower in energy than the lowest unoccupied molecular orbital (LUMO) of the electrolyte, while Fermi level of the cathode should be higher in energy than the highest occupied molecular orbital (HOMO) of the electrolyte (Roy and Srivastava 2015). In the context of solid-state electrolytes, LUMO (HOMO) is referred to as the conduction band minimum (valence band maximum). Additionally, the electrolyte, especially for the solid-state electrolyte, is expected to have good interfacial stability, facile ion conductions which are more stringent than electrodes, and excellent electronic insulations.

Due to the significant advances in computation power in recent years, computer simulations have become an important complementary tool kit to experiments for the understanding and design of materials. First-principles density functional theory (DFT), (Hohenberg and Kohn 1964) especially the framework developed by Kohn and Sham (1965) that evaluates various properties of materials by solving single-particle Kohn-Sham equation with minimal approximation, has been particularly successful in this context. First-principles calculations allow researchers to evaluate the key properties of materials from an atomistic level, and hence provide invaluable insights into the fundamental physical processes that may not be directly assessable by experiments. Moreover, computation-guided discovery of novel materials has made tremendous progress for energy applications, including new alkali-ion battery materials (Deng et al. 2016; Urban et al. 2016). This is facilitated not only by the increasing computer power but also the development of the automated high-throughput first-principles screening that allows rapid identification of the most promising candidates from a pool of hundreds of thousands of initial candidates.

In the following sections, we will give a general introduction of well-established first-principles modeling techniques used to elucidate the thermodynamics and kinetics of battery materials.

2   Thermodynamic Approaches

2.1   Equilibrium Voltage

The operating voltage is a key metric of a battery to evaluate the stored energy density. By definition, the voltage difference between electrodes of a battery is equal to the difference in electron electrochemical potentials of the electrodes:

$$V = -\frac{\eta_{e^-}^{cathode} - \eta_{e^-}^{anode}}{e} \tag{1}$$

where $\eta_{e^-}^{cathode}$ and $\eta_{e^-}^{anode}$ are the electrochemical potentials of electrons (Fermi levels) in the cathode and anode. Inserting alkali chemical potential in any material, $\mu_A^X = \eta_{e^-}^X + \eta_{A^+}^X$, into Eq. 1 yields:

$$V = -\frac{\left(\mu_A^{cathode} - \mu_A^{anode}\right) - \left(\eta_{A^+}^{cathode} - \eta_{A^+}^{anode}\right)}{e} \tag{2}$$

where $\eta_{A^+}^{cathode}$ and $\eta_{A^+}^{anode}$ are the electrochemical potentials of alkali-ion in the cathode and anode. When a battery is at thermodynamic equilibrium ($\eta_{A^+}^{cathode} = \eta_{A^+}^{anode}$), the voltage difference between electrodes can be expressed in terms of the difference in alkali chemical potential:

$$V = -\frac{\mu_A^{cathode} - \mu_A^{anode}}{zF} \tag{3}$$

where z is the amount of charge transferred, and F is the Faraday constant. By definition, $\mu_A = dG/dN_A$ is the change of the Gibbs free energy of the electrode material with alkali-ion concentration (Aydinol et al. 1997). The voltage is thus a function of free energy change of the overall cathode/anode reactions (Nernst equation):

$$\overline{V} = -\frac{\Delta G_r}{zF} \tag{4}$$

where $\overline{V}$ is the average voltage by considering over a finite amount of reactions between cathode and anode.

Assuming that the entropic and volumetric effects are small at low temperature, the Gibbs free energy change for alkali insertion is often approximated by the DFT total energy difference for two alkali-ion concentrations $\Delta G_r = \Delta E_r + p\Delta V_r - T\Delta S_r \approx \Delta E_r$. Taking the lithium-layered oxide cathode Li_xMO_2 with a Li metal anode as an example, the overall reaction is given as follows:

$$Li_{x2}MO_2 + (x_1 - x_2)\,Li \rightarrow Li_{x1}MO_2 \tag{5}$$
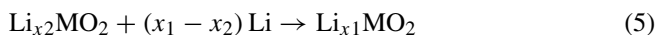
with $x_1 > x_2$. Then the reaction voltage can be calculated as:

$$V_{(x_1,x_2)} \approx -\frac{E\left(\mathrm{Li}_{x_1}\mathrm{MO}_2\right) - E\left(\mathrm{Li}_{x_2}\mathrm{MO}_2\right) - (x_1 - x_2)\, E\left(\mathrm{Li}\right)}{(x_1 - x_2)\, F} \tag{6}$$

The energies of $\mathrm{Li}_x\mathrm{MO}_2$, MO_2, and Li can be obtained via DFT total energy calculations. This approximation is not limited to insertion or intercalation electrodes but can also be applied to conversion or alloy reactions.

Most early DFT calculations of average voltage for battery materials were based either on the local density approximation (LDA) or the generalized gradient approximation (GGA) for the exchange-correlation potential (Kohn and Sham 1965; Langreth and Mehl 1983). However, for the modeling of reactions involving changes in electronic states, neither LDA nor GGA are good approximations due to improper cancelation of self-interaction error. Many TM oxides have strongly localized d electrons at the metal center, which leads to the large error of voltage predicting based on LDA or GGA functional. In this regard, an inclusion of a Hubbard U in DFT calculations (aka DFT $+ U$) was developed in the 1990s that significantly improves the DFT results for the system with strong correlated electronic structure (Anisimov et al. 1991, 1993). The method is controlled by Hubbard U parameter, which can be derived for each species either self-consistently (Cococcioni and de Gironcoli 2005) or more often, via fitting to experimental properties such as the oxide formation energy. The idea is borrowed from the Hubbard model in solid-state physics, to describe the transition between conducting and insulating systems. The larger the U value, the greater the penalty for shared electrons, driving the system toward integer electronic occupations (Cococcioni and de Gironcoli 2005). With proper choice of the U parameter, DFT $+ U$ calculations significantly improve the accuracy of predicted voltages over pure LDA or GGA for TM oxides (Zhou et al. 2004). More recently, DFT calculations using screened hybrid functionals, such as the Heyd-Scuseria-Ernzerhof (HSE) functional, (Heyd and Scuseria 2004; Heyd et al. 2003, 2006) have been shown to yield a similar accuracy as GGA $+ U$ without the need for an adjustable U parameter, though at significantly higher computational cost. A detailed comparison of calculated voltages for lithium TM oxides with GGA, GGA $+ U$, and HSE06 can be found in an earlier work by Chevrier et al. (2010)

2.2   Voltage Profile

Calculating the voltage profile requires the knowledge of stable intermediate phases during alkalination/de-alkalination process. The formation energy calculations allow one to determine the ground state energy of the stable phase at all intermediate concentrations. Taking the layered transition metal oxide $\mathrm{Li}_x\mathrm{CoO}_2$ as an example, the formation energy with respect to the stable end members (LiCoO_2 and CoO_2) is expressed as:

$$E_f\left(\mathrm{Li}_x\mathrm{CoO}_2\right) = E\left(\mathrm{Li}_x\mathrm{CoO}_2\right) - x\, E\left(\mathrm{LiCoO}_2\right) - (1 - x)\, E\left(\mathrm{CoO}_2\right) \tag{7}$$
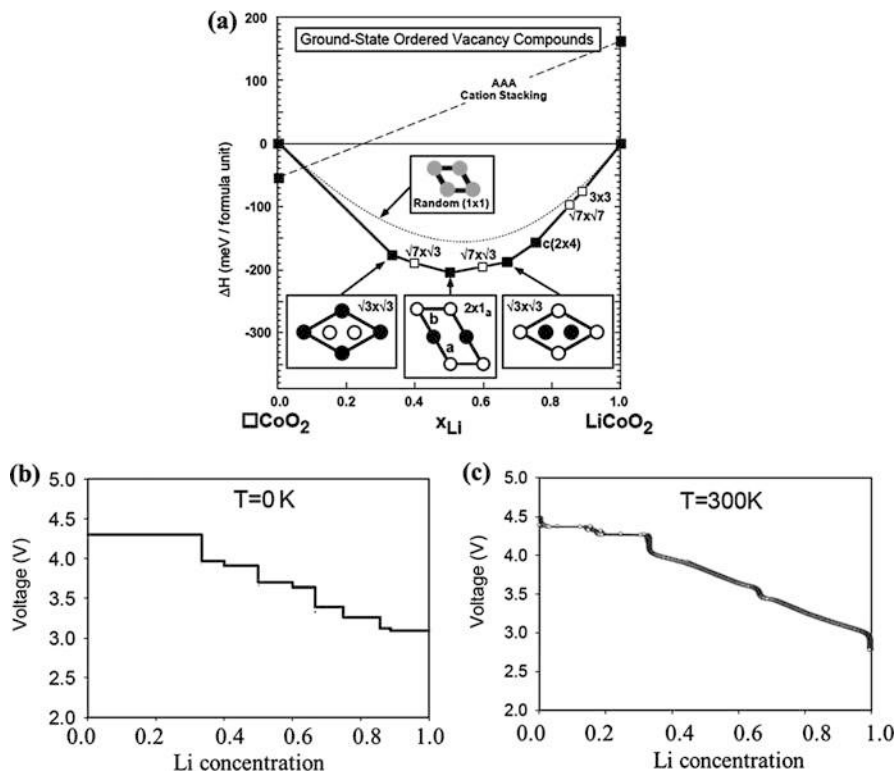
Fig. 2 (**a**) Stability convex hull (black line) of Li_xCoO_2. Filled square, DFT results; open square, cluster expansion results; filled circles, Li atoms; empty circles, Li-vacancy. (**b**) Predicted Li intercalation voltage of the Li_xCoO_2/Li cell as a function of Li concentration at T = 0 K. (**c**) Equilibrium voltage profile at T = 300 K calculated from Monte Carlo simulations. (Reproduced from Ref. (Wolverton and Zunger 1998) with permission. Copyright (1998) American Physical Society)

The resulting DFT-computed formation energy vs. composition curve (convex hull) is generated as shown in Fig. 2a. In this pseudo-binary stability diagram, the intermediate phases that are thermodynamically stable with respect to the end members ($LiCoO_2$ and CoO_2) form the convex hull. When the energy of an intermediate phase is above the convex hull (e.g., the dotted line in the figure), it implies that such intermediate phase is energetically unstable against its adjacent stable intermediate phases (filled and open squares in the figure) on the convex hull. At each concentration, different configurations of Li-vacancies and oxygen stacking sequence are considered. Examples of some of the predicted Li-vacancy ordering in Li_xCoO_2 are also illustrated in the figure. After the stable intermediate phases are identified, stepwise average voltage can be computed based on Eq. 6, which leads to the voltage profile in Fig. 2b.

It should be noted that the DFT-computed voltage is an approximation without considering entropic effects, and the accessible alkali orderings are also limited by

the supercell size utilized. These approximations work reasonably well for average voltages, especially for two-phase reactions involving ordered phases (e.g., olivine $LiFePO_4$). To obtain a full voltage profile and account for the configurational entropy (important for solid solution regions), one approach is to fit the DFT energies of a large number of alkali concentrations and orderings for a host material to a cluster expansion Hamiltonian. As shown in Fig. 2a, additional stable intermediate phases of Li_xCoO_2 (indicated by open squares on the black line) are discovered by using such approach. The key assumption here is that the contribution to free energy from the other degrees of freedom (e.g., vibrational, electronic) is insignificant and can be coarse-grained out. Monte Carlo simulations, e.g., using the Metropolis algorithm, can then be carried out using the Hamiltonian to sample the configurational space and evaluate finite temperature voltage profile. Details of the cluster expansion method and Monte Carlo simulation can be found in some excellent reviews (Urban et al. 2016; Meng and Arroyo-de Dompablo 2009). As an example, Wolverton and Zunger applied first-principles-based cluster expansion and Monte Carlo methods to model the Li-vacancy orderings in Li_xCoO_2, resulting in the temperature-dependent voltage profiles (see Fig. 2b and c for $T = 0$ and 300 K). Temperature effect mainly results in a continuous voltage slope, which smooths out the voltage steps that are present in the 0 K limit.

2.3   Electronic Structure

The relative energies of electronic levels for different components in alkali-ion battery determine the limitation of the stored energy. Figure 3a schematizes the relation between the equilibrium voltage and electronic structures of the electrode, electrolyte. At open circuit condition, the energy difference of Fermi levels between the anode and cathode builds electron electrochemical potentials. Assuming there
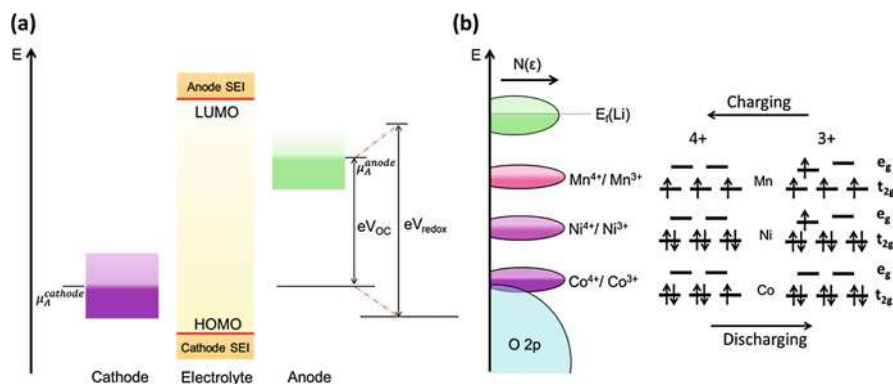
Fig. 3 (**a**) Relative energies of the electrolyte window and the electrode chemical potentials $\mu_A^{cathode}$ and μ_A^{anode} with no electrode/electrolyte reaction. (**b**) Relative positions of energy levels and typical electron configurations in layered Ni-Mn-Co oxide cathode

is no electrode/electrolyte reaction, this electron potential difference equals the difference in alkali chemical potential between the anode and cathode, which is measured as open circuit voltage (V_{OC}). Once charging current is applied, electrons are further pumped out from the valence band of cathode through the circuit to the conduction band of the anode, resulting in the downshift and upshift for the Fermi level of cathode and anode, respectively. Alkali ions are extracted from cathode to the anode with the oxidation change of other ions, e.g., TM ions to balance the charge. The whole process is fully reversible during the discharging process in an ideal case. The redox potential (V_{redox}) measured during charging/discharging process is limited by the "window" of the electrolyte. If the potential of the anode and cathode shift beyond the LUMO and HOMO of the electrolyte, the electrolyte is reduced on the anode and oxidized on the cathode to form the solid electrolyte interphase (SEI). The ideal SEI is an ionic conductor and an electron insulator, which prevents further decomposition of the electrolyte and thus increases stable voltage window.

The electronic levels of the electrode involved in the redox reaction arise from the d-states of the TM cations in most cases. The Fermi level energy as well as the density of states (DOS) near the Fermi level for redox center can be obtained from DFT calculations. Figure 3b qualitatively summarizes the positions of the redox active energy levels for Co^{4+}/Co^{3+}, Ni^{4+}/Ni^{3+}, and Mn^{4+}/Mn^{3+}, as well as the oxygen $2p$ band in TM oxides. In most of the structures including layered, spinel, and disordered rock salt, TM ions occupy octahedral sites that are coordinated by six oxygen atoms. Based on crystal field theory, TM five d orbitals split into two sets of levels labeled t_{2g} and e_g as shown in Fig. 3b. Each of the level can hold two electrons with opposite spin. The commonly observed electronic configurations of d orbitals for Co^{4+}/Co^{3+}, Ni^{4+}/Ni^{3+}, and Mn^{4+}/Mn^{3+} are also illustrated in Fig. 3b. The lone electron in e_g level triggers Jahn-Teller distortion, breaking the degeneracy between the two e_g orbitals to lower the overall energy for the case of Mn^{3+} and Ni^{3+}. When the battery is charged, TM cations are oxidized to 4+ from 3+, and they are reduced back to +3 oxidation state as they receive electrons from the anode during discharging. More importantly, the Co^{3+}/Co^{4+} level is the lowest in energy because Co redox involves the more stable t_{2g} orbitals, while Ni and Mn redox involves the less stable e_g orbitals. This relative alignment of energy levels directly indicates redox reaction sequence of the electrode containing different TM ions since lower energy levels correspond to higher redox voltages.

Although the electronic structure is in general complex, it provides another key insight into the available number of electrons near Fermi level. If an alkali-ion is extracted from the cathode host structure, an electron is pumped out from the valence band at the same time. Therefore, the more the available number of electrons are, the higher the alkali-ion storage capacity is. In classical layered cathode, e.g., $LiCoO_2$ and $LiNi_{0.5}Mn_{0.5}O_2$, oxygen ions are coordinated by three Li and three TM ions as shown in Fig. 4a. In this configuration, a much larger projected DOS originating from the TM states than from the O states between 0 and $-1.5\,eV$ below Fermi level (commercial organic liquid electrolyte window) is found, indicating that the redox reactions are dominated and restricted by the TM $3d$ electrons. On the
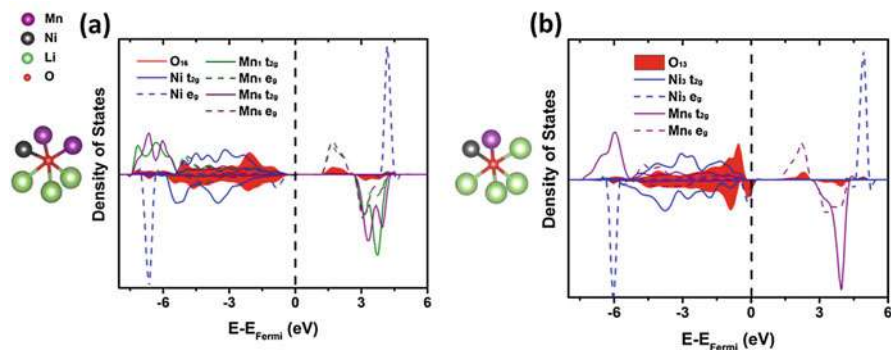
Fig. 4 An example of an oxygen site in the fully lithiated structure (**a**) without a Li-O-Li linear bond and (**b**) with a Li-O-Li linear bond, required for labile oxygen states

other hand, Li-excess is created by the substitution of some TM by Li, leading to Li-O-Li configurations in many of the new exciting cathode materials. It shows an increased oxygen projected DOS below the Fermi level for oxygen with a linear Li-O-Li bond based on GGA $+ U$ calculations (see Fig. 4b). This labile oxygen observed computationally with increased projected DOS below the Fermi level paves the completely new research on anionic redox and breaks the bottleneck for reaching higher energy density in alkali-ion battery (Grimaud et al. 2016; Okubo and Yamada 2017).

The electronic band gap deduced from the calculated DOS is also useful to investigate the electronic properties of a potential electrode or solid electrolyte material. While high electronic conductivity for the electrode material is a necessary condition for high rate performance of alkali-ion battery, low electronic conductivity is required for the potential electrolyte materials. Similar to the voltage calculation, the introduction of a U correction term in the GGA method could substantially improve the accuracy of calculated band gaps (Zhou et al. 2004). Other helpful information that can be extracted from electronic structure analysis includes chemical bonding, orbital hybridization, magnetic properties, etc.

2.4 Stability Analyses

2.4.1 0 K Phase Stability

The pseudo-binary stability diagram mentioned in the previous subsection essentially works for electrode materials. In a general situation, the thermodynamic stability of any given compound (e.g., solid electrolyte) can be analyzed by evaluating its decomposition energy referenced to the linear combinations of stable compounds in this compositional space. This can be achieved by constructing the convex hull using the DFT energy of all known compounds in this compositional space. In the case of olivine LiFePO$_4$, for example, the related stability analysis

requires the DFT energies of all existing $Li_xFe_yP_zO_u$ compounds (x, y, z, and u are nonnegative integers) including itself in Li-Fe-P-O quaternary compositional space. The stability metric under this circumstance is called the energy above the hull (E_{hull}), (Sun et al. 2016; Ong et al. 2008) which measures how stable a compound is by measuring how high its energy is above the convex hull. It is always nonnegative, and it equals zero for stable compounds. The higher the E_{hull}, the more unstable the compound tends to be. With the advent of large open-source materials databases such as the Materials Project (Jain et al. 2013), a common approach is to leverage on pre-computed DFT energies in these databases for compounds other than that of primary interest, which significantly reduce the computational effort involved, especially for high-dimension compositional space. E_{hull} is frequently adopted as an efficient descriptor for phase stability and synthesizability in high-throughput first-principles screening studies to rapidly rule out structures that are unlikely stable at ambient temperatures (high E_{hull}).

2.4.2   Surface Stability

The real surface configuration rarely coincides with the ideal lattice plane which is a geometrical abstraction of the crystal structure. In reality, the stable surface can contain charged species adsorbed from the environment, charge density rearrangement, and stepped or kinked surface profile. This makes determination of surface thermodynamics very challenging. For this reason, first-principles calculations can be a useful tool to determine facet-specific surface energies. The standard model to calculate surface energies is a two-dimensional surface slab with bulk stoichiometry oriented to expose the facet of interest, with a sufficiently large vacuum layer added on top of the slab to ensure that there is minimal interaction between the slab and its periodic images along the surface normal. The k-point grid for surface slab is thus generated as $k_1 \times k_2 \times 1$, which assures no energy dispersion in the direction out of the surface.

Standard procedures of generating surface slabs include basis transformation, atomic coordinates rescaling, termination choice, and polarity check. Surfaces are typically defined using the Miller index notation by the conventional unit cell. However, first-principles calculation requires the basal plane to be coincident with the constructed slab surface. The basis of the conventional unit cell has to be transformed so that (001) plane of the new basis is parallel to the Miller plane of interest. The third vector should be as orthogonal to the basal plane as possible with a relatively short length. The atomic coordinates are then redefined based on the new basis. For a given surface orientation, a number of termination layers usually exist due to different atomic configurations and can be chosen by placing the basis at different positions along the third vector. Note that a charged surface with a perpendicular dipole moment in a repeated unit (type III according to surfaces classification by Tasker) produces a polarizing electric field in the bulk, leading to infinite periodic cell energies (Tasker 1979). For example, the (111) face of NiO ionic crystal has alternating layers of Ni^{2+} and O^{2-} ions, which forms a permanent dipole moment perpendicular to the surface. Atomic reconstruction by moving half

the charged species on one surface to the other side to cancel out the 2D-dipole is a common approach (the so-called Tasker III to Tasker II reconstruction) to stabilize type III surface (Bruno 2013).

From DFT calculations, the surface energy can be computed using the following expression:

$$\gamma = \frac{1}{2A}\left[E\left(slab\right) - N \cdot E\left(bulk\right)\right] \qquad (8)$$

where A is the surface area of the slab, E(slab) is the surface slab energy, E(bulk) is the bulk energy per atom, and N is the number of atoms in the surface slab. The prefactor 1/2 comes from the fact that each surface slab contains two equivalent surfaces. The Wulff shape, which is the equilibrium shape of the crystal, can be constructed using the lowest surface energies of different facets. The larger the surface area of a facet contributes to the Wulff shape, the more stable that facet tends to be. As an example, surface properties of olivine $LiFePO_4$ cathode were studied by Wang and co-workers (Wang et al. 2007). The two low-energy surfaces (010) and (201) are found to make up almost 85% of the surface area of Wulff crystal shape. More importantly, redox potential for the (010) surface was calculated to be much lower than the bulk value, which shows the anisotropic property of surface redox potentials. This study pinpoints the effect of surface energies and crystal shape on the electrochemistry for electrode materials composed of particles at nanoscale. Surface electronic structure, surface reconstruction, and adsorbate interactions can also be studied based on surface stability calculations.

2.4.3   Electrolyte/Electrode Interface Stability

Understanding of electrolyte/electrode interfaces has become a recent focus of the greatest importance for the performance improvement of electrochemical energy storage systems. For alkali-ion batteries with liquid organic electrolyte, SEI formation due to electrolyte decomposition is critical to cycling performance and safety. For all solid-state alkali-ion batteries with sulfide-type electrolytes, the rate-determining process is usually observed around the cathode/electrolyte interfaces. To predict the electrode/electrolyte interface stability under large variations in environmental conditions is one of the grand challenges for battery study. First-principles calculation using DFT-based methods has been developed to investigate most stable interfacial matching of two crystal solids (Haruyama et al. 2014). In a traditional manner, the procedure to investigate the interface formed by two compounds comprises three major steps: (i) identify the lowest-energy surface facet of each material as detailed in surface stability calculations; (ii) patch the two surface slabs using a reasonably large supercell (with a vacuum layer of ~10 Å) to ensure as small lattice parameter misfit as possible; and (iii) enumerate all possible configurations for the interface supercell including slide of one surface slab with respect to the other, and perform DFT relaxations. The alkali-ion sites and alkali-ion transfer properties in the most stable interface can then be further investigated under different thermodynamic equilibrium conditions.

The explicit interface calculations often require very high computational cost because of the supercell used. For studies of electrolyte/electrode interfacial stability, a critical step is to identify the possible reaction products formed at the interface. In the case of solid electrolyte/electrode interfaces, there are two efficient analysis methods based on two different thermodynamic approximations for such purposes (Zhu et al. 2015a; Richards et al. 2016; Tang et al. 2018).

- *Fast ionic diffusion limit*: In this approximation, the diffusing ions in the solid electrolyte are assumed to be the main mobile species. Under such conditions, the interface can be modeled as an open system with respect to the mobile species. The relevant thermodynamic potential is the grand potential. Taking the Li-ion solid electrolyte as a concrete example, the Li grand potential is expressed as $\varphi = E - \mu_{Li}N_{Li}$, where E, N_{Li}, and μ_{Li} are, respectively, the internal energy, the number of Li atoms in the open system, and the Li chemical potential. When the metallic Li is taken as the anode, the Li solid electrolyte/anode interface is modeled as the solid electrolyte at high $\mu_{Li} \approx \mu_{Li0}$, and the Li solid electrolyte/cathode interface is modeled as the solid electrolyte at low $\mu_{Li} \approx \mu_{Li0} - V$, where V is the cell voltage and μ_{Li0} as the chemical potential in metallic Li.

- *Multi-species equilibrium*: The second method assumes full thermodynamic equilibrium between the solid electrolyte and electrode in varying electrolyte/electrode compositional ratios x (between 0 and 1). Specifically, the equilibrium composition at a given ratio x corresponds to $c_{electrolyte}x + c_{electrode}(1 - x)$ with c_y being the composition of compound y. To estimate the likely reaction products at different x values, a pseudo-binary stability diagram that is like the aforementioned stability diagram for electrode materials is constructed. In this pseudo-binary stability diagram, however, the x values correspond to the compositional ratio x, whereas y values are the associated reaction energy which reaction products have the same composition as $c_{electrolyte}x + c_{electrode}(1 - x)$. This approach tends to be more reflective of the conditions during actual synthesis and assembly conditions.

Note that the approaches presented here are within thermodynamic approximations. To capture the kinetic effects on the formation of electrolyte/electrode interfaces, computationally more demanding approaches, e.g., ab initio molecular dynamics can be adopted (see Sect. 3.2.6).

2.4.4   Defects and Dopability

Extrinsic doping is a common strategy to further optimize battery material properties, such as phase stability, ionic and electronic conductivity, and cost. It is crucial to predict the feasibility of introducing a dopant into a crystal structure based on first-principles calculation. The size of the supercell model is inversely proportional to the dopant concentration. Note that different types of doping, including substitutional, interstitial, and anti-site, need to be considered for the supercell construction. Generally speaking, the dopability of a species depends on

its ionic radius and oxidation state relative to existing species in the crystal. In the case of charge balance doping, the neutral dopant formation energy can be defined as follows:

$$E_f\,(\text{doped}) = E\,(\text{doped}) - E\,(\text{pristine}) - \sum_{i=1}^{N} n_i \mu_i \qquad (9)$$

where $E(\text{doped})$ and $E(\text{pristine})$ are the energies of the structure with and without the neutral dopant, respectively, μ_i is the atomic chemical potential of species i, n_i is the number of atoms of species i being added or removed, and N is the total number of species in the doped structure. Synthesis condition can be selected based on the energy calculation to tune species chemical potential so that the formation energy can be low enough to prevent any species precipitation. Additionally, the doped structure should remain stable without any decomposition during synthesis.

3   Kinetics Approaches

There are two widely used first-principles approaches that can assess the kinetics of alkali-ion diffusion in battery materials: transition state theory-based methods such as the nudged elastic band approach and ab initio molecular dynamics simulations. In the following, we will present the key concept of each approach for ionic diffusion studies as well as their limitations.

3.1   Transition State Theory and the Nudged Elastic Band Method

In materials where the equilibrium lattice sites are well defined, the ionic diffusion can be investigated using transition state theory. From an atomistic viewpoint, diffusion can be viewed as a series of hopping events of ions between neighboring sites. In each hop, an ion migrates from one site to its neighboring vacant site via an activated state (aka the transition state), which has the highest Gibbs free energy along the minimum energy path between the two sites. Based on the transition state theory (Deng et al. 2016; Urban et al. 2016), the rate k at which the hopping event occurs can be expressed as:

$$k(T) = \upsilon^*(T) e^{-\frac{\Delta G_a(T)}{k_B T}} \qquad (10)$$

$$\Delta G_a(T) = G_a(T) - G_i(T) \qquad (11)$$

Here, $\Delta G_a(T)$ is known as the activation barrier, and it is the difference of the Gibbs free energies at the activated state ($G_a(T)$) and initial state ($G_i(T)$). $\upsilon^*(T)$ is the effective attempt frequency at temperature T. In the dilute limit of carrier

Fig. 5 An example of NEB computed migration barriers. The two example structures are LiCoO$_2$ (LCO) and NaCoO$_2$ (NCO). (Reproduced from Ref. (Ong et al. 2011) with permission. Copyright (2011) Royal Society of Chemistry)
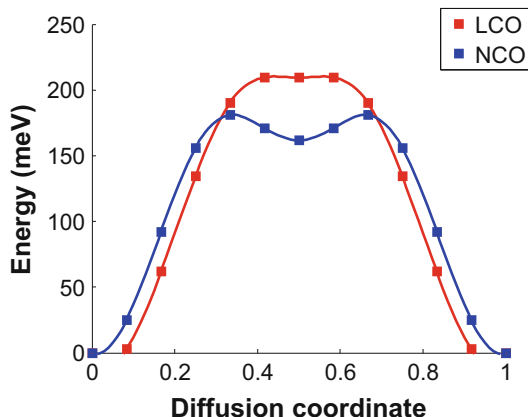
concentration, the diffusivity can be approximated as $D(T) \sim d^2 k(T)$ where d is the hopping distance (assumed to be known). Moreover, the attempt frequency $\upsilon^*(T)$ is often assumed to be temperature-independent, and the typical values fall between 10^{11} and 10^{13} s^{-1} (Urban et al. 2016). When the diffusivity follows the Arrhenius relation (see Eq. 18 below), the activation barrier in Eqs. 10 and 11, $\Delta G_a(T)$, can be assumed to be temperature independent. When the entropic effects become negligible, $\Delta G_a(T)$ can be further approximated as DFT energy difference between the activated state and the initial state ($\Delta G_a(T) \approx \Delta E_a$).

For ionic diffusion studies, the nudged elastic band (NEB) method (Jonsson et al. 1998) is a widely used approach to determine the 0 K activation barriers of the ion migration paths for a given atomic configuration. This approach requires the initial and final states of the migration path (aka end points) as input, from which a given set of intermediate states (aka images) are generated to serve as the initial guess of the migration path. The minimum energy path is determined by minimizing the atomic forces in all the images subject to the constraint of a harmonic coupling (in terms of harmonic springs) between adjacent images. In the end, the activation barrier (ΔE_a) is computed as $\Delta E_a = E_{max} - E_i$ where E_{max} is the highest energy along the optimized migration path and E_i is the energy of the initial state. An example of NEB results is shown in Fig. 5 for layered oxides ACoO$_2$ (A = Li and Na) (Ong et al. 2011).

In a typical NEB calculation, it essentially comprises the following four steps:

1. Initial structural optimization calculations of the pristine crystal structure.
2. Structural optimizations of the two initial and final state structures.
3. An interpolation between the initial and final states to generate a set of images along the migration path. The linear interpolation is the most commonly used scheme for such purpose.
4. An NEB calculation to find the lowest-energy migration path, from which the highest energy is determined, and hence the activation barrier.

In the studies of ionic diffusion, the NEB method can be applied to evaluate the barrier of any given migration path. This makes it particularly useful to probe the effects of local environment (e.g., local arrangement of the framework) on the ionic diffusion. Indeed, the NEB calculations have been employed to understand the diffusion mechanisms in several important battery materials, e.g., layered $LiCoO_2$ and olivine $LiFePO_4$ cathodes (Deng et al. 2016; Urban et al. 2016). Moreover, the migration barriers from NEB calculations can also serve as inputs for other techniques, including the kinetic Monte Carlo (KMC) simulations. This is a Monte Carlo simulation method that studies the time evolution of natural processes using lattice models, e.g., the diffusion of alkali ions in the battery materials. Using the local environment-dependent barriers from NEB calculations, a first-principles cluster expansion can be carried out to parameterize these barriers for Monte Carlo simulations. More details about the KMC method, including the basic steps in the simulations, can be found in literature (Deng et al. 2016; Urban et al. 2016).

3.1.1   Limitations of the NEB Method

In studies of alkali-ion batteries, the NEB method has been mainly applied to the bulk systems (pristine electrodes and solid electrolytes). In the case of noncrystalline systems, e.g., electrolyte/electrode interfaces, much higher computational efforts are inevitably required due to the large number of distinct ionic migration paths that need to be considered due to the lower symmetry. There are two further limitations of the NEB method compared to the ab initio molecular dynamics simulations (see Sect. 3.2).

First, the start and end points of the migration path must be specified as inputs for NEB calculations. For alkali battery materials with complex diffusion properties, e.g., when ionic motions are highly concerted (common in superionic conductor electrolytes), the identification of the relevant migration events and local environment become very challenging. In this scenario, ab initio molecular dynamics (see Sect. 3.2) becomes the more appropriate technique to use. For this reason, NEB calculations are primarily performed in dilute carrier limits, e.g., single alkali-ion vacancy migration in an otherwise fully alkaliated structure or single alkali migration in an otherwise alkali-ion-free framework structure.

Second, NEB calculations essentially output the migration barrier at 0 K, i.e., the temperature effects are not considered in the calculations. Besides, to estimate an accurate hopping rate associated with the migration path, and hence the diffusivity, the prefactor in Eq. 10 must be computed explicitly. However, it is often assumed to be some value between 10^{11} and 10^{13} s^{-1}.

3.2   Ab Initio Molecular Dynamics Simulations

Molecular dynamics (MD) simulates the atomic motions at finite temperatures by integrating the Newtonian equations of motions. A key input of this approach is the description of the interatomic interactions, whereas the main output is the trajectories of the atomic motions. Depending on how the interatomic interactions are

obtained, MD simulations can be categorized into two general classes: classical MD and ab initio MD (AIMD) simulations. In classical MD simulations, the interatomic interactions are expressed in terms of analytical functions with parameters fitted using either experimental or first-principles data, whereas in AIMD simulations, the interatomic interactions are obtained by directly solving the Schrödinger equation with approximations.

Although classical MD simulations are computationally less expensive than the ab initio counterpart, it does suffer from serious drawbacks stemming from the use of empirical potentials, e.g., poor transferability across different chemistries, and the difficulty in dealing with complex interatomic interactions with simple function forms. On the other hand, AIMD simulations can generally be applied across broad chemical spaces, and the accuracy of the interatomic interactions is essentially limited by the underlying approximations of the ab initio approach.

There are two main variants of AIMD models today: Born-Oppenheimer (BO) variant and Car-Parrinello (CP) variant (Car and Parrinello 1985; Hutter 2012). For battery-related studies, the former variant is more commonly adopted. The theory behind these AIMD techniques is beyond the scope of this chapter and is extensively covered by many excellent textbooks and reviews (Hutter 2012; Marx and Parrinello 1996). The main disadvantage of AIMD methods is the significantly higher computational cost compared to the classical MD, and also the NEB method (Sect. 3.1). This places constraints on the accessible cell size as well as time scale in AIMD simulations. However, these can be mitigated by the advances in computer power and the nature of the problem under investigation.

The key output from AIMD simulations is the trajectories of ionic motions, from which the diffusion properties of materials, e.g., ionic diffusivity and conductivity, can be evaluated. The AIMD technique has gained tremendous successes in the studies of alkali-ion battery materials, particularly for the understanding and discovering novel alkali-ion superionic conductors. In the following, we will first sketch the well-established analysis methods and how they are applied to understand the ionic diffusions in battery materials. We will also discuss practical considerations on the parameter selection, the applications of AIMD in electrolyte/electrode interfaces, and the limitation of this technique. We should stress that these analysis approaches are agnostic of the types of MD simulations (i.e., classical and ab initio), and they are also applicable to other classes of materials, e.g., oxygen conductors in solid oxide fuel cells.

3.2.1 Ionic Diffusion and Conduction Coefficients

For a given 3D crystal structure with N mobile ions, the self-diffusion coefficient can be computed from AIMD simulations by evaluating the velocity-velocity autocorrelation function (Green-Kubo relation):

$$D^* = \frac{1}{3N} \int_0^\infty dt \sum_{i=1}^N \left\langle \overrightarrow{v_i}\,(t_0) \cdot \overrightarrow{v_i}\,(t + t_0) \right\rangle_{t0} \tag{12}$$

where $\vec{v}(t)$ is the velocity of ion i at time t and angular bracket in the integrand stands for an ensemble average over the initial time argument t_0.

This quantity is more frequently estimated from the ionic displacements via the Einstein equation:

$$D^* = \frac{1}{2dt} \lim_{t \to \infty} \frac{\partial \Delta \vec{r}(t)^2}{\partial t} \tag{13}$$

$$\Delta \vec{r}(t)^2 = \frac{1}{N} \left\langle \sum_{i=1}^{N} \left[\vec{r}_i(t + t_0) - \vec{r}_i(t_0) \right]^2 \right\rangle_{t0} \tag{14}$$

Here, $\vec{r_0}(t)$ is the position of mobile ion i, and $\Delta \vec{r}(t)^2$ is known as the mean square displacement (MSD) of the mobile ions over time t as an ensemble average over the initial time argument t_0.

We should stress that the two expressions (Eqs. 12 and 13) for the self-diffusion coefficient D^* are equivalent. However, due to the short accessible time scales in AIMD simulations that are limited up to a few hundred picoseconds in most cases, the convergence speed becomes an important factor to consider when computing the diffusion coefficient. In this context, Eq. 13 is more commonly adopted for computing D^* because it tends to converge more rapidly than that of Eq. 12 due to the averaging of the square displacements. Moreover, the long-time tail of the integral in Eq. 12 may also cause numerical inaccuracy when evaluating the diffusion coefficient in a short simulation time scale.

With the diffusion coefficient at temperature T in hand, the ionic conductivity can be estimated using the Nernst-Einstein equation expressed below,

$$\sigma(T) = \frac{\rho z^2 F^2}{RT} D^*(T) \tag{15}$$

Here, ρ and z are, respectively, the molar density and charge of the mobile ions in the simulation cell. For example, z $= +1$ for alkali ions such as Li^+ and Na^+. F and R are the Faraday constant and the gas constant, respectively. We should point out that since the computed D^* corresponds to the tracer diffusivity in experiments, the conductivity derived from Eq. 15 assumes no correlations in ionic motions. Here, the correlation essentially describes how the displacement of one ion depends on those of other ions during the diffusion.

In many situations, such ionic correlation can be measured in terms of Haven ratio H_R, (Murch 1982)

$$H_R = \frac{D^*}{D_\sigma} \tag{16}$$

where D_σ is known as charge diffusivity defined below:

$$D_\sigma = \frac{1}{6N} \lim_{t \to \infty} \frac{\partial}{\partial t} \left\langle \left\{ \sum_{i=1}^{N} \left[\vec{r}_i\,(t+t_0) - \vec{r}_i\,(t_0) \right] \right\}^2 \right\rangle_{t0} \qquad (17)$$

The Haven ratio equals one when the ionic motions are uncorrelated, while it is less than 1 for correlated ionic motions. For fast ionic conductors, H_R typically falls between 0.3 and 0.6 (Morgan and Madden 2014; Bron et al. 2013). Given the fact that the convergence of D_σ in any molecular dynamics simulations is usually much slower than that of D^*, there have only been few AIMD studies that attempt to explicitly estimate the H_R value.

3.2.2 Activation Energy

Apart from the ionic diffusivity and conductivity, another key quantity of interest is the activation energy E_a. Under the assumption that there are no phase transitions, one can perform AIMD simulations at multiple temperatures and fit the temperature-dependent diffusivity D to the Arrhenius relationship:

$$D = D_0 e^{-E_a/kT} \qquad (18)$$

Here, D_0 stands for the diffusivity at temperature $T \to \infty$, and k denotes the Boltzmann constant. One can obtain E_a via a linear fitting of the log of D vs. 1/T (see Fig. 6a as an example). We should note that AIMD simulations are usually performed at elevated temperatures, e.g., above 600 K, to increase the number of diffusion events. The diffusivity at ambient temperatures can also be obtained from Eq. 18 via linear extrapolation. Nevertheless, we should stress that the scheme implies that the fundamental diffusion mechanisms remain unchanged between the entire temperature range of interest.
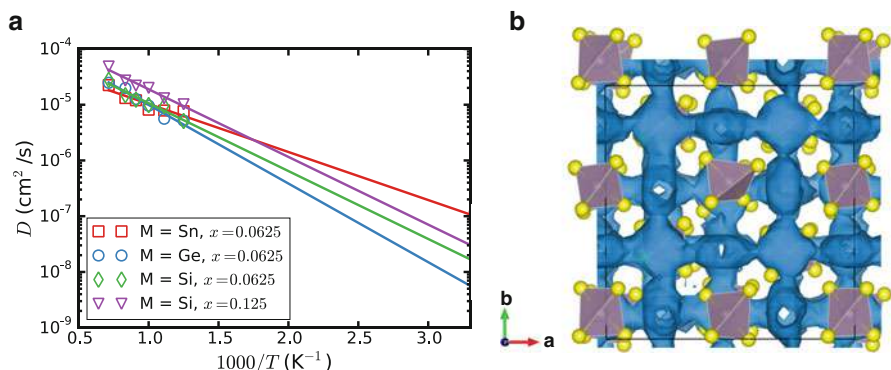
Fig. 6 (**a**) Arrhenius plot of doped structures $Na_{3+x}M_xP_{1-x}S_4$ (M = Si, Ge or Sn). (**b**) Isosurface (blue) of Na^+ probability density distribution P for $Na_{3.0625}Si_{0.0625}P_{0.9375}S_4$ at 800 K. (Reproduced from Ref. (Zhu et al. 2015b) with permission. Copyright (2015) American Chemical Society)

3.2.3 Analysis of the Diffusion Process

Besides the ionic diffusivity and conductivity, the ionic trajectory data also provide rich details about diffusion process of ions, including the energy landscape in the battery material and the related diffusion pathways for the mobile ions. The latter information can be obtained by calculating the probability density function $P\left(\overrightarrow{r}, T\right)$, which is a function of both the temperature and a spatial 3D grid (typically a uniform grid). Specifically, for an input trajectory data with a given time scale and temperature T, P is computed by first counting the total number of mobile ions occupying each grid point using a distance-based algorithm. This is followed by a normalization such that $\int_0^{\Omega} P d \overrightarrow{r} = 1$ with Ω being the volume of the unit cell. Two useful pieces of information can be extracted from the probability density function: (1) low-energy sites (corresponding to higher $P\left(\overrightarrow{r}, T\right)$ values) at where the mobile ions prefer to reside and (2) pathways for the ionic diffusion. As an example, Fig. 6b demonstrates the Na^+ probability density function in cation-doped sodium superionic conductor Na_3PS_4. We should point out that the occupancies of the symmetrically distinct sites can also be estimated using this function, which provides useful message for further understanding the ionic diffusion process.

The Van Hove correlation function is a well-known technique to assess the ionic correlations during the diffusion process. Van Hove correlation function is often split into two parts, self-part G_S and distinct-part G_d, and they are expressed as follows:

$$G_s\left(r, t\right) = \frac{1}{4\pi r^2 N_d}\left\langle\sum_{i=1}^{N_d}\delta\left(r - |\overrightarrow{r}_i\left(t + t_0\right) - \overrightarrow{r}_i\left(t_0\right)|\right)\right\rangle_{t0} \tag{19}$$

$$G_d\left(r, t\right) = \frac{1}{4\pi r^2 \rho N_d}\left\langle\sum_{i\neq j}^{N_d}\delta\left(r - |\overrightarrow{r}_i\left(t_0\right) - \overrightarrow{r}_j\left(t + t_0\right)|\right)\right\rangle_{t0} \tag{20}$$

Here, $\delta(\cdot)$ refers to the 1D Dirac delta function. N_d and r are, respectively, the number of mobile ions in the unit cell and the radial distance. ρ is the number density for the mobile ions that is used as the normalization factor such that $G_d \to 1$ when $r \gg 1$. In terms of physical interpretation, the self-part $G_s(r, t)$ can be considered as the probability density of a mobile ion that diffuses away from its initial position by a distance r after time t. The distinct-part $G_d(r, t)$ describes the radial distribution of $N_d - 1$ ions after time t with respect to the initial reference mobile ion. Note that when $t = 0$, $G_d(r, t)$ is reduced to the static pair distribution function.

3.2.4 Practical Considerations in AIMD Simulations

Since the AIMD techniques are computationally expensive, special care is required when selecting the relevant simulation parameters so that it achieves a good balance between the computational cost and the realistic results. In general, the selection of these parameters depends strongly on (i) the specific system under investigation, e.g., fast or slow conductors, and (ii) the type of problem of interest,

e.g., comparative evaluation for systems with similar chemistries/topologies, or accurate diffusivity calculations. For any AIMD simulations of interest, key parameters include the simulation cell size, total simulation time, time step, simulation temperature, the ensemble, and so on. An inclusive summary about the parameter selection can be found in literature for interested readers (Shin and Saal 2018).

3.2.5   Modeling Electrolyte/Electrode Interfaces

AIMD simulations can also be used to explicitly model the electrolyte/electrode interfaces, which play a crucial role in the battery performance during charge-discharge process. To obtain more realistic results, the simulation supercell size for such systems is required to be much larger than the supercell for bulk materials, especially along the lattice direction normal to the interface. In principle, the diffusion properties at the interfaces may be estimated via AIMD simulations. However, that requires significant computational efforts due to the required large supercell and long simulation time scales. More often, AIMD simulations are applied to simulate the structural evolutions during alkaliation/de-alkaliation process (Leung and Budzien 2010) and the chemical reactions between the electrolyte (solid or liquid) and electrode materials from which the likely interfacial reaction products can be identified (Tang et al. 2018). In those simulations, a smaller supercell size with a shorter time scale may be utilized to speed up the AIMD simulations and obtain realistic results in a qualitative manner. Analysis tools have been developed in recent years that aim to efficiently determine the interfacial products from AIMD simulations. Among them include the radial distribution function (RDF) analysis. In that method, the RDFs of all possible products in their crystalline phases are compared to that of the interface to identify the likely reaction products. We should note that all these existing techniques are based on various approximations, and they are required for further improvements.

3.2.6   Limitation of AIMD Simulations

Similar to the NEB method, AIMD simulations are mainly applied to the *bulk* systems for their diffusion properties, especially the alkali-ion superionic conductors. For alkali-ion superionic conductors, key quantities such as ionic diffusivity and conductivity can often be estimated within the accessible AIMD time scale that is typically up to a few hundred picoseconds. For bulk electrode material such as alkali transition-metal oxide cathodes, the presence of transition metals often requires spin polarization calculations. Also, the ionic conduction in the electrode materials is usually much slower than in the superionic conductors, implying that longer simulation time scale may be needed for properly converged ionic conductivity. Therefore, diffusion studies using AIMD simulations for these materials often require substantially higher computational cost, and key diffusion properties such as alkali-ion migration barriers and ionic conductivity are often estimated from NEB calculations (see Sect. 3.1) instead. Nevertheless, AIMD simulations can still be performed for the electrode materials for other purposes, e.g., to estimate the diffusion pathways of the alkali ions, site occupancy, etc.

As a final note for AIMD simulations, the ionic diffusivity (and activation energy) of a material at ambient temperatures is essentially obtained via linear extrapolation from those obtained at elevated temperatures. This implies that the diffusion mechanisms between the ambient temperatures and elevated temperatures must stay unchanged. For materials where such condition does not hold, the NEB method should be utilized instead.

4   Conclusion and Outlook

This chapter provides a general overview of the well-established first-principles methods used in battery-related studies (electrodes, electrolytes, and their interfaces) from both the thermodynamic and kinetics aspects. These include many key properties of alkali-ion battery: equilibrium voltage, phase stability, electronic structures, and ionic diffusion properties. First-principles calculations need to have closed-loop interaction with experimental design and data. Simulations provide insight into development of new materials and aid in the interpretation of experimental observations. Experimental results in turn prove the validity of the theoretical model and proposed mechanism. Despite the remarkable advances of the first-principles methods in recent years, there remain some pressing challenges in modeling technique when it comes to (1) noncrystalline systems, e.g., the presence of intrinsic defects and the electrolyte/electrode interfaces, and (2) the design of high-voltage high-capacity cathode materials.

First, it should be noted that most of the first-principles methods presented here, e.g., voltage calculations and NEB methods, have mainly been applied to perfect crystalline solids. From the experimental perspective, however, intrinsic defects such as dislocations, stacking fault, and micro-strain are always present in the synthesized material. These intrinsic defects can have significant impact on the electrochemical performance of the battery material. In principle, imperfections (e.g., dopants) can be modeled via the construction of supercell, and the size of the supercell depends mainly on the defect concentration. Given that there are often more than one possible configuration upon the introduction of defects, an efficient structural enumeration is needed to be performed, which is followed by DFT calculations for each possible distinct configuration. Hence, it eventually leads to the lowest-energy doped structure. This procedure is often computationally very expensive, especially in the dilute defect concentration.

Solid electrolyte/electrode interface is another important example of noncrystalline first-principles modeling, which is highly challenging at the current stage. To obtain realistic results, the required supercell needs to be sufficiently large, especially the length along the normal of the interface that aims to minimize the artificial interactions between periodic images. In principle, the AIMD method is suitable for interfacial diffusion studies at finite temperatures. Nevertheless, the required long simulation time scales as well as the length scales indicates that such simulations are almost computationally infeasible.

For cathode materials, it should be noted that this chapter is mainly devoted to intercalation electrode materials, wherein energy storage is possible through a reversible intercalation in a stable host framework. Although anionic redox has been recently discovered and confirmed, the capacity of this intercalation type electrode material is mainly limited to one electron per transition metal. Conversion and alloy reactions in electrode material are thus proposed to achieve higher capacities. More oxidation states of TM are feasible in conversion and alloy-type electrodes, which allows more electrons to reversibly change during electrochemical cycling. Similar to intercalation reactions, the thermodynamics of conversion/alloy reactions can be investigated from first-principles calculations by computing the total energy of all the involved compounds. Additional complexity does exist since conversion/alloy reactions produce multiple phases or domains with different chemical composition, where interface and nanosize effects (i.e., requiring noncrystalline modeling) often dominate the thermodynamic properties of materials.

The solutions to the aforementioned limitations and challenges are multifold. First, computational power will continue to grow rapidly, which can allow for longer simulations and larger system sizes and chemical complexity. Second, multi-scale approaches that aim at bridging the multiple time and length scales are expected to further improve in the future. That enables us to probe even more realistic models with higher computational accuracy. Third, the continued development of automation software tools will also facilitate the studies of battery materials and significantly minimize the researchers' time to monitor the simulations. Fourth, building open-access materials science databases from standardized first-principles high-throughput calculations can also facilitate the first-principles modeling. More importantly, with the rapid increase of the computation data, machine learning models can be trained with a higher prediction power, enabling researchers to discover hidden design rules for battery materials.

References

Anisimov VI, Zaanen J, Andersen OK (1991) Band theory and Mott insulators: Hubbard U instead of Stoner I. Phys Rev B 44:943–954. https://doi.org/10.1103/PhysRevB.44.943

Anisimov VI, Solovyev IV, Korotin MA et al (1993) Density-functional theory and NiO photoemission spectra. Phys Rev B 48:16929–16934. https://doi.org/10.1103/PhysRevB.48.16929

Aydinol MK, Kohan AF, Ceder G et al (1997) Ab initio study of lithium intercalation in metal oxides and metal dichalcogenides. Phys Rev B 56:1354–1365

Bron P, Johansson S, Zick K et al (2013) Li10SnP2S12 – an affordable lithium superionic conductor. J Am Chem Soc 135:15694–15697. https://doi.org/10.1021/ja407393y

Bruno M (2013) The reconstruction of dipolar surfaces: a preliminary step for adsorption modeling. Cryst Res Technol 48:811–818. https://doi.org/10.1002/crat.201200707

Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. Phys Rev Lett 55:2471–2474. https://doi.org/10.1103/PhysRevLett.55.2471

Chevrier VL, Ong SP, Armiento R et al (2010) Hybrid density functional calculations of redox potentials and formation energies of transition metal compounds. Phys Rev B 82:75122. https://doi.org/10.1103/PhysRevB.82.075122

Cococcioni M, de Gironcoli S (2005) Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. Phys Rev B 71:35105. https://doi.org/10.1103/PhysRevB.71.035105

Deng Z, Mo Y, Ong SP (2016) Computational studies of solid-state alkali conduction in rechargeable alkali-ion batteries. NPG Asia Mater 8:e254. https://doi.org/10.1038/am.2016.7

Grimaud A, Hong W, Grimaud A, Hong WT, Shao-Horn Y, Tarascon J-M (2016) Anionic redox processes for electrochemical devices. Nat Mater 15:121–126. https://doi.org/10.1038/nmat4551

Haruyama J, Sodeyama K, Han L et al (2014) Space-charge layer effect at interface between oxide cathode and sulfide electrolyte in all-solid-state lithium-ion battery. Chem Mater 26:4248–4255. https://doi.org/10.1021/cm5016959

Heyd J, Scuseria GE (2004) Efficient hybrid density functional calculations in solids: assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional. J Chem Phys 121:1187. https://doi.org/10.1063/1.1760074

Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened Coulomb potential. J Chem Phys 118:8207. https://doi.org/10.1063/1.1564060

Heyd J, Scuseria GE, Ernzerhof M (2006) Erratum: "Hybrid functionals based on a screened Coulomb potential" [J. Chem. Phys. 118, 8207 (2003)]. J Chem Phys 124:219906. https://doi.org/10.1063/1.2204597

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864. https://doi.org/10.1103/PhysRevB.7.1912

Hutter J (2012) Car-Parrinello molecular dynamics. Wiley Interdiscip Rev Comput Mol Sci 2:604–612. https://doi.org/10.1002/wcms.90

Jain A, Ong SP, Hautier G et al (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:11002. https://doi.org/10.1063/1.4812323

Jonsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Classical and quantum dynamics in condensed phase simulations. World Scientific, Singapore, pp 385–404

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133

Langreth DC, Mehl MJ (1983) Beyond the local-density approximation in calculations of ground-state electronic properties. Phys Rev B 28:1809–1834. https://doi.org/10.1103/PhysRevB.28.1809

Leung K, Budzien J (2010) Ab initio molecular dynamics simulations of the initial stages of solid-electrolyte interphase formation on lithium ion battery graphitic anodes. Phys Chem Chem Phys 12:6583–6586. https://doi.org/10.1039/b925853a

Lewis NS (2007) Powering the planet. MRS Bull 32:808–820. https://doi.org/10.1557/mrs2007.168

Marx D, Parrinello M (1996) Ab initio path integral molecular dynamics: basic ideas. J Chem Phys 104:4077. https://doi.org/10.1063/1.471221

Meng YS, Arroyo-de Dompablo ME (2009) First principles computational materials design for energy storage materials in lithium ion batteries. Energy Environ Sci 2:589. https://doi.org/10.1039/b901825e

Morgan BJ, Madden PA (2014) Relationships between atomic diffusion mechanisms and ensemble transport coefficients in crystalline polymorphs. Phys Rev Lett 112:145901. https://doi.org/10.1103/PhysRevLett.112.145901

Murch G (1982) The haven ratio in fast ionic conductors. Solid State Ionics 7:177–198. https://doi.org/10.1016/0167-2738(82)90050-9

Okubo M, Yamada A (2017) Molecular orbital principles of oxygen-redox battery electrodes. ACS Appl Mater Interfaces 9:36463–36472. https://doi.org/10.1021/acsami.7b09835

Ong SP, Wang L, Kang B, Ceder G (2008) Li-Fe-P-O_2 phase diagram from first principles calculations. Chem Mater 20:1798–1807. https://doi.org/10.1021/cm702327g

Ong SP, Chevrier VL, Hautier G et al (2011) Voltage, stability and diffusion barrier differences between sodium-ion and lithium-ion intercalation materials. Energy Environ Sci 4:3680. https://doi.org/10.1039/c1ee01782a

Richards WD, Miara LJ, Wang Y et al (2016) Interface stability in solid-state batteries. Chem Mater 28:266–273. https://doi.org/10.1021/acs.chemmater.5b04082

Roy P, Srivastava SK (2015) Nanostructured anode materials for lithium ion batteries. J Mater Chem A 3:2454–2484. https://doi.org/10.1039/C4TA04980B

Shin D, Saal J (2018) Computational materials system design. Springer, Cham

Sun W, Dacek ST, Ong SP et al (2016) The thermodynamic scale of inorganic crystalline metastability. Sci Adv 2:e1600225–e1600225. https://doi.org/10.1126/sciadv.1600225

Tang H, Deng Z, Lin Z et al (2018) Probing solid–solid interfacial reactions in all-solid-state sodium-ion batteries with first-principles calculations. Chem Mater 30:163–173. https://doi.org/10.1021/acs.chemmater.7b04096

Tarascon J-M, Armand M (2001) Issues and challenges facing rechargeable lithium batteries. Nature 414:359–367. https://doi.org/10.1038/35104644

Tasker PW (1979) The stability of ionic crystal surfaces. J Phys C Solid State Phys 12:4977–4984. https://doi.org/10.1088/0022-3719/12/22/036

Urban A, Seo D-H, Ceder G (2016) Computational understanding of Li-ion batteries. npj Comput Mater 2:16002. https://doi.org/10.1038/npjcompumats.2016.2

Wang L, Zhou F, Meng YS, Ceder G (2007) First-principles study of surface properties of Li Fe P O$_4$: surface energy, structure, Wulff shape, and surface redox potential. Phys Rev B 76:165435. https://doi.org/10.1103/PhysRevB.76.165435

Wolverton C, Zunger A (1998) First-principles prediction of vacancy order-disorder and intercalation battery voltages in Li$_x$CoO$_2$. Phys Rev Lett 81:606–609. https://doi.org/10.1103/PhysRevLett.81.606

Zhou F, Cococcioni M, Marianetti CA et al (2004) First-principles prediction of redox potentials in transition-metal compounds with LDA + U. Phys Rev B 70:235121. https://doi.org/10.1103/PhysRevB.70.235121

Zhu Y, He X, Mo Y (2015a) Origin of outstanding stability in the lithium solid electrolyte materials: insights from thermodynamic analyses based on first-principles calculations. ACS Appl Mater Interfaces 7:23685–23693. https://doi.org/10.1021/acsami.5b07517

Zhu Z, Chu I-H, Deng Z, Ong SP (2015b) Role of Na$^+$ interstitials and dopants in enhancing the Na$^+$ conductivity of the cubic Na$_3$PS$_4$ superionic conductor. Chem Mater 27:8318–8325. https://doi.org/10.1021/acs.chemmater.5b03656

Transport in Frustrated and Disordered Solid Electrolytes

51

Boris Kozinsky

Contents

Abstract

Materials design for next-generation solid-state Li-ion batteries require atomistic-level understanding of ionic transport mechanisms. It is not well understood what the required ingredients and design principles are for enabling high ionic conductivity of inorganic materials at room temperature. Strong ion-ion interaction, geometric frustration, disorder, and collective motion seem to be emerging as common themes in recent investigations of super-ionic materials. This chapter presents an overview of the essential features of these effects, their relationships, and implications on ionic transport. We illustrate applications of atomistic computational tools in combination with techniques of statistical

---

B. Kozinsky (✉)
John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
e-mail: bkoz@seas.harvard.edu

mechanics, group theory, and computational geometry for analyzing dynamics and transport mechanisms in the context of understanding and optimization of leading oxide Li-ion conductors.

1 Introduction

Ability of complex oxide materials to conduct ions lies at the heart of key modern technologies with high technological impact. Solid-state batteries, chemical sensors, solid oxide fuel cells, gas separation membranes, and memristors rely on ionic transport for their function. In particular, safe high-energy batteries for automotive and stationary energy storage can unlock a market of several hundred billion dollars, making it arguably the largest technology impact enabled by new functional materials. Today's battery electrolytes are based on liquid organic solvents that are combustible and prone to decompose and passivate the interfaces with electrodes, causing performance degradation. It would be a major technological step forward to replace the liquid electrolyte in a battery with a solid ion conductor that is chemically and mechanically stable against the Li metal anode, whose high gravimetric capacity can double the energy density of conventional cells with graphite anodes. Over the past two decades, significant progress has been achieved in optimizing solid-state Li-ion conductors. However, only a few classes of fast Li-ion conductors are known, and they have been discovered and optimized primarily by haphazard trial and error. Highest ionic conductivities are found in sulfides, particularly crystalline $Li_{10}GeP_2S_{12}$ (12 mS/cm at room temperature) and glassy $xLi_2S\text{-}yP_2S_5$ (3 mS/cm). Oxides are more chemically and electrochemically stable but have lower conductivity, for instance garnets $Li_{6.55}L_3Zr_2Ga_{0.15}O_{12}$ with conductivity of 1.6 mS/cm and NASICON $Li_{1.3}Al_{0.3}Ti_{1.7}(PO_4)_3$ with 1 mS/cm (Meesala et al. 2017). The great technological importance but limited number of fast ion conductors has inspired recent efforts to identify new families of materials more suitable for application in solid-state batteries, but this is inherently challenging given the nonlinear dependence of conductivity on atomic structure and composition. Accelerated materials development requires identification of general design rules and mechanistic principles enabling fast ion transport. Various qualitative descriptions have been proposed over the last decade, such as sufficient volume of conduction pathways and high anion polarizability, but these have limited predictive power. In order to make rational progress in the search and optimization of fast ion conductors, it is important to develop general physics-based models and validate their prediction using statistically significant data sets. Formulating a general microscopic theory of super-ionic transport in solids is difficult because ion-ion interactions and correlations are especially strong when carrier concentration is high. Even qualitative mechanistic principles and design rules that govern "superionic" behavior are not yet established, and generalization from only a few known cases is difficult. This is in stark contrast with the success of band theory and Fermi liquid model in perturbatively describing electron interactions and transport in weakly correlated semiconductors and metals. There is evidence that solid-state

ionic transport in the best conductors exhibits collective features reminiscent of correlated fluids, where ionic transport is typically much faster than in regular solids. Despite the long history of the field, complex effects of many-body interactions on transport in atomic fluids are generally not well captured by analytical models, making explicit atomistic simulations particularly important. In this chapter, we illustrate how combinations of techniques of statistical mechanics, group theory, and molecular dynamics can help develop systematic understanding of ionic transport in strongly interacting and dynamic systems. We give an overview of practical computational approaches useful for analyzing transport mechanisms, studying geometric frustration, and estimating correlation in crystalline and disordered materials and provide examples of their application to understand solid oxide Li-ion electrolytes.

2 Simulation Methods

Fast ion conduction arises due to a delicate balance between the interatomic forces acting on the mobile ions. On one hand, the mobile carrier sublattice needs to be sufficiently coupled to the host lattice in order to ensure overall thermodynamic and mechanical stability of the material. On the other hand, the interaction of the mobile ions with their host cannot be too strong so as to allow for low enough activation barriers in order for ionic current to flow at device operating temperatures. The strong ionic interactions of Li ions with the host lattice determine the topology of sites that can be occupied, typically triangles, tetrahedra, or octahedra formed by the anions (e.g., O, S, N, or F). In the classical picture of solid-state diffusion, an ion spends the vast majority of the time rattling in its site, occasionally escaping and hopping to a neighboring vacant site through a polyhedral face, as illustrated on Fig. 1a. The probability of this hopping event is described by transition state theory as activated by thermal fluctuations of the host lattice, giving the Arrhenius equation for temperature-dependent diffusivity:

$$D(T) = D_0 \exp\left(-\Delta E_a \big/ kT\right) \tag{1}$$

Activation energy ΔE_a corresponds to the height of the barrier in the energy landscape separating the two local minima sites. In the case of a single ion moving in a static periodic potential of an empty lattice, where the transport mechanism is known, this barrier can be computed readily using the nudged elastic band (NEB) method that uses a total energy model, such as density functional theory (DFT) to find the minimum energy path passing through the unstable saddle point (Jónsson et al. 1998). This approach, however, is only simple when Li ions are dilute interstitial defects in an ideal crystal lattice and do not interact while hopping independently of each other. In the best ionic conductors, Li-ion concentrations are high, which means that Li ions are in close contact and Li-Li interactions are not negligible. In such strongly interacting scenarios, forces confining Li ions in
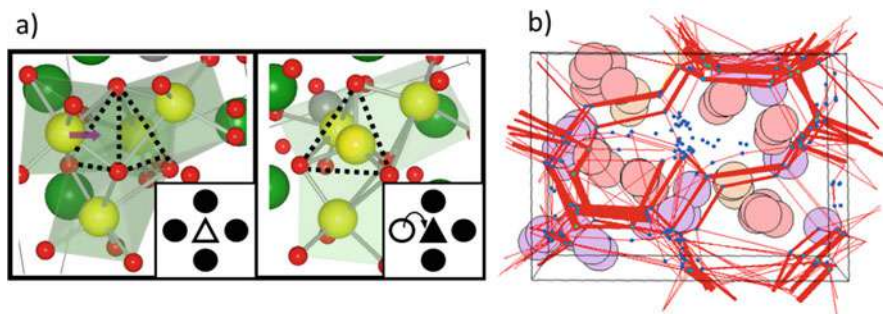
Fig. 1 (**a**) Ionic transport mechanism in Li garnets, showing Li migration from an oxygen-coordinated octahedral site to a neighboring tetrahedral site through a face. Geometric hull analysis determines that the occupancy changed and flags this as a hopping event (Courtesy of Sneha Akhade). (**b**) Voronoi decomposition of a crystal automatically identifies polyhedral sites that are used by the geometric transport analysis algorithm. The thickness of the graph edges connecting the Voronoi sites corresponds to the number of hops detected in the MD simulation analysis. (Courtesy of Leonid Kahle)

their sites are counteracted by electrostatic repulsion from nearby Li ions, and Li ions can displace each other. This leads to a self-blocking interaction where at most one Li ion can occupy a polyhedral site. As a result, there is a significant amount of correlation between mobile ions, meaning that the energy landscape of the Li sublattice (occupancy energies and transition barriers) is complex and depends on local configuration. Configuration dependence of the energy barrier can be introduced using the method of cluster expansions (Van der Ven et al. 2001), but also in this case, the transport mechanisms must be known a priori in order to compute barriers in all configurations relevant for macroscopic transport. If it is possible to obtain the energy barriers, the next step is to construct a kinetic Monte Carlo (KMC) model based on the complete catalog of all possible transitions and their probabilities.

An alternative practical approach to study complex ion conductors is molecular dynamics (MD), where interatomic forces are used to simulate the actual lattice time evolution according to Newton's equations of motion, without taking as input any information about transport mechanisms (Alder and Wainwright 1962). MD has emerged as a powerful tool for examining and optimizing solid ionic conductors, providing insights into transport mechanisms that are not possible to obtain experimentally. MD also relaxes the assumption of KMC models of time-independent transition probabilities, which allows realistic treatment of ion transport in dynamic environment such as liquids and polymers. At the same time, MD is significantly more computationally expensive and statistically inefficient than stochastic methods, since transport is dominated by rare hopping events, and thermal vibrations consume the vast majority of computation time. Typical simulations using ab initio molecular dynamics, where forces are obtained from DFT, are limited to about 200 atoms and 100 picoseconds taking roughly 2 weeks of wall time, using

the typical time step of 1–2 femtoseconds. The standard approach is to perform simulations at significantly elevated temperatures usually above 700 K, in order to rapidly sample ionic motion, and then extrapolate diffusivity to room temperature, assuming that the Arrhenius dependence in Eq. (1) holds. This approach requires caution when transport mechanisms are temperature dependent and if phase transitions occur at lower temperatures. In classical MD, forces are determined from analytical force fields, such as Buckingham and Morse pair potentials, sometimes with core-shell polarization and angle-dependent terms included. Classical MD has the advantage of being several orders of magnitude faster than ab initio dynamics. The limitation is that every material requires careful selection and validation of the functional form and parameters of the force field. Special care is needed especially for ionic transport simulations, since diffusivity is determined by energies of low-symmetry configurations corresponding to transition states, while available force fields are typically parameterized to represent equilibrium properties.

3   Transport Mechanism Analysis and Coarse-Graining

A key advantage of using molecular dynamics to study ionic transport in ceramic electrolytes is the ability to directly characterize atomistic transport mechanisms, which consist of very fast and rare hopping events determining the long-range macroscopic transport character of the material. In order to extract information about the transport mechanisms, it is necessary to separate vibrational motion from hopping events. The goal is therefore to map the dynamics of Li onto a discrete lattice of sites in order to analyze the hopping sequences and statistics. One option for extracting hop events from the MD simulation is to keep a record for every Li of its nearest-neighbor atoms, based on a carefully chosen distance threshold, and to detect significant changes in the neighbor lists. This method is easy to implement but has the disadvantage that trajectory discretization can depend on the distance threshold and neighbor list, since different types of sites have different anion coordination numbers. A more robust approach is to first determine the topology of the host lattice and partition it into nonoverlapping polyhedra formed by immobile anions. This can be obtained from crystallographic information and/or from Voronoi decomposition using anions as the vertices. Every Li ion is then assigned to an available polyhedral site in each MD snapshot. A simple way is to use a 3D convex hull construction (e.g., using the Qhull code) taking as input the positions of the Li ion of interest and all anion vertices of the candidate site. If the Li ion is inside a particular site, then the convex will contain only the anions, and if Li is outside, then its position will be part of the convex hull. This method is reliable, since there are no free parameters, with the assumption that the host lattice does not rearrange. For example, in garnet ionic conductors, this analysis clarified that transport occurs only by tetrahedral to octahedral to tetrahedral site face-crossing hops, contrary to earlier conjectures of direct octahedral-octahedral edge-crossing hops (Fig. 1a). Using MD as a starting point for mechanistic investigations and coarse-graining dynamics simulations is especially useful when transport mechanisms in a complex

new material are not known a priori. In this way, the transport analysis can be automated and used for high-throughput investigation of many diverse structural variations. In practice, this automatic procedure starts from high-temperature ab initio or classical MD to quickly sample important hopping scenarios. Once the MD trajectory is discretized, the site-to-site transition sequences and timings can be extracted to obtain the complete transport mechanism sampled during the MD simulation. Frequency statistics directly from MD are often insufficient to accurately estimate hopping probabilities, especially for the more rare events. In such cases, hopping rates for each event can be estimated using transition state theory, which relies on barrier height estimation using the NEB method. Finally, the catalog of transition probabilities can then be used to define a coarse-grained dynamics models using KMC, mentioned above, which simulates hopping as a stochastic Markov process while keeping track of real system time. However, this last step may not be straightforward in some cases we consider below.

The resulting energy landscape determined by this analysis workflow contains important information about the nature of ionic transport. For example, Fig. 2 shows an example of an MD trajectory of a garnet Li-ion conductor that was first discretized using the geometric method described above and combined with NEB calculation of barriers of the pathways connecting the extracted hop endpoints. It
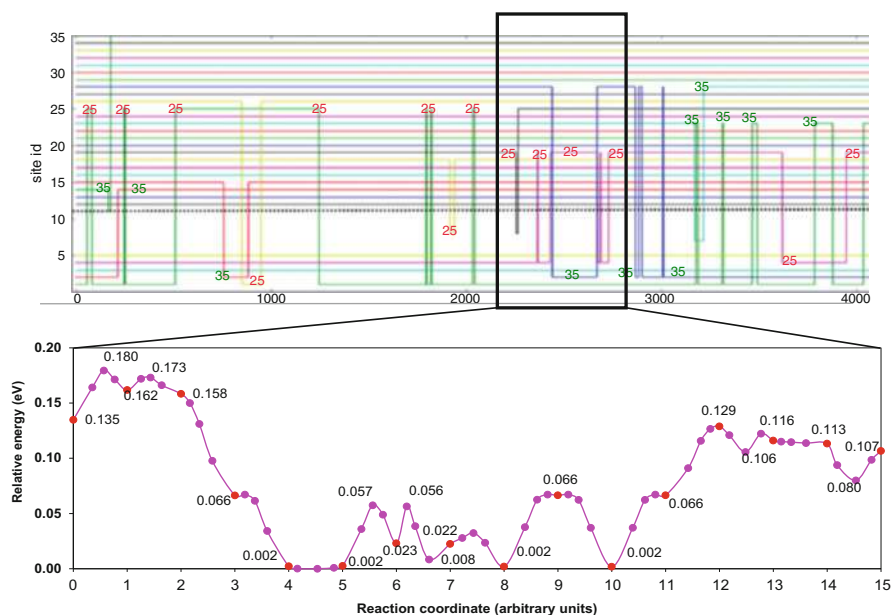
Fig. 2 Result of geometric discretization of an ab initio molecular dynamics simulation of a Li-garnet structure performed at elevated temperature (800 K) in order to sample a significant number of hops. Endpoint configurations of detected transition paths are used as inputs into NEB computations of migration barriers

is noteworthy that the resulting energy landscape is qualitatively different from the classical picture of a single-ion hopping in a periodic potential landscape of the host lattice. Instead, Li configurations explore a complex energy landscape in a collective way, and some events do not even exhibit barriers. Transport proceeds through a sequence of moves involving multiple Li ions moving downhill or uphill in energy, and constructing a kinetic coarse-grained model may require an alternative approach to the conventional individual ion hop picture.

4   Correlation and Collective Motion

The outputs of both KMC and MD simulations are trajectories, or sequences of atomic positions and corresponding energies and forces on each atom. Information on positions (or velocities) can be used to calculate transport properties, hopping mechanisms, and correlation functions. Ionic conductivity can be obtained using the Green-Kubo formalism, starting from

$$\sigma = \frac{N^2 q^2}{6kTV} \lim_{t \to \infty} \frac{d}{dt} \langle |\mathbf{R}(t) - \mathbf{R}(0)| \rangle = \frac{q^2}{kT} \frac{N}{V} D_\sigma \qquad (2)$$

where N is the number of diffusion Li ions, $q = 1$ is the charge, V is unit cell volume, and $\mathbf{R}(t)$ is the vector connecting the center of mass of all Li ions with that of the host material at time t. The reference to the host material center of mass is especially important in simulations of multicomponent systems where the host exhibits diffusive or fluctuating dynamics (Wheeler and Newman 2004). The second equality in (2) is the Nernst-Einstein equation defining for convenience a "conductivity" diffusion coefficient D_σ which, unlike conductivity itself, is not directly measured. A useful directly measurable quantity is the tracer diffusion coefficient which is given by

$$D^* = \lim_{t \to \infty} \frac{1}{6} \frac{d}{dt} \frac{1}{N} \sum_{i=1}^{N} \left\langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \right\rangle \qquad (3)$$

where $\mathbf{r}_i$ are the positions of the Li ions and the quantity under the derivative is called the mean square displacement (MSD). It can be shown that $D_\sigma = D^*$ when all cross-correlations between displacements of different Li ions vanish. This is a good commonly used approximation to obtain ionic conductivity from MD simulations of dilute liquid electrolytes. However, in solid electrolytes, this is not the case due to significant correlations, as mentioned earlier. The importance of correlations can be estimated from the Haven ratio which is defined as $H = D^*/D_\sigma$. Typical Haven ratios for fast Li-ion conductors range between 0.25 and 0.7. The inverse of the Haven ratio has been interpreted as the average number of ions that interacts to move collectively (Doliwa and Heuer 2000). Despite the non-negligible effect of correlations, in practice MSD is the most frequently used

quantity for estimating ionic conductivity from MD since it converges much faster with simulation time than the fully correlated conductivity obtained from the center of mass displacement. The reason is that in the first case, the estimate of the MSD is based on N-independent Li displacements, while in the second case, only the single displacement value of the center of mass is used to estimate the conductivity. Therefore, the statistical variance of the estimate of D^* is smaller than that of the conductivity by a factor of order N^2.

Haven ratio can be directly measured (e.g., from dc conductivity and NMR tracer diffusivity) but only provides a time-averaged indication of ionic correlations. To understand the microscopic nature of correlations, a detailed examination of atomic trajectories is necessary, as discussed in the previous section. In some garnet crystals, Li ions exhibit pronounced collective, or concerted, motion (see Fig. 3a), where multiple ions are displaced at the same time (Meier et al. 2014) rather than individual independent hops. In Li_3PO_4 this type of concerted motion was found to be responsible for low activation energies of 0.2 eV, while direct hopping has a migration barrier of 0.6 eV (Du and Holzwarth 2007). Concerted motion and hopping time correlations were also observed in $Li_xLa_{2/3-x/3}TiO_3$ (Catti 2011) as well as sulfides and NASICON ion conductors using ab initio molecular dynamics (He et al. 2017). In each case, it was noted that collective migrations had lower-energy barriers compared to mechanisms involving single-ion hops.

To understand the appearance of collective modes, consider a string of ions moving in a single-particle potential energy surface with amplitude ΔE_s, illustrated in Fig. 3b. If the ion-ion interaction is strong enough and constrains the ions to be spaced apart in a way that is not commensurate with the energy landscape maxima and minima, then the group as a whole experiences a smoother energy landscape, due to a "window-averaging" effect of the original energy landscape. Such collective modes, if they can form and move, will statistically dominate transport relative to
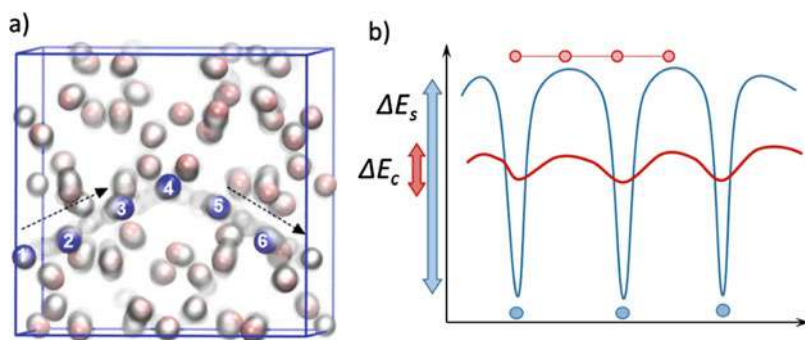
Fig. 3 (**a**) String-like collective motion involving multiple Li ions in tetragonal Li-garnet oxides (Meier et al. 2014). (**b**) Effective flattened energy landscape (red) experienced by a string of ions moving in an external lattice potential (blue). Dilute ions or configurations commensurate with the external potential minima will experience much higher activation energies in their motion, making collective motions more likely

the individual ion hops that are much less frequent due to their higher barriers. This situation can only happen when nearest-neighbor Li interactions are significant, i.e., when the fraction of occupied Li sites is high. This simple picture is an example of the phenomenon of frustration, as discussed in detail below, where strong short-range interactions prevent ions from settling into local minima, ensuring collective mobility. It is interesting that collective "string-like" highly correlated motion has also been noted in other classes of materials, particularly in regimes close to melting and glass transitions (Gebremichael et al. 2004). It is not clear whether collective motion is a necessary ingredient for high ionic mobility or whether it is simply the lowest-energy excitation in highly correlated situations. Their existence points to the possibility that super-ionic conductors are close to a sublattice phase transition where the Li sublattice exhibits liquid-like behavior, motivating the discussion of phase transitions in the next section.

A practical way to quantify position and time correlations of ionic jumps and to detect such collective modes is the distinct part of the van Hove correlation function, defined as

$$G_d\left(r, t\right) = \frac{1}{N} \sum_{i \neq j} \left\langle \delta\left(r - \left|\mathbf{r}_i(0) - \mathbf{r}_j(t)\right|\right) \right\rangle \tag{4}$$

It describes the conditional probability of finding a pair of distinct atoms such that they are separated by distance of r after a time delay of t. $G_d(r, t = 0)$ is just the usual pair distribution function. $G_d(r = 0, t)$ indicates probability of one ion being replaced by another ion in time t, which can be interpreted as a collective simultaneous event when t is comparable to the inter-site transition time scale. In addition to being useful for gaining intuition about ionic transport correlation, the Van Hove correlation function is measurable experimentally. Its Fourier transform is the intermediate scattering function that can be measured by neutron spin echo spectroscopy. This quantity provides a valuable opportunity for direct validation of MD simulations against experimental data.

5   Geometric Sublattice Frustration

The overall ability of the Li sublattice configuration to rearrange with low transition barriers is the necessary ingredient of high ionic conductivity. Since ionic transport is a statistical thermally activated phenomenon, the probability of such rearrangements is proportional to the number of available configurations. It is clear, therefore, that in order to maximize conductivity, we need to maximize the sublattice entropy. As with any system in thermal equilibrium at temperature T, the ionic conductor will minimize the free energy and therefore tend to find low-energy configurations of the Li sublattice. If the ground state is well ordered, the only available mechanism for transport is through creation of mobile defects. In regular periodic crystals, the activation energy required for ionic hopping $E_a = E_{\text{def}} + E_{\text{hop}}$ is a sum of the energy

required to first create a mobile defect and the migration barrier height. However, thermal energy scale at room temperature is rather low, with kT only on the order of 25 meV, making it difficult to overcome strong ionic interactions and disturb a stable ground state configuration in a crystal. This is the reason that in crystalline ionic solids, diffusion is much slower than in liquids. The goal in designing a fast ionic conductor, therefore, is to ensure that the ground state is intrinsically degenerate or nearly degenerate, so that it is able to rearrange with minimal energy cost. In materials science terms, this qualitatively corresponds to engineering a high intrinsic concentration of mobile carriers by varying the composition and structure. In mathematical terms, this can be described as symmetry breaking and maximizing the amount of geometric frustration in the Li sublattice, preventing the system from settling into a low-entropy arrangement.

We generally define *frustration* as competition between local interactions (Li-Li and Li-anion), resulting in near degeneracy and disorder in the ground state configuration of the Li sublattice. Locally, this situation comes about when Li occupancy of a lattice site cannot be decided based on minimizing total interaction energy with its neighbor sites. Globally, frustration manifests itself when there is not a single way to minimize energy by ordering mobile ions on their sublattice.

Intuitively, the role of frustration is to flatten the energy landscape of Li ions, preventing locking of mobile carriers in deep energy minima corresponding to stable ordered arrangements (Fig. 3b). Li ions in the frustrated sublattice are then forced to explore the large number of degenerate states in a narrow energy window. Effective frustration, which maximizes the sublattice entropy, results in a high concentration of charge carriers and high interstate transition rates, the two necessary ingredients for fast ion conduction.

Geometrically frustrated systems have a long history in solid-state physics, traditionally studied in the context of antiferromagnetic spin models on various lattices with short-range interactions. Due to the rich physics and emergence of unexpected phenomena, frustration in magnetism is an increasingly active research field in condensed matter theory (Balents 2010). Complexities arising from the presence of extensive ground state degeneracy lead to unusual thermodynamic and transport effects. These include suppression of finite temperature order-disorder freezing transitions, giving rise to disordered spin liquid ground states. Recently attention to frustrated lattice systems has increased due to the emergence of Majorana fermions, topological quantum excitations resembling relativistic particles (Kitaev 2006).

In order to illustrate the concept of frustration and its implication on charge transport, let us consider a simple model: a two-dimensional antiferromagnetic Ising model with nearest-neighbor interactions on a triangular lattice. As illustrated on Fig. 4a, once opposite spins occupy two of the three corners of a triangle, there is no unique choice for the third spin direction that minimizes the energy. To connect this peculiar behavior to charge transport, we can map the lattice spin variables to charge occupancy, so that spin up corresponds to occupied site and spin down to unoccupied, as shown on Fig. 4a. The antiferromagnetic spin interaction physically corresponds to electrostatic repulsion of charges on neighboring sites.
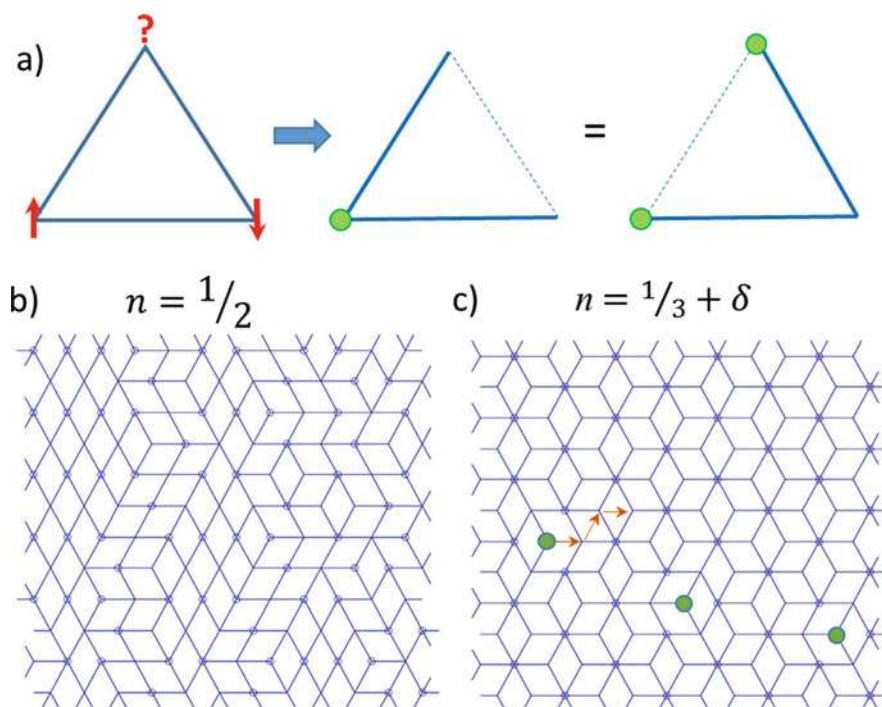
Fig. 4 (**a**) Frustrated configuration of spins on a triangle is mapped onto site occupancy in a lattice gas model. The two configurations have the same energy if only nearest-neighbor interactions are considered. (**b**) One of many possible ground state configurations at ½ occupancy of the triangular antiferromagnetic Ising model. (**c**) A well-ordered low-entropy ground state configuration exists for occupancy fractions 1/3. At slightly higher occupancy, entropy increases, since additional charges can freely move through the ordered background, forming a dilute gas

This construction is representative of a class of lattice gas models with on-site exclusive constraints. Frustrated lattice models are commonly discussed in the context of spin or electron charge degrees of freedom, but it is natural to extend this understanding to crystalline ionic conductors. In fact, the triangular lattice gas model describes Li-ion configurations in transition metal layered oxides $Li_x(Ni,Co,Mn)O_2$ $(0 < x < 1)$, which are by far the most commonly used cathode materials in Li-ion batteries today. In these materials, the crystal structure is made up of layers of oxygen-coordinated edge-sharing octahedral sites filled with transition metals, separated by layers of octahedral Li sites that form a triangular lattice, such that Li ions move in and out of the layers during battery cycling.

Relevant temperature scales in this lattice gas model are approximately given in terms of the nearest-neighbor interaction strength V. The simplest to understand is the high-temperature behavior $kT \gg V$, corresponding to a disordered gas phase in which charges can move via independent hops, only subject to the exclusive site occupancy constraint. At low temperature $kT \ll V$, the system settles into the

ground state configuration (solid phase). At intermediate temperatures, the system leaves the ground state manifold, and a population of mobile defects exist, e.g., all three triangle corners occupied or empty.

Due to local frustration, the ground state at half-filling ($n = 1/2$, i.e., zero average magnetization) is highly degenerate, allowing many configurations with the same lowest energy (illustrated in Fig. 4b). This model with only nearest-neighbor interactions can be solved exactly to show that there is actually no ordering phase transition at any finite temperature (Wannier 1950). In other words, the spin lattice never truly freezes and remains in a glassy state. We should note that this is true because the nearest-neighbor antiferromagnetic interaction is incompatible with the triangular lattice symmetry in the case of $n = 1/2$, but there exists a well-ordered ground state at $n = 1/3$, where a pronounced freezing transition occurs at finite temperature and consequently entropy and transport are strongly suppressed. This emphasizes that "local frustration" does not necessarily guarantee ground state degeneracy. Instead, *ground state degeneracy is a function of both carrier concentration and global lattice symmetry*, as elaborated further in the next section.

Remarkably, the frustrated ground state without any "defect" excitations has substantial ability to rearrange, due to local charge motions. Monte Carlo simulations of thermodynamics and transport indicate that this system appears to possess a sufficient amount of residual entropy in the $n = 1/2$ ground state to allow for finite conductivity, even in the presence of long-range interactions (Levitov and Kozinsky 1999). When long-range screened Coulomb interactions are included for $n = 1/2$, the degeneracy is broken, and an ordering phase transition appears but at much lower temperature than in the case of $n = 1/3$, indicating that frustration effectively suppresses freezing. In this regime, charge transport has collective character of a correlated fluid due to strong short-range correlations between charges which constrain individual charge migrations, preserving the degenerate ground state manifold of configurations. The change in conductivity behavior while cooling down from the disordered into the correlated fluid and solid phases can also be seen in Fig. 5. The low-temperature curves indicate that the conductivity vanishes more quickly near the densities of a simple fraction form ($n = 1/4, 1/3, 1/2$). This corresponds to freezing of the system into states commensurate with the lattice at these values of n. The qualitative features of the phase diagram and its influence on charge transport do not seem to be affected by including disorder (Novikov et al. 2005). So far the model only considers energy of configurations and does not include kinetic details of inter-site transition. Even when more detailed hopping mechanisms are included specifically to describe transport of Li in Li_xCoO_2-layered oxides using cluster expansions and kinetic Monte Carlo, the main effects of concentration dependence of conductivity are preserved (Van der Ven et al. 2001). This simple 2D model demonstrates the key principle that charge transport without defect excitations is possible in an ideal symmetric lattice, as long as frustration exists that establishes degeneracy and disorder at non-commensurate carrier concentrations. In a later section, we will see that this principle allows us to understand and design more complex ionic conductors.
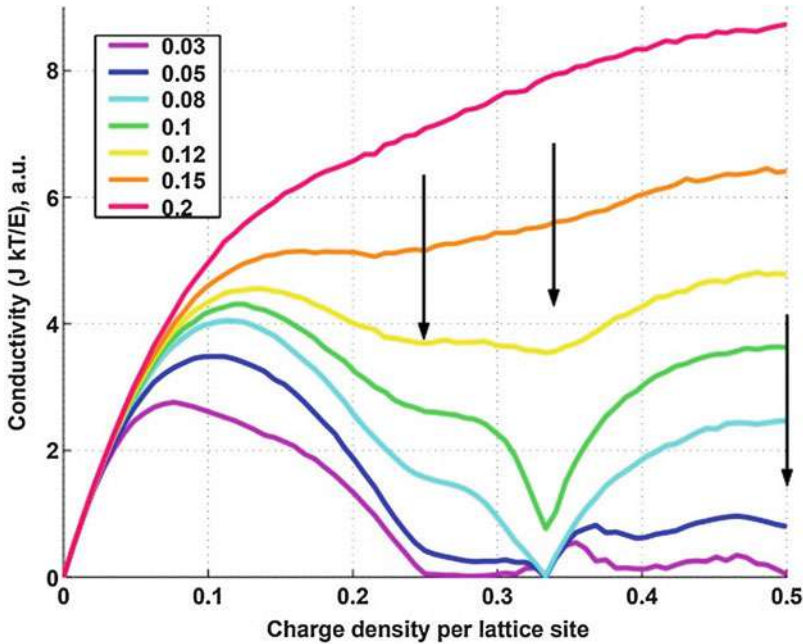
Fig. 5 The scaled conductivity in the triangular lattice gas model with long-range screened Coulomb interaction, as a function of charge occupancy fraction *n* for several temperatures. The temperature values are given in the units of V(a), the nearest-neighbor interaction. Arrows mark the features corresponding to the freezing phase transitions at n = 1/4, 1/3, 1/2. Dashed line corresponds to the high-temperature limit. The symmetric high-occupancy half of the concentration range is not shown. (Levitov and Kozinsky 1999)

5.1 Algebraic View of Sublattice Frustration

We now turn to the discussion of ways to identify and design materials with frustrated mobile sublattices. In mathematical terms, the appearance of a strong degeneracy of the ground state requires that there is no low-entropy highly symmetric configuration commensurate with the host lattice. At the same time, frustration arises due to strong short-range interaction, which requires high Li-ion concentration in the material. From this point of view, when optimizing ionic conductivity of a particular crystal, *Li concentration* is the key tuning parameter. In the light of the discussion in the previous section, our goal is to identify high-symmetry ground states of the sublattice configuration, into which Li ions can freeze. Ability to do this has wide applicability in general, especially for Li-containing crystals, since Li has a very weak X-ray scattering cross-section and Li site occupancies are very difficult to assign without resorting to neutron diffraction of single-crystal samples. The nature of ionic conduction is also very difficult to observe, only indirectly using solid-state NMR techniques. In the

context of fast ion conductors, knowledge of where ordered configurations appear in the composition space is critical to guide optimization away from these Li concentrations, thus maximizing frustrated disorder.

In order to systematically and rigorously identify ordered ground state configurations as a function of Li occupancy, group theoretic tools are particularly effective. We illustrate this approach in the case of Li-garnet family of ionic conducting crystals with room temperature ionic conductivity of up to 1.6 mS/cm. Combined with its electrochemical stability in contact with Li metal and high mechanical toughness, this makes it an attractive and widely investigated candidate for a solid electrolyte in a Li metal battery. The prototypical garnet structure is body-centered cubic (group Ia-3d No. 230) with the chemical formula $Li_xB_3C_2O_{12}$. This structure is remarkably robust in regard to changes in the B and C cation composition (B = La, Ca, Ba, Sr, Y, Pr, Nd, Sm-Lu, and C = Zr, Ta, Nb, Nd, Te, W) and is able to accommodate Li concentrations x ranging from 3 to 7 per formula unit. In the vast majority of crystalline oxides, Li ions occupy octahedral and/or tetrahedral sites. Garnets are no exception; there are two distinct sets of sites: octahedral (O, Wyckoff position 48 g) and tetrahedral (T, at position 24d). The numbers indicate multiplicities, i.e., 48 O sites are equivalent and can thus be either all filled or all empty in the prototype group of the host lattice. Each T site shares a face with four O sites, while each O site shares a face with two T sites, thus forming a percolating Li site network of pathways for ionic conduction.

The arrangement of Li ions among available crystallographic sites is the main degree of freedom that controls ionic transport. However, complete brute-force analysis of Li orderings is impossible: for instance, in the simple cubic cell representation of the $x = 7$ garnet, there are 72 Li sites, of which 56 are occupied, resulting in 10^{15} possible arrangements. In order to deal with the combinatorial space of configurations, it is useful to apply crystal group analysis to classify simply ordered Li arrangements for each composition x. The procedure starts with exhaustive crystallographic projections of the prototype high-symmetry group of the host lattice (Ia-3d No. 230 in case of garnets) onto lower-symmetry subgroups. This allows to split the high-symmetry Wyckoff positions (48 g and 24 d), corresponding to all filled or empty T and O sites, into sets of smaller multiplicities, so that they can be separately occupied. The resulting structures are the highest symmetry configurations, commensurate with the host lattice space group, where Li occupies a fraction of the sites. For instance, group No. 198 $P2_13$ is a cubic subgroup of the parent group, where the T site with the original prototype Wyckoff position 24d is split into two Wyckoff positions 12b corresponding to two subsets of 12 sites related by symmetry operations. Each subset can then be occupied independently. There are many ways of breaking the parent prototype symmetry, and going down to the triclinic group P1 corresponds to removing all symmetry and enumerating all possible configurations, which is not feasible. Therefore, it is practical to constrain the search to structures that have high enough symmetry to be likely ground states and can be represented using small unit cells in an ab initio simulation. Mathematically this corresponds to calculating the tree of maximal subgroups of the parent prototype group of the host lattice, limiting them to certain Bravais lattices

(e.g., only cubic or tetragonal) and a low k-index. The k-index of a subgroup is the multiplication factor relating the volume of the primitive cell of the subgroup with respect to the primitive cell of the original prototype structure. We can then analyze the splittings of Wyckoff positions in each subgroup and conjugacy class in order to identify symmetrically related sets of Li sites and their multiplicities. A simple 2D example of crystal symmetry breaking is illustrated on Fig. 6a. This procedure has the advantage of rigorously evaluating all possibilities starting with the highest symmetry and usually results in only a few hundred high-symmetry configurations with various Li concentrations, a significant reduction compared to the original space of possibilities.

To implement the subgroup projection in practice, it is convenient to use symmetry tools provided by the Bilbao Crystallographic Server (Aroyo 2006). For the case of garnets, a partial tree of maximal subgroup projections down to orthorhombic symmetry and k-index 2 is illustrated on Fig. 6b, which lists symmetry equivalent sets of sites with their Wyckoff multiplicities. Crystal structures are constructed of the simple cubic cell where each independent set of Wyckoff positions is either included or excluded in the structure. The total energy of each configuration is computed using density functional theory. For the garnet family, the result is that there are unique low-energy configurations at certain integer values of Li composition (x = 3, 4, 6, 7). High-symmetry configurations also exist at other integer compositions but are very close in energy (less than 3 meV per atom), resulting in
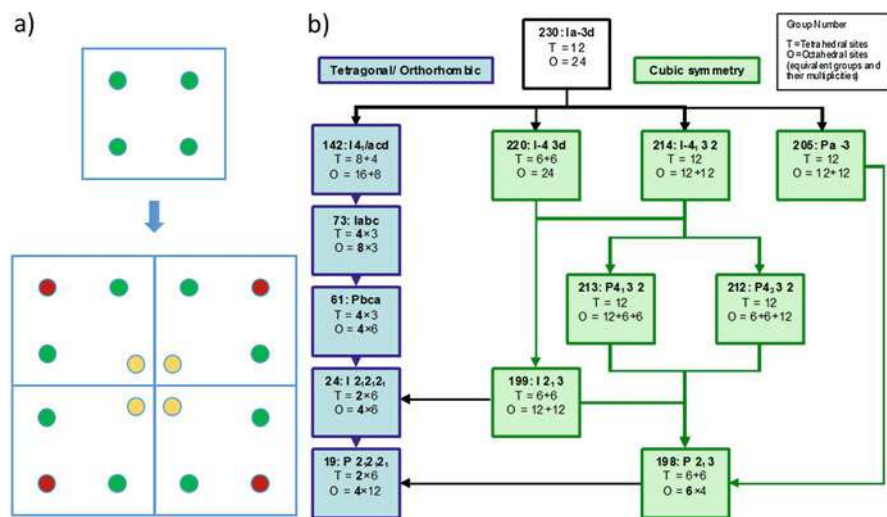
Fig. 6 (**a**) Illustration of symmetry breaking preserving square symmetry in a larger supercell (k-index 4). Different colors represent sets of equivalent positions that can be removed or displaced according to the subgroup Wyckoff positions. (**b**) Partial group-subgroup relationship tree that exemplifies a typical set of paths from the parent group No 230 to lower-symmetry subgroups. T and O represent tetrahedral and octahedral sets of sites, respectively, together with their multiplicities
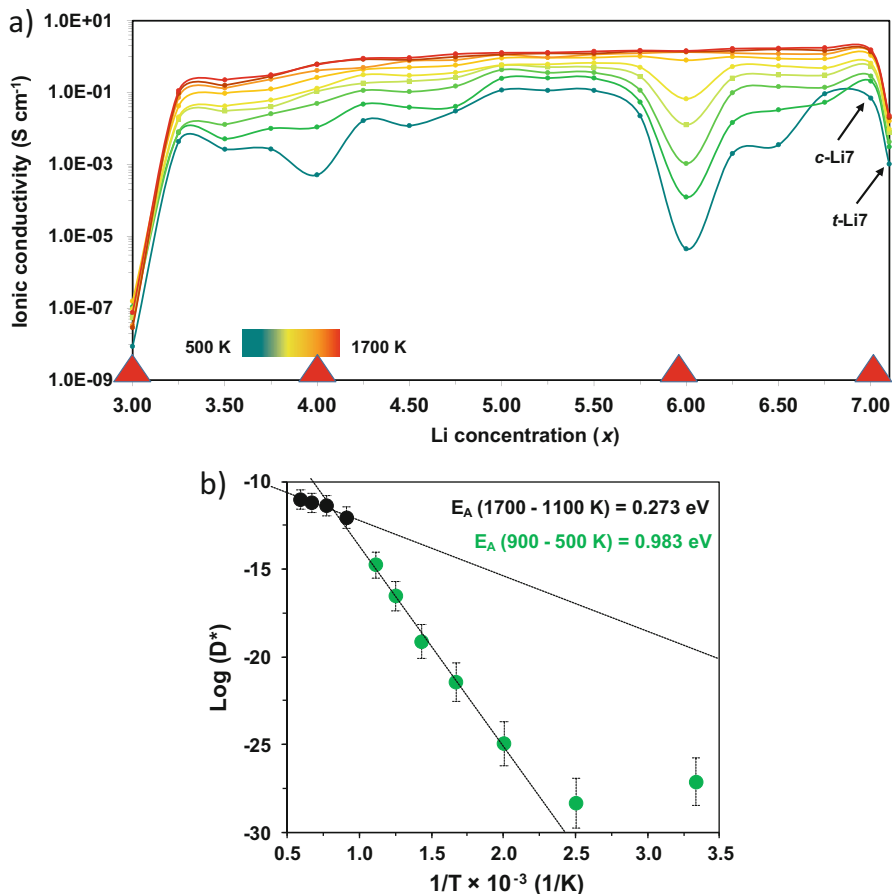
Fig. 7 (**a**) Ionic conductivity as a function of Li content x in the garnet crystal, from MD simulations. Red triangles indicate compositions at which ordered ground states were identified by the subgroup projection method. (**b**) Logarithm of the tracer diffusion coefficient versus inverse temperature for the $x = 6$ structure, illustrating the order-disorder phase transition at about 900 K. Black circles represent high-temperature values, and green triangles represent low-temperature values

disordered structures at room temperature. At non-integer compositions, no low-entropy ground states are found at all by the symmetry analysis, at least within the constraint of a cubic structure with fewer than 200 atoms in the unit cell. Since Li-Li interactions are weak beyond the length scale of the already large unit cell (13 Å), this constraint is physically justified. Compositions where ground states have high degeneracy and frustration due to symmetry incompatibility are expected to have the highest ionic conductivity. Indeed, as illustrated in Fig. 7a, conductivity is noticeably suppressed near precisely the compositions where well-ordered ground state Li configurations exist, with non-negligible excitation energies. On the other

hand, much shallower energy landscapes and degeneracy result in better ionic mobility in disordered configurations that arise in situations of frustration, where order-disorder phase transitions are suppressed (Kozinsky et al. 2016), and much lower conductivity when order-disorder transitions appear at high temperature, e.g., at $x = 6$ (Fig. 7b).

6   Extrinsic Disorder: Amorphous Materials and Random Alloys

An alternative way to avoid ordered Li sublattice configurations is to break the translational symmetry of the host lattice. Extrinsic disorder can be implemented in several ways. By melting and rapidly cooling, some crystalline materials can be turned into glasses that remain in the state of quenched disorder indefinitely at room temperature. This process typically preserves the local coordination and bonding features of the crystal but destroys the long-range order. Examples of materials where ionic conductivity in crystalline form is significantly lower than in amorphous structure are Li_3PO_4 and $LiAlSiO_4$. Geometric frustration between the building blocks of the host lattice (not Li sublattice as discussed earlier) can result in a glassy disordered morphology of the material (e.g., water ice). Disorder can also be introduced by random site occupancy in regular crystals, preserving the long-range lattice structure. For instance, experiments and MD simulations show that random alloying in $Ba_{1-x}Ca_xF_2$ enhances F-ion conductivity by a factor of 10^5 (Duvel et al. 2017). Similarly, classical molecular dynamics indicated that Li_4SiO_4-Li_3PO_4 crystalline alloy mixtures exhibit 2–4 orders of magnitude higher ionic conductivity than either endpoint (Deng 2015).

The effect of extrinsic host disorder on ionic transport is not well understood, despite significant interest (Dyre et al. 2009), and numerical modeling plays a central role in this direction. Open questions remain about the nature of order-disorder phase transitions in lattices with quenched extrinsic disorder. A useful model to represent extrinsic quenched disorder is the "random-field Ising model," where each site experiences a random local chemical potential (Berretti 1985). For strong enough disorder, suppression and eventual elimination of order-disorder transitions are observed, but the cause is very different from geometric frustration. Monte Carlo simulations are used to examine different regimes that are not covered by analytical renormalization group techniques, far away from critical points. There is also incomplete understanding of how quenched disorder relates to ionic transport. Qualitative features of transport in disordered systems are illuminated using the random barrier model (Dyre 1988). In this model, a single particle is hopping on a lattice, where barrier heights between sites are drawn from a smooth probability distribution. In 2D and 3D, long-range transport happens along percolation paths composed of hops with low migration barriers. Thus diffusion at low temperatures is controlled by a single critical percolation energy barrier, explaining why Arrhenius behavior is observed in systems with a wide range of activation energies. In large-scale MD simulations of disordered systems, care must

be taken to ensure that tracer diffusivity is sampled for long enough time for the system to equilibrate and reach a truly long-range diffusive transport regime, as opposed to Li exploring a confined percolation cluster in a sub-diffusive mode. This can be done by checking whether the mean square displacement increases linearly with time, as expected from Fick's law and Einstein relation. At higher temperatures, a wider range of pathways and a larger fraction of activated mobile ions may start contributing to the transport. In experiment deviations from Arrhenius behavior are observed in glass ionic conductivity at high temperatures, but the reason is still not well understood (Kincs and Martin 1996).

6.1 Host Dynamics

In addition to static extrinsic disorder, in some of the best ion conductors, the host itself has mobile degrees of freedom that couple in nontrivial ways with the mobile ions. Host dynamics on the time scale of Li migration can effectively introduce fluctuations of the Li configuration energy landscape, preventing long-term locking in local minima. This effect is well known in liquid electrolytes, where solvation shells of mobile species are dynamic, resulting in very high diffusivity values. The most widely studied solid electrolytes relying on host dynamics are polymer salt solutions, particularly the family of polyethylene oxide (PEO)-based polymers. In these electrolytes, Li ions are coordinated by the polymer, and conduction mainly happens due to segmental motion of the flexible chains, with Li occasionally hopping from one chain to another. As a result, PEO polymer electrolytes only have substantial ionic conductivity above their glass transition temperature, when the host polymer has significant local mobility, similar to a viscous liquid. Due to the dynamic nature of the host, molecular dynamics is the only option for simulation, and due to the size of these amorphous systems (>5000 atoms) and long equilibration times (>10 ns), classical molecular dynamics is the method of choice, typically using force fields developed for organic molecules. In some inorganic materials, dynamic host behavior is also important in assisting ionic transport by preventing mobile ions from ordering and by increasing the rate of exploration of the configuration landscape. In oxides host dynamics are rare due to strong ionic interactions, but some examples are found in covalently bonded oxide polyanion materials. These types of dynamic mechanisms are material-specific, and due to the mixed ionic-covalent nature of bonding, ab initio molecular dynamics is mostly used to study the interplay between structure dynamics and ionic transport. For instance, high proton conductivity is demonstrated by ab initio MD to be mediated by fast SO_4 anion in-place rotations in $CsHSO_4$ proton electrolyte (Wood and Marzari 2007). In the recently discovered class of closo-borate fast ion conductors such as $Li_2B_{12}H_{12}$, the $(BH)_x^{2-}$ anion centers are ordered on a regular lattice due to long-range electrostatic interaction, but their icosahedral geometry is incompatible with the lattice symmetry (Kweon et al. 2017). This geometric frustration causes the anions to exhibit orientation disorder and in-place rotation. Li ions jump in the extremely dynamic energy landscape, driven by its fluctuations, and explore a large

number of available dynamic sites. This type of "dynamic frustration" is a promising mechanism to enhance ionic mobility explore and requires further understanding.

7 Outlook

There is converging understanding from recent molecular dynamics and experimental work that ionic transport in fast Li-ion conductors at room temperature is different from the classical picture of independent migrating defects and that concepts based on single particles are not entirely relevant. Evidence of importance of many-particle collective effects comes from molecular dynamics simulations and from measurements of the Haven ratio. Correlated transport phenomena are prominent, but their origins need deeper microscopic and statistical understanding. In a surprising number of cases, mobile sublattice disorder seems to dramatically improve ionic conductivity, whether it is due to frustration, extrinsic distortion and alloying, or dynamic coupling to the host fluctuations. In particular, Li sublattice frustration with maximal residual entropy may be a promising recipe for high ionic conductivity at low temperature, due to both high carrier concentration and their collectively mobile behavior. The reasons and mechanisms are not yet well known and may not be the same in different materials, but the complex regimes of high Li-ion concentration and strong interaction deserve a general systematic consideration. This further elevates the importance of explicit molecular dynamics approaches and development of fast transferable ways to accurately simulate complex ion dynamics, especially dealing with rare hopping events. Toward the goal of establishing a common framework of understanding room temperature fast ion transport, we underscore the relevance of rigorous methods developed in other fields for describing sublattice order-disorder phase transitions and transport in correlated liquids and disordered media.

References

Alder BJ, Wainwright TE (1962) Phase transition in elastic disks. Phys Rev 127:359–361

Aroyo MI et al (2006) Bilbao Crystallographic Server I: Databases and crystallographic computing program, Z. Krist 221(1):15–27

Balents L (2010) Spin liquids in frustrated magnets. Nature 464:199–208

Berretti A (1985) Some properties of random ising models. J Stat Phys 38(3–4):483–496

Catti M (2011) Short-range order and Li ion diffusion mechanisms in $Li_5La_9\square_2(TiO_3)_{16}$ (LLTO). Solid State Ionics 183:1–6

Deng Y (2015) Structural and mechanistic insights into fast lithium-ion conduction in Li_4SiO_4–Li_3PO_4 solid electrolytes. J Am Chem Soc 137(28):9136–9145

Doliwa B, Heuer A (2000) Cooperativity and spatial correlations near the glass transition: computer simulation results for hard spheres and disks. Phys Rev E 61:6898

Du YA, Holzwarth NA (2007) Li ion diffusion mechanisms in the crystalline electrolyte Li_3PO_4. J Electrochem Soc 154(11):A999–A1004

Düvel A et al (2017) Is geometric frustration-induced disorder a recipe for high ionic conductivity? J Am Chem Soc 139:5842–5848

Dyre JC, Maass P, Roling B, Sidebottom DL (2009) Fundamental questions relating to ion conduction in disordered solids. Rep Prog Phys 72:046501

Dyre JC (1988) The random free-energy barrier model for ac conduction in disordered solids. J Appl Phys 64:2456

Gebremichael Y, Vogel M, Glotzer S (2004) Particle dynamics and the development of string-like motion in a simulated monoatomic supercooled liquid. J Chem Phys 120:4415

He X, Zhu Y, Mo Y (2017) Origin of fast ion diffusion in super-ionic conductors. Nat Commun 8:15893

Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Classical and quantum dynamics in condensed phase simulations. World Scientific, Singapore pp 385–404

Kincs J, Martin SW (1996) Non-Arrhenius conductivity in glass: mobility and conductivity saturation effects. Phys Rev Lett 76:70

Kitaev A (2006) Anyons in an exactly solved model and beyond. Ann Phys (N Y) 321:2. https://www.sciencedirect.com/science/article/pii/S0003491605002381

Kweon KE et al (2017) Structural, chemical, and dynamical frustration: origins of superionic conductivity in *closo*-borate solid electrolytes. Chem Mater 29(21):9142–9153

Kozinsky B et al (2016) Effects of sublattice symmetry and frustration on ionic transport in garnet solid electrolytes. Phys Rev Lett 116:055901

Levitov L and Kozinsky B (1999) Charge ordering and hopping in a triangular array of quantum dots. arXiv:cond-mat/9912484v1

Meesala Y, Jena A, Chang H, Liu RS (2017) Recent advancements in Li-ion conductors for all-solid-state Li-ion batteries. ACS Energy Lett 2(12):2734–2751

Meier K, Laino T, Curioni A (2014) Solid-state electrolytes: revealing the mechanisms of Li-ion conduction in tetragonal and cubic LLZO by first-principles calculations. J Phys Chem C 118:6668–6679

Novikov DS, Kozinsky B, Levitov L (2005) Correlated electron states and transport in triangular arrays. Phys Rev B 72:235331

Van der Ven A, Ceder G, Asta M, Tepesch PD (2001) Phys Rev B 64:184307

Wheeler DR, Newman J (2004) Molecular dynamics simulations of multicomponent diffusion. 1. equilibrium method. J Phys Chem B 108(47):18353–18183

Wannier GH (1950) Antiferromagnetism. The triangular Ising net. Phys Rev 79:357–364

Wood BC, Marzari N (2007) Proton dynamics in superprotonic $CsHSO_4$. Phys Rev B 76:134301

Solid Oxide Fuel Cell Materials and Interfaces

52

Franziska Hess, Aleksandar T. Staykov, Bilge Yildiz, and John Kilner

Contents

F. Hess (✉)
Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: fhess@mit.edu

A. T. Staykov
International Institute for Carbon Neutral Energy Research (WPI-I2CNER), Kyushu University, Fukuoka, Japan
e-mail: alex@i2cner.kyushu-u.ac.jp

B. Yildiz
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: byildiz@mit.edu

J. Kilner
International Institute for Carbon Neutral Energy Research (WPI-I2CNER), Kyushu University, Fukuoka, Japan
Department of Materials, Imperial College London, London, UK
e-mail: j.kilner@imperial.ac.uk

Abstract

Perovskite materials are widely used as catalysts and electrocatalysts in energy conversion devices. Their bulk transport properties can be tuned flexibly for many applications. While the ion and electron transport properties in their bulk are generally well-understood, the surfaces remain much less explored and challenging. Surface instabilities give rise to compositional deviations in the near-surface region or even the precipitation of secondary phases. These deteriorate the surface catalytic properties under operating conditions, thus reducing efficiency and lifetime of devices.

In this contribution, the bulk and surface chemistry of perovskite oxides for solid oxide fuel cell applications are explored for selected compounds relevant for application. Starting with a short description of the bulk defect chemistry of perovskite oxides, the challenges for both theory and experiment in treating these materials are highlighted. Surface instabilities and a selection of mechanisms for surface stabilization are explored, such as electronic reconstruction, ionic segregation, and the formation of stable AO surface terminations. Finally, the O$_2$ dissociation is examined at two different stable AO terminations, where it is revealed that the nature of the active sites differs for the SrO and LaO terminations on SrTiO$_3$ and La$_2$NiO$_{4+\delta}$, respectively, so that the O$_2$ dissociation follows fundamentally different mechanisms on these surfaces.

1   Introduction

Perovskite oxides are an important class of materials that are widely applied in solid oxide fuel cells (SOFCs) and solid oxide electrolysis cells (SOECs) as electrode or electrolyte materials (Mahato et al. 2015; da Silva and de Souza 2017). They are also used as catalysts for electrochemical and thermochemical splitting of H$_2$O and CO$_2$ (Scheffe and Steinfeld 2014; Rao and Dey 2016) and as catalysts for production of fuels, such as H$_2$ and syngas (Agrafiotis et al. 2015). These oxides achieve high oxygen ion conductivity at high temperatures as well as high electronic conductivity, both of which are important properties for SOFC electrode materials. The exchange of oxygen at the surface limits the overall rate of the oxygen reduction reaction (ORR) in fuel cells and the oxygen evolution reaction (OER) in electrolysis and thermochemical splitting of H$_2$O and CO$_2$. In general, these reactions will be referred to as oxygen exchange.

Perovskite oxides can be represented by the general formula ABO$_3$, where A is a large cation (typically from the alkaline earth or rare earth groups) and B is a small cation (typically a transition metal). Due to the flexibility of the perovskite structure, the cations in the A- and B-sublattices can be easily substituted by other

ions, thus making it possible to tailor the bulk properties (such as electronic or ionic conductivity or the band gap) as required. The perovskite lattice was originally identified for the mineral $CaTiO_3$. The structure of the perovskite lattice depends on the proper ratio between the radii of the A-site and B-site cations (Abbes and Noura 2015; Goldschmidt 1926). Owing to the large ionic radii, alkaline and alkaline earth elements from the first and second group of the periodic table are often the choice for the A-site cation. They are characterized with low ionic charge: $+1$ or $+2$. B-site elements with small ionic radii are often the transition metals, which are characterized with large ionic charge. The B-site element, i.e., the transition metal, is situated in the center of an octahedral coordination sphere of oxide ions, where one oxide ion bridges two B-site centers. The B-site – O – B-site angle is characteristic for the perovskite lattice varying between 180° and 140° depending on the ratio of radii of A-site and B-site cations and the degree of doping. The tilted octahedral structure forms the BO_2 sublattice of the perovskites. It is worth noting that despite the perovskites showing mostly an ionic character, some degree of covalency is observed due to overlap between the d-orbitals at the B-sites and the 2p-orbitals at the oxide ions (Yagi et al. 2015). That leads to a deviation from the formal charges at the B-sites and the oxide ions and is the foundation for the high electronic conductivity found in some perovskites, e.g., $LaCoO_3$. The perovskite lattice is not limited to oxides but is a more general crystallographic structure that allows the incorporation of organic fragments, e.g., the hybrid perovskites recently applied in the solar cells (Wu et al. 2017).

One of the most important and best-studied perovskite materials for energy conversion applications is $La_{1-x}Sr_xMnO_3$ (LSM). It is a promising material for SOFC/SOEC electrodes that exhibits superior electronic conductivity, oxygen reduction kinetics, thermal stability, and compatibility (e.g., thermal expansion coefficient) with the commonly used solid electrolyte materials, such as yttrium-stabilized zirconia (YSZ) or gadolinium-doped ceria (GDC). As the oxygen exchange at the surface is typically rate-limiting, tuning of the electrode surface is a key aspect in the development and optimization of materials for energy conversion applications. With Sr-doped perovskite materials, a fundamental problem arises because these materials have a tendency to form passivating Sr-rich layers on the surface due to segregation of the dopant strontium ions (Li et al. 2017). This surface layer inhibits the oxygen exchange at the surface, thus increasing polarization resistance and reducing the energy efficiency of the SOFC cell.

A major application of complex perovskite materials is as air electrodes in high-temperature electrochemical solid oxide devices. This is because of their catalytic activity for molecular oxygen activation and molecular oxygen dissociation reactions on cathodes for solid oxide fuel cells, as well as molecular oxygen evolution reaction from anodes in solid oxide electrolyzers (Ishihara 2009). This catalytic activity is characterized by a high value of the surface exchange coefficient for oxygen (k), measured by techniques such as isotopic oxygen exchange, which gives a measure of the flux of oxygen crossing the surface at chemical equilibrium. A proper understanding of the underlying principles of these processes is crucial for the further optimization of the materials used in these solid oxide electrochemical devices.

While studying and understanding the ionic transport through the solid oxide electrolyte require expertise, mainly in the field of materials science, the reactions on the electrode surfaces add many additional levels of complexity. These reactions occur at the solid-gas interface, and their understanding requires expertise in the wider fields of materials science, surface science, and chemical catalysis (Burriel et al. 2014; Li et al. 2015). In order to be able to elucidate the surface processes on perovskite electrodes, one should be able to correctly understand the precise surface composition, surface reconstruction, the difference in properties between the surface and the bulk, as well as the evolution of the molecular chemistry of the catalytic reactions. This requires an active collaboration between material scientists, in materials and device preparation; spectroscopists, applying ion beam-based techniques such as low-energy ion scattering (LEIS) and secondary ion mass spectroscopy (SIMS) (Druce EES 2014; Chen et al. 2015; Téllez et al. 2015); and the efforts of theoreticians applying first-principles quantum mechanical methods to elucidate the molecular catalysis at atomistic scales (Mastrikov et al. 2018; Akbay et al. 2016; Staykov et al. 2015). Thus, to further our understanding of the oxygen reduction reaction on perovskite surfaces, it requires knowledge of the molecular chemistry of oxygen, understanding of the high-temperature surface reconstruction, the bulk material properties, and the catalytic chemistry on the perovskite surface.

2   Regimes in the Bulk Defect Chemistry of LSM

Many properties of perovskite materials, such as oxygen exchange and diffusion coefficients, band gap, and electronic conductivity, are strongly related to the bulk and surface defect chemistries of these materials. The dominant defects in oxides change with $p(O_2)$, and for perovskite oxides, five defect regimes can be identified as shown in Fig. 1a.

Nonstoichiometry, which introduces a high concentration of cation or anion vacancies, can be efficiently probed using thermogravimetric analysis (TGA) after equilibrating the sample in Ar–O_2, CO–CO_2, and H_2/H_2O gas mixtures (resulting in well-defined oxygen chemical potentials $\mu(O_2)$). For $LaMnO_3$, the characteristic nonstoichiometry isotherm, the curve of $3 \pm \delta$ vs. $Log(p(O_2)/p^0)$, is S-shaped as shown in Fig. 1b, with δ indicating the amount of oxygen excess ($+\delta$) or deficit ($-\delta$). Around the turning point, the composition is stoichiometric with an A/B/O ratio of 1/1/3. At higher (oxidation) or lower (reduction) $p(O_2)$, different defects dominate as indicated in Fig. 1a (top) with consequences for oxygen stoichiometry, oxygen vacancy concentration, and electronic conductivity as shown in the boxes in Fig. 1a.

Under reducing conditions (low $p(O_2)$ and/or high T), which correspond to the $3 - \delta$ branch in Fig. 1b, perovskite oxides form oxygen vacancies, thus reducing the transition metal cations (defect regime V):
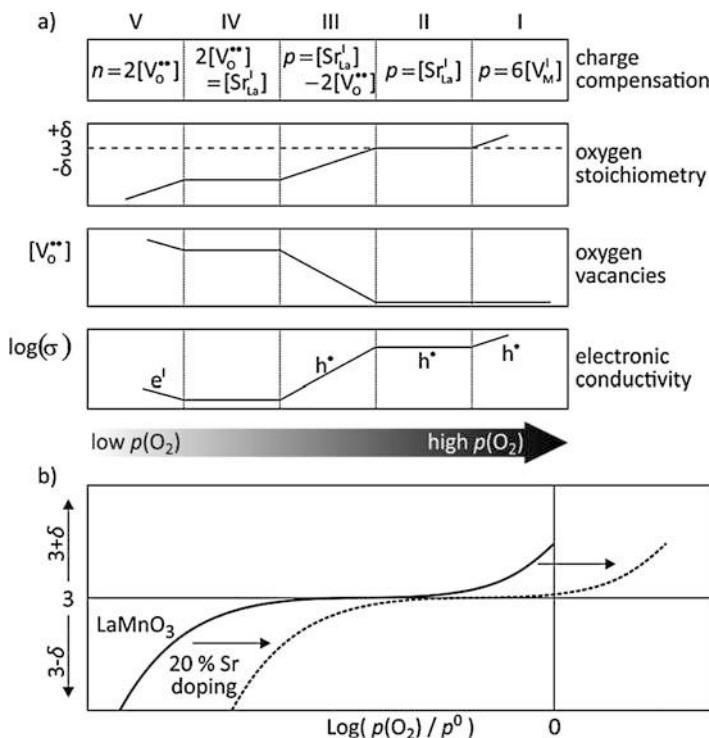
Fig. 1 Schematic diagram of defect chemistry in acceptor-doped perovskite oxides. (**a**) nonstoichiometry, oxygen vacancy concentration, and electronic conductivity as a function of $p(O_2)$, divided in five defect regimes. (**b**) Nonstoichiometry of $LaMnO_3$ (solid curve) and 20% Sr-doped $LaMnO_3$ (dashed curve) as a function of $Log(p(O_2))$. (Adapted from De Souza and Kilner 1998)

$$2Mn^x_{Mn} + O^x_O = 2Mn'_{Mn} + V^{\cdot\cdot}_o + \frac{1}{2}O_2 \tag{1}$$
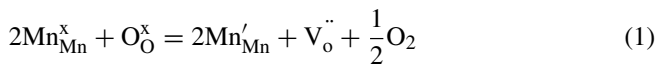
In the case of $LaMnO_3$, Mn^{3+} is reduced to Mn^{2+}. Under oxidizing conditions (high $p(O_2)$ and/or low T), $LaMnO_3$ incorporates additional oxygen into the lattice by forming cation vacancies in equal proportion on A- and B-sites (defect regime I):

$$6\,Mn^x_{Mn} + \frac{3}{2}\,O_2 = V'''_{La} + V'''_{Mn} + 6\,Mn^{\cdot}_{Mn} + 3O^x_O \tag{2}$$
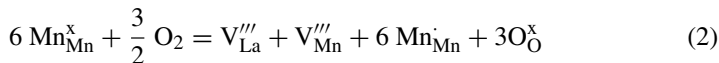
To compensate the high negative charge of the cation vacancies, 6 Mn^{3+} cations are oxidized to Mn^{4+}. The incorporation of additional oxygen is easily identified from the $3 + \delta$ branch as shown in Fig. 1b (solid line).

Doping $LaMnO_3$ with Sr^{2+} also leads to the formal oxidation of Mn^{3+} to Mn^{4+} (defect regime II):

$$Mn_{Mn}^x + La_{La}^x + Sr^{2+} = Mn_{Mn}^{\cdot} + Sr_{La}' + La^{3+} \tag{3}$$
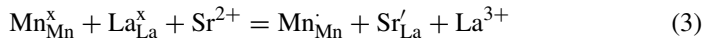
The oxidation of manganese through Sr doping suppresses the formation of cation vacancies, thus shifting the nonstoichiometry curve to higher $p(O_2)$ (cf. Fig. 1b, dashed line). Twenty percent doping is sufficient to make the oxygen-excess region inaccessible at $T > 1300$ K Mizusaki Solid State Ionics 2000. However, the oxidation of Mn^{3+} to Mn^{4+} facilitates the formation of oxygen vacancies, $V_o^{\cdot\cdot}$ (defect regime III):

$$2Mn_{Mn}^{\cdot} + O_O^x = 2Mn_{Mn}^x + V_o^{\cdot\cdot} + \frac{1}{2}O_2 \tag{4}$$
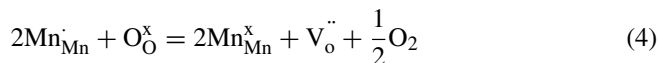
The occurrence of different charge states of manganese, Mn^{2+}, Mn^{3+}, and Mn^{4+}, can be considered implicitly by introducing the charge disproportionation equilibrium:

$$2Mn_{Mn}^x = Mn_{Mn}' + Mn_{Mn}^{\cdot} \tag{5}$$

Defect regime IV does not occur in acceptor-doped $LaMnO_3$, as doping is compensated by Mn oxidation, rather than the formation of oxygen vacancies.

2.1 Formation of Oxygen Vacancies

Bulk oxygen vacancies play a crucial role for oxygen transport properties in SOFC cathodes and electrolytes. Surface oxygen vacancies have also been proposed as a decisive factor for the activity of LSM in the dissociation of O_2 (Mastrikov et al. (2018); Kuklja et al. 2013). At technologically relevant doping levels of 20%, 1 bar O_2, and 900 °C, the oxygen vacancy concentration is on the order of 10^{-8} (Berenov et al. 1999). As a result, LSM20 has a slow oxygen diffusivity in comparison with other transition metal perovskite-based MIECs like LSCF. Before considering the modeling of oxygen vacancies, experimental results for LSM shall be summarized first in order to assess the reliability of first-principles calculations.

2.1.1 Experimental Investigation of Oxygen Vacancies in LSM

For determining defect formation energies from experiments, there are two popular approaches. In the first, the reaction energy of a defect formation reaction is obtained by fitting the temperature dependency of the corresponding equilibrium constant. These equilibration constants can be obtained by fitting the defect concentrations to nonstoichiometry data obtained from experiments. The thus-obtained energy is not a pure vacancy formation energy, as several effects (ionic and electronic) occur simultaneously. Consider the oxygen vacancy formation (reduction) in Eq. 4. Using the bracket notation for site fractions, the equilibrium constant can be written as

$$K_{\text{red}} = \frac{\left[\text{Mn}_{\text{Mn}}^{\text{x}}\right]^2 \left[\text{V}_{\text{o}}^{\cdot\cdot}\right] p_{\text{O}_2}^{1/2}}{\left[\text{Mn}_{\text{Mn}}^{\cdot}\right]^2 \left[\text{O}_{\text{O}}^{\text{x}}\right] p_0^{1/2}} = e^{-\frac{\Delta_{\text{red}} G_0}{RT}} \text{, and} \tag{6}$$

$$-R\, T \ln\left(K_{\text{red}}\right) = \Delta_{\text{red}} H - T\, \Delta_{\text{red}} S. \tag{7}$$

Throughout the text, p_0 indicates the standard pressure (10^5 Pa), T the temperature, R the universal gas constant, and $\Delta_i H$, $\Delta_i S$, $\Delta_i G_0$, and $\Delta_i K$ the molar enthalpy, entropy, standard Gibbs enthalpy (at $p^0 = 10^5$ Pa), and equilibrium constant of reaction i. Only the reaction enthalpy can be measured in this fashion, but not individual defect formation energies. The result is strongly intertwined with the applied defect model, and different defect models can give very different values for the formation of the same type of defect or even a different dependency on the Sr-doping content.

The second approach involves determining the ion mobility in a tracer diffusion experiment, where the vacancy concentration is implicitly contained and can be determined by using an approximation (Berenov et al. 1999). Again, the defect formation energy can be obtained by considering the temperature dependency of the defect concentration in similar fashion.

Naturally, these two approaches do not necessarily give the same results, as the two approaches operate under very different conditions and apply different approximations. Tracer diffusion experiments are typically conducted at $p(\text{O}_2) = 1$ bar, resulting in very low oxygen vacancy concentrations in LSM (on the order of 10^{-8}). The resulting values may thus be considered as dilute oxygen vacancy formation energies. However, under these conditions, cation vacancies V_{A}''' dominate in LSM, especially for low Sr-doping content. This possibly results in reduced $\text{V}_{\text{o}}^{\cdot\cdot}$ formation energy due to opposing defect charges and favorable defect association.

In TGA, however, the nonstoichiometry is detected by mass change. For reasonably fitting a defect model to TGA data, the defect concentration must be at least a few percent. Common fitting approaches neglect defect-defect interactions, but they do play a role at defect concentrations in the percent range. For these reasons, there are systematic deviations between the results from tracer diffusion and TGA.

A comparison of oxygen vacancy formation energies determined by TGA (blue) and tracer diffusion (green) is shown in Fig. 2. The literature sources used to compile this figure are listed in Table 1 (dilute concentration) and Table 2 (high concentration).

For both methods, data are available from different literature sources for different Sr-doping levels. The results within the two groups (low and high concentration) are remarkably consistent, which allows us to treat the data obtained by these two methods as whole sets of data, regardless of source. The only exception within the high-concentration group is Mebane et al. (2008), which uses a very different fitting model and is thus plotted separately in Fig. 2 (yellow squares).
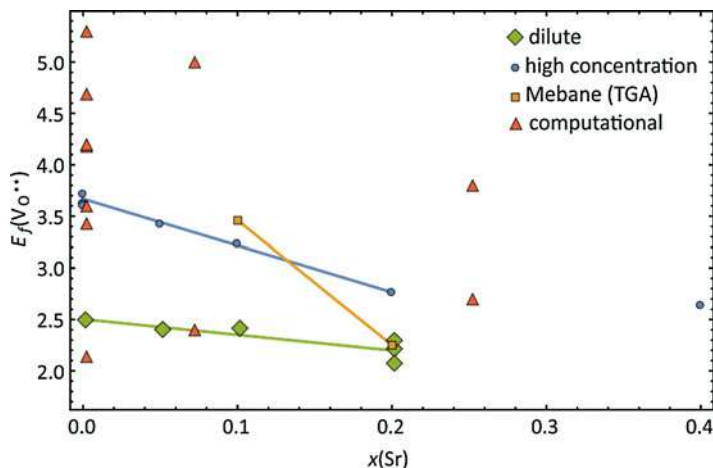
Fig. 2 Overview of literature values for the energy of oxygen vacancy formation from Table 1 (green diamonds), Table 2 (blue disks and yellow squares), and Table 3 (red triangles)

Table 1 Dilute oxygen vacancy formation energies estimated from oxygen tracer experiments

$x(Sr)$	References	E_{V_O}/eV	Note
0.00	Berenov et al. (1999)	3.04	Estimate from reported oxygen vacancy concentrations
0.05	Yasuda et al. (1996)	2.94	
0.10		2.41	
0.20		2.25	
0.20	De Souza et al. (2000)	2.07	Estimate from apparent activation energy of oxygen diffusion coefficient

Table 2 Experimentally determined oxygen vacancy formation energies at high $V_O^{\cdot\cdot}$ concentrations

$x(Sr)$	E_{V_O}/eV	Method and references
0.05	3.44	Chemical diffusion (Yasuda and Hishinuma 1996)
0.10	3.25	
0.20	2.78	
0.10	3.48	TGA, random defect model with vacancy-vacancy repulsion (Mebane et al. 2008)
0.20	2.27	
0.00	3.73	TGA, random defect model (Kuo et al. 1989)
0.00	3.64	TGA, random defect model (Zuev and Tsvetkov 2010)
0.00	3.62	TGA, defect cluster model (van Roosmalen and Cordfunke 1994)
0.40	2.65	TGA, random defect model (Takacs et al. 2016)

Table 3 Computationally determined oxygen vacancy formation energies

x(Sr)	References	E_{V_O}/eV
Cubic		
0	Olsson et al. (2016)	2.4
0	Curnan and Kitchin (2014)	4.2–4.3
0	Chen et al. (2011)	4.18
0.07	Sharma et al. (2015)	$\approx 2.4^a$
0.07	Sharma et al. (2015)	$\approx 5.0^b$
0.25	Chen et al. (2011)	3.8
0.25	Piskunov et al. (2008)	2.7
1	Curnan and Kitchin (2014)	2.9–3.5
Orthorhombic		
0	Lee and Morgan (2012)	3.43
0	Lee et al. (2009)	3.6
0	Kuklja et al. (2013)	4.69
0	Curnan and Kitchin (2014)	5.2
1.	Curnan and Kitchin (2014)	4.5
Rhombohedral		
0	Curnan and Kitchin (2014)	4.9
1.	Curnan and Kitchin (2014)	4.3

[a]V_O in center of slab, with Sr dopants in bulk
[b]V_O in center of slab, with Sr dopants at surface

For pure LMO, the oxygen vacancy formation energy at high concentration (blue dots) is about 3.6 eV. Tracer experiments at dilute concentration (green diamonds) give a much smaller value of 2.5 eV. This apparent discrepancy is resolved by the very different experimental conditions as outlined above.

However, both methods show remarkably similar dependence of E_{V_O} on Sr content: the oxygen vacancy formation energy decreases with increasing Sr content. This is consistent with the experimentally observed shift of the nonstoichiometry curve to higher $p(O_2)$ with increasing x(Sr) (Fig. 1b), which means that oxygen vacancies form more easily at higher Sr content.

The models used to fit nonstoichiometry data usually do not take into account the defect-defect interaction energy. The only exception is the TGA analysis by Liu and Mebane (orange curve), which introduces an interaction term describing the strain energy of close-by defects. As a result, the model gives very different values from other TGA analyses, as well as a much stronger dependency of the Sr content. Since there are only two points available, it is unclear if the relationship of E_{V_O} with x(Sr) is linear.

2.1.2 Modeling of Oxygen Vacancies by Density Functional Theory Calculations

Computations of the oxygen vacancy formation energy in LSM are also available (red triangles in Fig. 2). Compared to the experimental results, the computational values are remarkably inconsistent. Some of the inconsistencies can be explained

Table 4 Overview of computational parameters and resulting vacancy formation energies. Data are compiled from Curnan and Kitchin (2014)

Parameter	Range tested	A	$\left\lvert \Delta\left(E_{V_O}\right)\right\rvert$
Supercell[a]	$(1\times1\times1)$, $(1\times1\times2)$, $(2\times2\times1)$, $(2\times2\times2)$	La	0.5 eV
		Sr	0.5 eV
Space group	$Pm\bar{3}m$ (cubic), $R\bar{3}c$ (rhombohedral), $Pbnm$ (orthorhombic)	La	0.7 eV
		Sr	0.2 eV
Magnetism	FM, AFM	La	0.5 eV
		Sr	0.7 eV
U_{eff} (Mn 3d)	0–4.5	La	0.1 eV
		Sr	0.3 eV
O_2 correction	$E(O_2)$		$\frac{1}{2}\Delta E\,(O_2)$

[a]E_{V_O} usually decreases with increasing supercell size. In the study by Curnan and Kitchin (2014), the opposite was observed for the cubic crystal structure of $LaMnO_3$ (but not of $SrMnO_3$). This is unphysical and reflects the instability of the cubic structure, which is not a minimum on the potential energy surface for A = La.

by the use of different exchange functionals (generalized gradient approach (GGA) with Hubbard U correction (+U) or hybrid functionals), choice of the U parameter, assumed space group, choice of supercell size, and magnitude of O_2 overbinding (Wang et al. 2006) correction. Curnan et al. investigated the influence of these factors on the oxygen vacancy formation energies in $LaMnO_3$ and $SrMnO_3$. An overview of these results is shown in Table 4. These parameters influence the oxygen vacancy formation energies in various ways, which will be briefly discussed here.

The supercell size and shape dependency of E_{V_O} reflects the strength of defect-defect interactions. Within the examined range of supercell volumes (1–8 cubic unit cells), E_{V_O} scatters by about 0.5 eV for both $LaMnO_3$ and $SrMnO_3$, which is a large possible source of discrepancy when comparing different density functional theory (DFT) calculations with experiments but also when comparing different DFT studies with each other. It should also be kept in mind that $LaMnO_3$ has a rather low tolerance for oxygen vacancies. Pure $LaMnO_3$ decomposes already at a V_O concentration of about 5–10%, as inferred from TGA experiments, which corresponds to 1 vacancy in 3–7 unit cells. DFT calculations in smaller supercells do not reflect physical vacancy concentrations, thus overestimating the oxygen vacancy formation energy.

Structural distortions frequently occur in perovskites, reducing the symmetry from cubic to rhombohedral or orthorhombic. LSM is frequently modeled using cubic and orthorhombic space groups in DFT calculations, but their defect energetics are quite different. According to Curnan et al., the cubic structures usually show the lowest oxygen vacancy formation energies, regardless of A- and B-site cations. Introducing lattice distortions uniformly increases the oxygen vacancy formation energy, although the magnitude depends on the combination of A/B cations. For $LaMnO_3$, which is strongly distorted in the $R\bar{3}c$ and $Pbnm$ space groups due to the

Jahn-Teller distortion characteristic for Mn^{3+}, E_{V_O} is increased by up to 0.7 eV. In $SrMnO_3$, where no Jahn-Teller distortion occurs, the influence of lattice distortions on the oxygen vacancy formation energy is of a much smaller magnitude (up to 0.2 eV). It should be noted, however, that the cubic structure is not a minimum on the potential energy surface (PES) in some cases (such as LSM). Calculating vacancy formation energies for a structure that is not a minimum on the PES is not physically meaningful and should be avoided.

Magnetism plays a strong role in transition metal-based perovskite oxides, and the magnetic ground state often shows transitions within a series of compounds. Within the $La_{(1-x)}Sr_xMnO_3$ series, the ground state is ferromagnetic for $0.09 < x < 0.48$. For $x \leq 0.09$, a canted antiferromagnetic spin order is realized, and for $x \geq 0.48$ the ground state 0 K is antiferromagnetic (Tokura and Tomioka 1999). The magnetic state influences the redox properties of manganites. Neglecting magnetization or choosing an inappropriate magnetic state can cause errors of 0.5 eV–0.7 eV.

GGA+U can be employed to improve the computed properties of perovskite compounds by enhancing electron localization. U_{eff} is typically applied to the d-orbitals of transition metals. In the case of Mn compounds, the optimal choice of U_{eff} depends on the oxidation number (Lee and Morgan 2012). Values between 3 and 5 eV are typically found in the literature. The oxygen vacancy formation energy is not particularly sensitive to U_{eff}, as varying the parameter between 0 and 4.5 eV changes the oxygen vacancy formation energy only by up to 0.3 eV. However, U_{eff} should be chosen carefully, as some of the observed trends (such as relative stabilities of crystal structures) may change with U_{eff}.

The highest reported O_2 overbinding correction among the considered studies is 1.36 eV, so that over- or undercorrection of the O_2 energy constitutes a large possible influence (up to 0.7 eV) on the obtained oxygen vacancy formation energy.

These considerations allow us to put the inconsistencies between the theoretical results from the literature into better perspective and give a reference for the expected deviations between studies employing a slightly different computational framework.

2.2   Oxidizing Regime

A prominent feature of LSM at low Sr-doping content is oxygen superstoichiometry under oxidizing conditions (low temperature, high $p(O_2)$). The oxygen superstoichiometry is achieved not by oxygen interstitials but rather by cation vacancies following Eq. 2. The equilibrium constants for oxygen incorporation can be obtained by fitting a defect model to nonstoichiometry data obtained by TGA in the same principle fashion as previously outlined for oxygen vacancy formation. The cation vacancy formation energies are determined from the temperature dependency
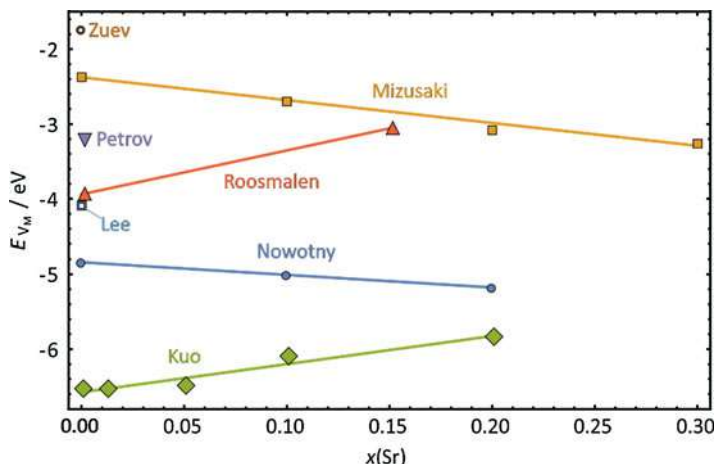
Fig. 3 Overview of literature values for the cation vacancy formation energy. (The figure is compiled from data in references listed in Table 5)

of these equilibrium constants. A selection of experimentally determined cation vacancy formation energies is shown in Fig. 3.

However, this approach is quite error-prone for cation vacancies for several reasons. The range of temperatures and $p(O_2)$ where cation vacancies occur in LSM is rather limited, especially for higher Sr-doping contents. The experiments must be able to measure very small stoichiometry variations very accurately, which is not always the case. For large cation vacancy concentrations, however, computational results suggest that defect-defect interactions are quite strong (Lee and Morgan 2012), making defect models that do not consider such interactions less reliable. Strong defect-defect interactions also make the usually applied ideal configurational entropy a debatable point because the defect distribution is not random when interactions are strong. In addition, several different defects, as well as entropy, need to be considered simultaneously, which can give rise to parameter compensation. As a consequence, the experimentally determined cation vacancy formation energies scatter substantially between literature sources, as shown in Fig. 3.

Each individual set of data in Fig. 3 shows a clear trend of E_{V_M} with respect to the Sr-doping content, but some sets show a decrease of E_{V_M} with $x(Sr)$, while others show an increase. Intuitively, an increase of E_{V_M} with $x(Sr)$ is expected due to the equal charge of defect and dopant. The absolute deviation between the data sets can be as high as 5 eV.

Since most computational studies focus on oxygen vacancy formation with the goal of modeling the oxygen transport, investigations of cation vacancy formation are scarce in the literature. In the case of LSM (for $x(Sr) \leq 0.20$), however, cation vacancies are the dominant defect under typical fuel cell conditions (1300 K, $p(O_2) = 1$ bar), so they should be considered when modeling fuel cell materials.

Table 5 Experimentally determined Cation vacancy formation energies in $La_{1-x}Sr_xMnO_3$ reported in literature

$x(Sr)$	E_{V_M}/eV	Method and reference
0.00	−6.54	TGA, random defect model (Kuo et al. 1989)
0.01	−6.54	
0.05	−6.50	
0.10	−6.10	
0.20	−5.85	
0.00	−2.35	TGA, random defect model with vacancy exclusion volume (Mizusaki et al. 2000)
0.10	−2.67	
0.20	−3.06	
0.30	−3.24	
0.00	−4.84	TGA, random defect model (Nowotny and Rekas 1998)
0.10	−5.01	
0.20	−5.18	
0.00	−3.94	TGA, random defect model (van Roosmalen and Cordfunke 1994)
0.15	−3.06	
0.00	−3.22	TGA, random defect model (Petrov et al. 2000)
0.00	−1.73	TGA, random defect model (Zuev and Tsvetkov 2010)
0.00	−4.07	DFT with vacancy-vacancy repulsion (Lee and Morgan 2012)

Lee and Morgan examined cation vacancy formation with a DFT-based model, taking vacancy-vacancy interaction explicitly into consideration but neglecting the influence of Sr on the cation vacancy formation (Lee and Morgan 2012). They report $E_{V_M} = -4.07$ eV at zero concentration. To obtain the defect formation energy E_{def}, the interaction is added as a linear function of the V_M concentration c_{def}:

$$E_{def}\left(c_{def}\right) = E_{rxn} + I_{rxn}c_{def}, \tag{8}$$

where E_{rxn} is the zero-concentration defect formation energy and I_{rxn} is the interaction term. The cation vacancy-vacancy interaction is very high at 16.10 eV compared to the oxygen vacancy-vacancy interaction at 2.54 eV. This suggests that the cation vacancy formation is limited by vacancy-vacancy repulsion, rather than the availability of oxidizable Mn^{2+}/Mn^{3+} cations as implicitly assumed by the ideal mixing entropy approach applied in experimental studies when modeling nonstoichiometry data.

These results highlight that concentration-dependent defect formation energies should always be employed in computational studies in order to obtain reliable defect concentrations. However, at strong interactions, the use of the ideal mixing entropy is questionable due to correlations that arise in the distribution of defects. This can be overcome by using more accurate approaches, such as the quasichemical approximation or lattice-based Monte Carlo simulations.

3 Surfaces and Defect Segregation

Although the bulk properties of LSM are well-understood, thus allowing for systematic material optimization, the surface properties are not. The surface structure and composition change dynamically with temperature, oxygen partial pressure, and polarization. This often leads to significant deviations from the bulk properties in the boundary region close to the surface after annealing. Cation segregation as observed in aliovalently doped perovskites (such as $La_{1-x}Sr_xMO_3$ where M = Co, Mn, Fe) not only leads to the enrichment of dopant cations close to the surface (Druce et al. 2014) but also to the formation of Sr-enriched secondary phase precipitates at the surface that are non-conductive (electronically and ionically) (Chen et al. 2012) and severely impede the oxygen exchange on the surface, thus leading to degradation of the electrode (Wang and Jiang 2006; Jiang 2008). The fundamental understanding and tuning of the properties of the electrode-gas interface thus remains a major challenge.

3.1 Electronic Reconstruction at Polar Surfaces

LSM surfaces and interfaces suffer from the fundamental problem of surface polarity. The structure is comprised of AO and BO_2 layers stacked in [001] direction. In the host structure $LaMnO_3$, where the A and B cations both have a +3 charge, these AO and BO_2 layers have +1 and −1 net charge, respectively. Cleaving the structure as shown in Fig. 4 creates dipolar surfaces of Tasker type 3 (Tasker 1979) with diverging surface energy, which cannot be stable without undergoing significant charge rearrangement, defect segregation, or surface reconstruction. In materials like LSM, which can support a complex defect chemistry, the transition metal (TM) cation can change oxidation state, possibly reducing surface charge at a lower energy cost compared to ionic rearrangement. Such charge redistribution frequently occurs at polar surfaces of perovskite compounds, such as the $SrTiO_3(111)$ surface (Pojani et al. 1999), as well as at heterointerfaces of perovskite compounds with different charge types (Nakagawa et al. 2006). The MnO_2 termination of LSM undergoes a similar electronic reconstruction as discussed in the following.

Fig. 4 Cleavage of a 3–3 perovskite, e.g., $LaMnO_3$, forms two surfaces, terminated by AO and BO_2 layers. These terminations are polar, inducing segregation of defects with opposite charge to the surface and defects with equal charge into the bulk to reduce surface polarity
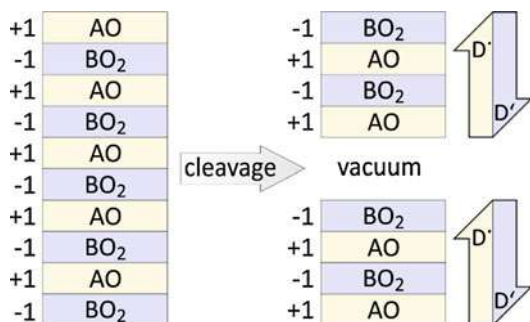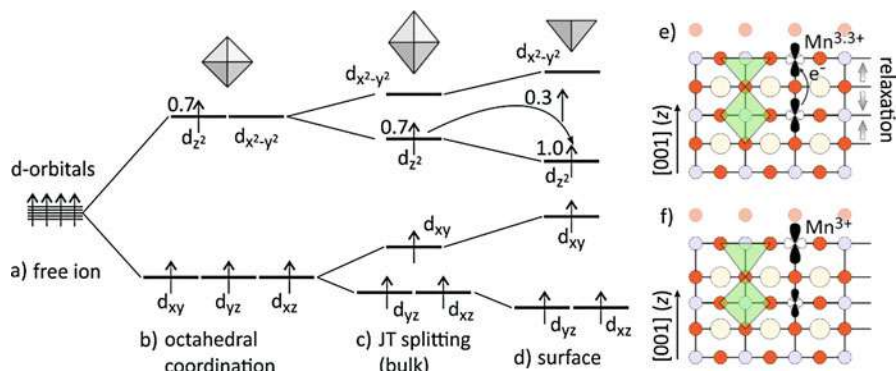
Fig. 5 Crystal field model of $Mn^{3.3+}$ in $La_{0.7}Sr_{0.3}MnO_3$. (**a**) Free Mn^{3+} ion, (**b**) crystal field splitting of $Mn^{3.3+}$ in octahedral coordination, (**c**) Jahn-Teller splitting as it occurs in bulk $La_{0.7}Sr_{0.3}MnO_3$, (**d**) approximation of undercoordinated Mn at the surface as a square-pyramidal complex. Since the $e_g\left(d_{z^2}\right)$ orbital energy is lowered compared to the underlying bulk-Mn, electron density is transferred from bulk to surface, reducing the surface-Mn to Mn^{3+} (Calderón et al. 1999; Zenia et al. 2005; Cui et al. 2014; Pesquera et al. 2012). (**e**) and (**f**) Surface and bulk coordination in LSM before (**e**) and after (**f**) relaxation of electronic and ionic structure (blue, Mn; yellow, La/Sr; red, O)

To understand the electronic reconstruction that occurs at the MnO_2 termination of LSM (Calderón et al. 1999; Zenia et al. 2005; Cui et al. 2014; Pesquera et al. 2012), consider first the electronic configuration of Mn in bulk LSM (Fig. 5b). For a Sr-doping level of 30%, LSM is ferromagnetic and semimetallic. The average oxidation number of Mn is +3.3. $Mn^{+3.3}$ is octahedrally coordinated, which splits the d-orbitals into t_{2g} (d_{xy}, d_{xz}, d_{yz}) and e_g ($d_{x^2-y^2}, d_{z^2}$) levels (Fig. 5b). In this high-spin $d^{3.7}$ configuration, Jahn-Teller splitting occurs in the bulk, which splits the t_{2g} and e_g levels (Fig. 5c). The energies of the antibonding d_{z^2} and $d_{x^2-y^2}$ levels are reduced and increased, respectively, which causes elongation of the bonds along the z direction (Jahn-Teller elongation).

At the surface, one O^{2-} is missing, so that the surface-$Mn^{3.3+}$ is approximately in a square-pyramidal coordination (Fig. 5c). This reduces the energy of the d_{z^2} orbital even further (Calderón et al. 1999). This reduction of the energy level allows the surface-Mn to accept 0.3 electrons from the subsurface layer, reducing the surface-Mn to Mn^{3+} while retaining the high-spin configuration. The subsurface-Mn is oxidized to $Mn^{3.6+}$, shortening its Mn-O bonds along the z direction upon subsequent ionic relaxation (Fig. 5e, f).

3.2   Segregation of Charged Defects

Since the surface-Mn is reduced to +3, the surface polarity persists despite this electronic reconstruction, which can attract or repel defects with opposing or equal charge, respectively. Similar considerations apply for other polar terminations, such as the pure LaO, mixed (La,Sr)O, and SrO terminations. These polar surface

terminations have been proposed to play an important role in dopant segregation and subsequent degradation of oxygen exchange activity through the formation of insulating areas on the electrode or precipitation of secondary phases.

Various computational studies have addressed defect segregation at (001) surfaces of perovskites, considering electrostatic and elastic driving forces for dopant segregation. The general conclusion drawn from these studies is that electrostatic contributions play the strongest role in defect segregation at polar surfaces (Harrison 2011; Lee and Morgan 2015; Lee et al. 2013). In a purely electrostatic picture, negatively charged defects (V_A''', V_B''', Sr_A', Mn_B') are drawn to the positively charged LaO termination in LSM, and positively charged defects $\left(V_o^{\cdot\cdot}, Mn_B^{\cdot}\right)$ are drawn to the negatively charged MnO_2 termination. Mn_B' and $Mn_B^{\cdot}$ represent Mn in 2+ and 4+ charge state, respectively. The previously discussed electronic reconstruction of the MnO_2 termination may be considered as such a segregation of electronic defects.

The electrostatic interaction is the dominant term in segregation at polar surfaces, as can be rationalized by comparing the oxygen vacancy segregation at nonpolar $SrTiO_3(001)$ surfaces with the polar $LaMnO_3(001)$ surfaces. The $V_o^{\cdot\cdot}$ segregation profiles at the BO_2 surfaces of $SrTiO_3(001)$ and $LaMnO_3(001)$ are reproduced in Fig. 6a and b from Lee and Morgan (2015). In the absence of a clear segregation trend, the profile for $SrTiO_3$ (a) shows a zigzag shape. The $V_o^{\cdot\cdot}$ segregation energy at $SrTiO_3(001)$ is strongly negative only for the termination layer. For deeper layers, the segregation energy is either close to 0 (for BO_2 layers) or slightly positive (for AO layers). The negative $V_o^{\cdot\cdot}$ segregation energy at the topmost BO_2 surface layers
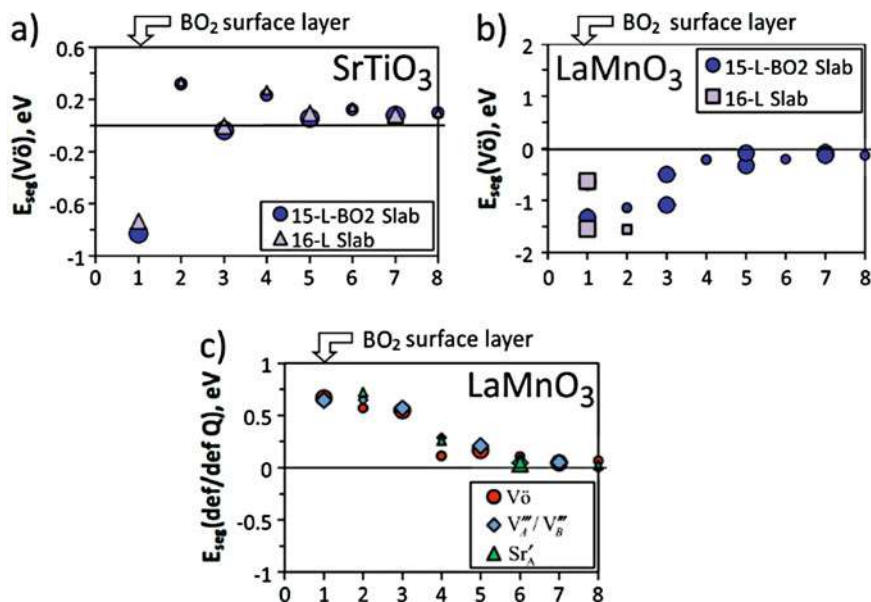
Fig. 6 Defect segregation energies of $V_o^{\cdot\cdot}$ at BO_2 surfaces of (**a**) $SrTiO_3$ and (**b**) $LaMnO_3$. (**c**) Defect segregation energies normalized by charge. (Figure adapted from Lee and Morgan 2015)

indicates the additional degrees of freedom at the surface, which allow the lattice to relax and accommodate the vacancy more easily compared to the bulk. In the AO bulk layers, removing one O^{2-} constitutes a larger perturbation compared to the BO_2 layers, thus reflecting finite-size effects of the slab calculations, which are most likely artificial (Lee and Morgan 2015).

The BO_2 surface of $LaMnO_3$ has a negative polarity, thus attracting positively charged defects like $V_o^{\cdot\cdot}$. As a consequence, the $V_o^{\cdot\cdot}$ segregation profile at $LaMnO_3$ in Fig. 6b) shows a clear trend of increasingly negative segregation as the $V_o^{\cdot\cdot}$ is moved closer to the surface.

Similar trends are observed also for other defects at the BO_2 surface of $LaMnO_3(001)$. Normalizing the defect segregation energy by the formal defect charge Q reveals a unique profile for $V_o^{\cdot\cdot}$, $V_{A/B}'''$ and Sr_{La}', as shown in Fig. 6. This is indirect evidence that the defect segregation is mostly governed by the electrostatic interaction between the polar surface and the defect charge. This electrostatic interaction, however, can be enhanced or counterbalanced by the coexistence of other point defects of opposing or equal charge close to the surface (Lee et al. 2013), and the trend of the classic electrostatic energy for coexisting point defects

$$E_{elec} = \sum -\frac{Q_1 Q_2}{4 \pi \varepsilon R_{12}} \tag{9}$$

was found to correlate perfectly with the trends obtained by DFT calculations in $LaMnO_3$ and $SmMnO_3$ (Lee et al. 2013), with Q_1 and Q_1 indicating the charges of defect and surface, ε the dielectric constant, and R_{12} the distance between defect and surface. More recently, experimental tuning of the oxygen vacancy concentration at the surface was shown to correlate with enhancement in chemical stability and electrochemical activity on Sr-substituted $LaCoO_3$ (Tsvetkov et al. 2016). However, which defects coexist at the surface in stable configurations at a given set of chemical potentials and whether or how these defects are responsible for the precipitation of secondary phases are still a matter of debate.

The elastic contribution to the segregation energy is more straightforward and has been investigated computationally for different combinations of A^{2+}- (Lee and Morgan 2015; Kwon et al. 2016) and B- (Kwon et al. 2016) site cations, as well as strain (Han et al. 2011). The change in free volume ΔV_f upon doping was found as a suitable predictor for cation segregation (Kwon et al. 2016), revealing a linear relationship between ΔV_f and segregation energy. The free volume quantifies the exclusion volume of the unit cell after subtracting the ions, modeled as hard spheres with defined radii, and is expressed as

$$\Delta V_f = \left(a_{doped}^3 - a_{host}^3\right) - \frac{4}{3}\pi \left(r_{dopant}^3 - r_{host}^3\right), \tag{10}$$

where a_{doped}^3 and a_{host}^3 are the unit cell volumes of doped and host structure and r_{dopant} and r_{host} are the ionic radii of dopant and host cations, respectively.

3.3   Stable Surface Terminations of Perovskite Oxides

A perovskite lattice can be cleaved in various low-index miller planes with different concentration of A-site and B-site cations within the surface layer. Thus, the surface catalytic activity would depend on the particular plane and the exposure of A-sites and B-sites. Theoretical studies on $SrTiO_3$ have shown that AO-terminated (001) surfaces and BO_2-terminated (001) surfaces are characterized with identical surface energy and both should be observed experimentally. Oxygen reduction reactions were studied computationally on La_2NiO_4 La-/Ni-terminated surfaces, concluding that Ni ions exposed to the surface are beneficial for the catalytic activity (Zhou et al. 2013). These theoretical results, however, have been contradicted by the vast majority of experimental observations which reported that under SOFC operational conditions, the surfaces of Sr-containing perovskites show significant Sr segregation (Lee et al. 2015 Phys. Rev B; Lee et al. 2015 PCCP). Scanning tunneling microscopy studies have shown that as a result SrO surface layers and Sr-rich nanoparticles have been formed (Chen Energy Environ. Sci. 2012). As layered materials, perovskites are well-suited to be studied with the LEIS technique (Druce Energy Environ. Sci. 2014). Unlike other characterization techniques, i.e., X-ray photoelectron spectroscopy (XPS), atomic force microscopy (AFM), etc., LEIS offers direct measurement of the chemical composition of the outermost surface layer and also offers the facility to obtain layer-by-layer depth profile of the outermost 20 nm of the material. Thus, the technique is sensitive enough to provide accurate information of the chemical composition of the surface termination and reconstruction of perovskite materials.

LEIS analysis of various functional perovskites used as electrodes in SOFCs at elevated temperatures has shown that all those materials are characterized by AO surface termination (Druce Energy Environ. Sci. 2014). Among the oxides investigated were $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$ (LSCF113, single perovskite), $GdBaCo_2O_{5+\delta}$ (GBCO-1125, double perovskite), and $La_2NiO_{4+\delta}$ (LNO-214, Ruddlesden-Popper structure). All those compounds are industrially relevant materials for application in solid oxide devices. In addition, single crystal $SrTiO_3$ was investigated as a model perovskite material. High-sensitivity LEIS depth profiling was performed to characterize the surface and near-surface composition of polycrystalline ceramics representative of the most promising perovskite-based mixed ionic and electronic conducting materials for fuel cell applications. The LEIS measurements are performed after grinding and polishing of the ceramic materials to produce fresh surfaces. The samples were subjected to high-temperature annealing treatments of 1000 °C for 12 h to simulate operating conditions of solid oxide electrodes. In all cases, the AO surface termination was observed. The experimentally observed surface terminations contradict the assumptions often made in computational modeling studies of the surface structure and oxygen exchange reaction. In the vast majority of computational modeling papers, the surfaces are assumed to be either a mixed AO- and BO_2-terminated or a BO_2-dominated surface to facilitate electron transfer to reduce molecular O_2 (Zhou et al. 2013; Eglitis 2009; Cai et al. 2006; Eglitis 2015).
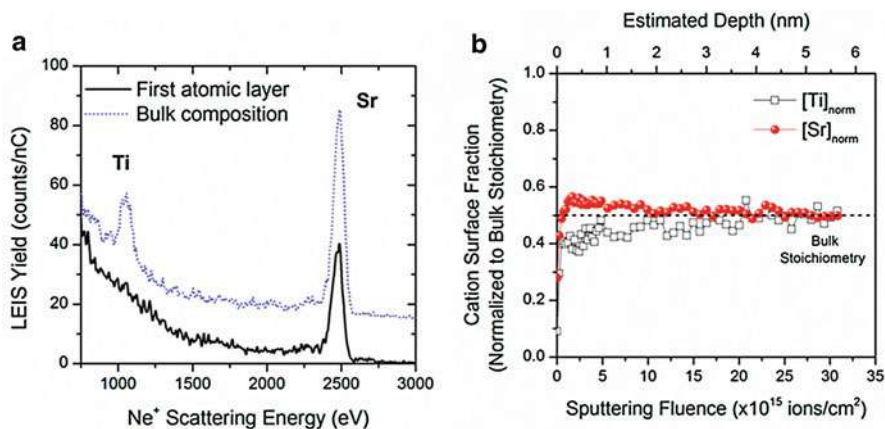
Fig. 7 Sequential LEIS spectra showing the evolution of the spectra with 500 eV Ar sputtering. The sample is a single crystal of $SrTiO_3$ after treatment at 1000 °C for 2 h in high purity oxygen. (Figure adapted from Staykov et al. 2015)

The investigated samples reveal through depth profile analysis that the very outermost surface is enriched in the A cation. Light sputtering reveals the B cations, which grow in intensity to show a second region which is slightly enriched in these species compared to the bulk stoichiometry. Finally, after sufficient sputtering, the bulk composition is reached, showing a plateau in the cation surface coverage, which was assumed to correspond to the bulk stoichiometry. An example set of spectra are shown in Fig. 7 for an $SrTiO_3$ sample heat-treated in oxygen (Staykov et al. 2015). The outermost layer shows an absence of the peak for the Ti cations, while the subsurface layers are Ti cations enriched. The bulk stoichiometry is observed after depths of 1–2 nm beneath the surface have been reached.

In addition to the information from surface analysis techniques, first-principles studies have shown AO-terminated surface reconstructions for polar perovskite surfaces, in particular $KTaO_3$ (Deacon-Smith et al. 2014; Setvin et al. 2018). The reason behind is that interlayer cation exchange stabilizes the polar perovskite surfaces. The Tasker theory for surface termination of ionic materials distinguishes between three possible surface types (Tasker 1979). Type one is characterized with planes containing positive and negative ions that completely compensate each other, and as a result, the individual planes are neutral. Type two is the result of alternation of one plane with either positive or negative charge followed by two planes characterized with half countercharge. Those surfaces should be cleaved in unit cells in order to avoid surface polarization. Type three surfaces are characterized with alternation of positively and negatively charged planes. As a result, a strong surface dipole moment arises which in reality is avoided by surface reconstruction, and if not compensated, it might lead to strong artificial effects in theoretical simulations. Woodley and co-authors have shown that cation exchange between the surface and subsurface layers of polar perovskites leading to AO surface termination and B-enriched sub-

surface region effectively eliminates the surface dipole momentum and is the natural mechanism to deal with surface polarity of ABO_3 perovskites (Deacon-Smith et al. 2014). The driving force for the cation exchange was the different electronic properties of the A- and B-site cations. A-site cations possess large ionic radii with usual charge of $+1$ or $+2$. In contrast B-site cations are characterized with small ionic radius and strongly localized point charge between $+3$ and $+5$. As a result, they build an octahedral symmetry coordination sphere with oxide ions which compensates the charge localization. When exposed to the surface, the B-site coordination sphere is broken, and surface-localized point charges give rise to dipole moments and high surface energies. By B-site/A-site cation exchange, the surface energy is minimized, the oxide coordination spheres of the B-site cations are restored, and the formation of a surface dipole is avoided (Deacon-Smith et al. 2014).

4 O_2 Activation on Perovskite Surfaces

In the gas phase, the oxygen molecule is stable in its triplet ground state with two unpaired electrons occupying degenerate π-molecular orbitals with an orthogonal spatial distribution. Any dissociation of the oxygen molecule progresses through its activated states, i.e., singlet oxygen, superoxo state (O_2^-), or peroxo state (O_2^{2-}). The electron configurations of the oxygen molecule, the superoxo state, and the peroxo state are shown in Fig. 8. The oxygen activation requires significant interaction between the molecule and the catalytic surface and possible electron transfer. The activated states of the oxygen molecule are characterized by different spins and bond lengths. The superoxo state has radical character and O−O bond of 1.3–1.4 Å, while the peroxo state is singlet and is characterized with an O−O
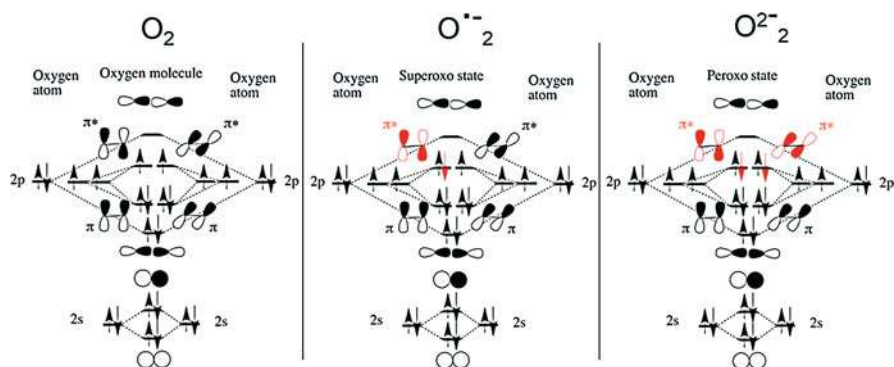
Fig. 8 Molecular orbital diagrams of dioxygen molecule (left), superoxo state of oxygen (center), and peroxo state of oxygen (right). The molecular orbitals are constructed by the atomic orbitals of two isolated oxygen atoms. The orbitals in black are those of the dioxygen molecule. A red color denotes the additional electrons transferred from the surface to the molecule (in order to obtain superoxo and peroxo states), as well as the molecular orbitals, which receive those additional electrons

bond length of 1.4–1.6 Å. These can be distinguished computationally, as well as experimentally, using electron paramagnetic resonance (EPR) and infrared (IR) spectroscopy. The superoxo state is characterized with high catalytic activity toward the oxidation reaction due to its radical character, while the peroxo state is known as a mild oxidant. Both the superoxo and the peroxo states are precursors for the molecular oxygen dissociation reaction.

4.1   Oxygen Activation on the SrO-Terminated $SrTi_xFe_{1-x}O_{3-\delta}$ Surface

Dissociation of molecular oxygen starts with adsorption on the substrate. Two orientations of the oxygen molecule are possible: end-on, with a single oxygen atom facing the substrate (perpendicular to the surface); and side-on, with two oxygen atoms facing the substrate (parallel to the surface) (Akbay et al. 2016; Staykov et al. 2015, 2008). The adsorption geometry is important for the further reaction mechanism and activation barrier for oxygen dissociation. Often, end-on adsorption geometry is realized in homogeneous organometallic and enzymatic catalysis, while side-on adsorption is preferred in the heterogeneous surface catalysis. While in the end-on adsorption geometry the oxygen molecule is attached only to one center and the substrate geometry has only minor effect on the dissociation mechanism, in the side-on adsorption, the two oxygen atoms are linked to different centers from the substrate and the substrate geometry, i.e., bond lengths, lattice constants, etc. can be of major importance.

As shown in Sect. 3.3, the surfaces of relevant perovskite materials are AO-terminated under typical SOFC conditions. Thus, it is important to address the question of the surface reactivity and catalytic properties of AO-terminated perovskite surfaces. For this purpose, the mechanisms for oxygen reduction reactions on a model material such as $SrTiO_3$ and industrially relevant material La_2NiO_4 will be reviewed. The simulations were performed with periodic plane wave density functional theory (DFT) implemented in the Vienna ab initio simulation package (VASP). The Perdew−Burke−Ernzerhof exchange-correlation functional (PBE) was employed using projector augmented wave pseudopotentials (Kresse and Hafner 1993; Kresse and Furthmüller 1996a, b; Blöchl 1994). DFT+U was used when necessary to localize electrons at the d-orbitals of transition metals. The details of those simulations are provided in the original research articles (Akbay et al. 2016; Staykov et al. 2015).

Oxygen reduction on an SrO-terminated $SrTiO_3$ surface was estimated to be an energetically unfavorable endothermic process. The reason is in the strong ionicity of Sr^{2+} and O^{2-}. Those surface ionic species do not participate in electron exchange with the molecular oxygen. However, perovskite materials are characterized with a moderate concentration of oxygen vacancies, and the simulation results summarized in Fig. 9 show that oxygen vacancies are slightly stabilized within the surface layer by 0.15 eV (Staykov et al. 2015).

Fig. 9 Subsurface to surface migration mechanism of an oxygen vacancy in $SrTiO_3$. (Figure adapted from Staykov et al. 2015)
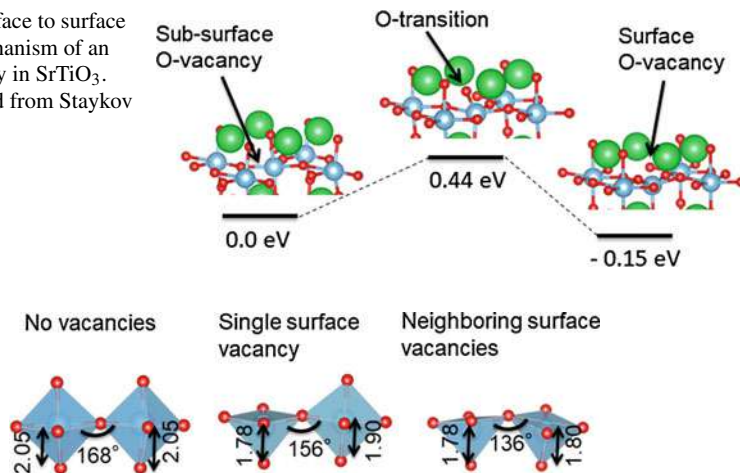
Fig. 10 Geometry of two neighboring Ti-octahedra in the subsurface layer of $SrTiO_3$, corresponding to the vacancy free case, the single oxygen vacancy case, and two neighboring oxygen vacancies. For simplicity, the Sr-ions are omitted in this representation. All distances are in Å. (The figure is compiled from data in Staykov et al. 2015)

The surface oxygen vacancy removes the apical oxygen in the terminating BO_6 octahedron and exposes the electronic density at a B-site cation center. This is important because it allows the surface adsorbed oxygen species to gain access to the partially delocalized electrons in the conducting –B–O–B– backbone network of the perovskites. Oxygen ion migration between the subsurface and the surface layers is characterized with an activation barrier of 0.44 eV. Interestingly, oxygen ion migration on the surface is characterized with a significantly higher activation barrier of 0.98 eV. The reason behind the higher activation energy necessary to transport vacancies on the surface compared to the bulk is that on the surface the Ti–O bond should be completely cleaved, while through the bulk the oxide ion shifts between different sites of the Ti coordination sphere, where one Ti–O bond is cleaved while another Ti-O bond is formed. Thus, the migrating oxide ion always shares electron density with Ti. Another important property of surface oxygen vacancies is that the energy is significantly decreased as vacancies occupy neighboring sites due to relaxation of the electron density due to decrease of the Ti–O–Ti angles close to the vacancy sites.

In Fig. 10, the Ti-O-Ti angle for vacancy free surface equals 168°. In the case of a single oxygen vacancy, the Ti ion is attracted away from the surface by the apical oxide ion below it, and the angle now equals to 156°. For two oxygen vacancies, both Ti ions are attracted below the surface resulting in an angle of 136° (Staykov et al. 2015). It is worth noting that the oxygen atom is in sp^3 hybridization with an optimal relaxed electron density when the four orbitals are pointing toward the apices of a tetrahedron. Two of those hybrid orbitals are occupied by electron pairs, and the other two give rise to the valence bonds of oxygen. Thus, the preferred
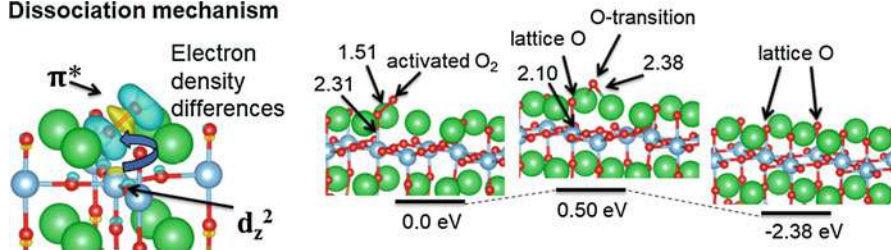
Fig. 11 Dissociation mechanism of oxygen molecule over a pair of surface oxygen vacancies on the SrO-terminated surface of SrTiO₃. (Figure adapted from Staykov et al. 2015)

oxygen valence bond geometry is V-shaped. The electron density around the oxygen atom is strongly strained in the perovskite lattice and is relaxed by coupling vacancies on the perovskite surface.

Having established the important electronic nature of the surface oxygen vacancy, it is now important to understand how this will interact with a gas-phase oxygen molecule. In contrast to the perfect surface, the oxygen molecule from the gas phase readily interacts with the surface oxygen vacancy. As a result, electron density is transferred from the B-O-B-O network, through the d_z^2 atomic orbital of the Ti, to the antibonding π^* orbital of the oxygen molecule (Staykov et al. 2018). As a result, the oxygen-oxygen bond is elongated to 1.5 Å, and oxygen is activated to its peroxo state. From the peroxo state, the oxygen molecule is dissociated as one oxygen atom occupies the lattice vacancy site and the second oxygen atom passes over a Sr-Sr bridge and occupies a nearby oxygen vacancy. The activation barrier for the oxygen dissociation is 0.5 eV, and the reaction is exothermic with 2.38 eV (Staykov et al. 2015). The dissociation mechanism is summarized in Fig. 11. In the case that a second surface vacancy is not available in the vicinity which can be occupied by the second oxygen atom, the surface oxygen exchange reaction can still occur, and a cross molecule can be formed between an oxygen atom from the gas phase and an oxygen from the surface.

From these results on the model system SrTiO₃, it is possible to infer that the same oxygen incorporation mechanism will be active in other perovskite materials, specifically those where bulk oxygen transport has been enhanced by the substitution on the A-site by Sr. Such materials can display very high values of the surface exchange coefficient (k) and have found extensive use as air electrodes and include the mixed ionic electronic conductors $La_{1-x}Sr_xCoO_{3-\delta}$, $La_{1-x}Sr_xMnO_{3-\delta}$, and $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$. As mentioned in Sect. 3.3, at the working temperature of these devices, the surfaces will be terminated by SrO; however, most of the materials will also have a substantial population of oxygen vacancies. Similar to SrTiO₃, the surface vacancies will provide access for the gas-phase oxygen molecules to the electronically conducting B-O-B network and hence the electron transfer needed for dissociation.

In a recent work on Sr-doped LaCoO₃, experimental evidence has been provided for the effect of surface oxygen vacancies and the exposure of the B-site of the

perovskite lattice for the activity of oxygen reduction reaction using real-time impedance spectroscopy (Rupp et al. 2017). They have successfully estimated that on SrO-terminated surface, oxygen reduction occurs at selected spots corresponding to surface oxygen vacancies and exposed Co ions. With an increase of both temperature and observation time, the surface was completely covered by Sr segregation leading to inert, inactive surface composition. As Co (Co_3O_4) was deposited to the surface, the catalytic activity was restored. The experiment of Rupp et al. clearly demonstrates that on the SrO-terminated perovskite surface, molecular oxygen dissociation occurs only on exposed B-site cations (Rupp et al. 2017). Thus, surface oxygen vacancies are crucial for the oxygen exchange reaction in such substituted perovskite materials.

4.2   Oxygen Activation on the LaO-Terminated La$_2$NiO$_{4+\delta}$ Surface

The oxygen vacancy facilitated oxygen dissociation mechanism on the SrO-terminated surface of $SrTiO_3$ leaves open the question about the large surface exchange coefficients observed on the LaO-terminated surface of $La_2NiO_{4+\delta}$ (Burriel et al. 2008; Tarancon et al. 2007; Skinner and Kilner 2000). $La_2NiO_{4+\delta}$ is a layered perovskite related to Ruddlesden-Popper oxide which has been observed to show the same AO termination as the simple perovskites after high-temperature exposure to O_2. The material is not Sr-substituted and is usually prepared as an oxygen-rich material where the oxygen excess is accommodated by oxygen interstitials. This high density of oxygen interstitials has the effect of suppressing the oxygen vacancy concentration through the Frenkel equilibrium, and thus the oxygen vacancy concentration is vanishingly small. In order to investigate this apparent contradiction in the earlier findings of the importance of the surface oxygen vacancy, a theoretical study of molecular oxygen dissociation on an LaO-terminated La_2NiO_4 surface was performed by Akbay et al. (2016). The simulation results show that unlike the AO termination in $SrTiO_3$, on the LaO surface, molecular oxygen can be dissociated without the support of surface oxygen vacancies. The dissociation mechanism is summarized in Fig. 12.

The oxygen molecule is chemisorbed over a La-La bridge on the LaO surface. Such an adsorption geometry was not observed on the SrO-terminated $SrTiO_3$. The La-O-O-La bridge geometry resembles the molecular oxygen adsorption on noble metals such as Pd (Staykov et al. 2008). In this geometry, the oxygen molecule is activated to the superoxo state upon electron transfer from the LaO surface to the antibonding π^* orbital of the oxygen molecule. The oxygen-oxygen bond is elongated to 1.33 Å. The chemisorption state is stabilized by 0.59 eV. The oxygen molecule is dissociated with an activation barrier of 1.1 eV, and the two surface oxygen species take the position of an La-O-La surface bridge. The charge on the surface oxygen species is -1. The overall reaction is slightly exothermic by 0.03 eV at 0 K. Under SOFC operating conditions, this process will proceed at an elevated temperature, and as a result, the activation barrier of 1.1 eV is feasible. In order to understand better the difference of the catalytic activity between the SrO and LaO
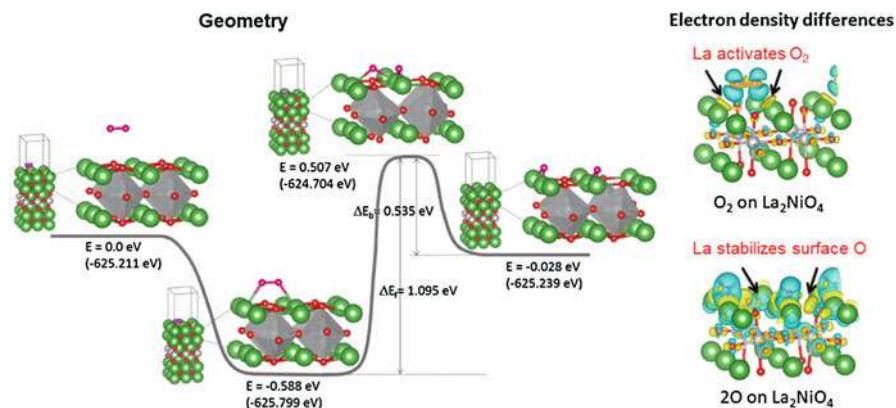
Fig. 12 Dissociation mechanism and electron density differences of oxygen molecule over the LaO-terminated surface of La_2NiO_4. (Figure adapted from Akbay et al. 2016)

surfaces, the electronic properties of La and Sr will be investigated. Bader charge analysis shows that Sr charge is close to $+2$ verifying that Sr is mostly fully ionized. However, the La charge was estimated to be also close to $+2$, which differs from its formal charge of $+3$. The reason for the deviation of the La charge from its formal charge is a partially covalent interaction between the d-levels of La and the p-levels of O. Such interaction is responsible for shared delocalized electron density between the A-sublattice and B-sublattice. As a result, the surface La is characterized with electronic density perpendicular to the surface plane (i.e., electron density similar to dangling bonds at the cleaved La-O bonds). Such electron density available on an LaO surface is responsible for the strong adsorption energy of the oxygen molecule and also stabilizes the oxo species after the dissociation of the molecular oxygen as it is summarized in Fig. 12.

The theoretical results described in Akbay et al. (2016), Staykov et al. (2015)) conclude that there is a fundamental difference in the oxygen reduction reaction mechanism on the SrO- and LaO-terminated perovskite surfaces. The reason behind this lies in the differences in electronic structure of Sr and La and, more generally, in the different electronic structure of earth alkaline elements and the lanthanoid series elements. While Sr, and in general all earth alkaline elements, contains two s elections in their valence shell which are easily ionized, La and most of the lanthanoids contain in addition d and f electrons whose ionization energy is rather different from that of s electrons. As a result, a significant deviation from the La formal charge in La_2NiO_4 was observed due to the partial covalent interaction between La and O. On the surface, it results in occupied d-levels of La which can interact with gas-phase molecular oxygen and lead to its dissociation. The lack of occupied d-levels in the valence shell of Sr leads to the inert properties of SrO toward molecular oxygen dissociation. Thus, surface oxygen vacancies in the $SrTiO_3$ lattice have been found essential for the oxygen dissociation. The dissociation mechanism includes electron transfer from the d-levels of Ti to the

oxygen antibonding π* orbital and the consequential cleavage of the oxygen-oxygen bond. A very similar mechanism is inferred for the common mixed ionic electronic conductors used extensively in high-temperature electrochemical devices. The alkaline earth dopants will provide a segregated outer surface layer which will be inert to the gas-phase molecules. The presence of surface oxygen vacancies will provide the catalytically active sites for the oxygen reduction reaction.

5   Summary

Perovskite materials have wide range of application as catalysts and electrocatalysts in energy conversion devices. Their flexibility is owing to the large number of compositions that can be synthesized, which allows for tuning the material's electronic, ionic, and catalytic properties to the desired applications.

The bulk properties, such as oxygen diffusivity, band gap, redox stability, and electronic conductivity, are governed by the bulk defect chemistry. Transition metal perovskites, such as $La_{1-x}Sr_xMnO_3$ (LSM) or $La_{(1-x)}Sr_xCoO_3$ (LSC), can often support a rich bulk defect chemistry with electronic and ionic defects. The dominant defect depends on $p(O_2)$ at a given temperature, and five defect regimes can be identified for perovskite oxides. For ionic defects, two major regimes can be distinguished: (1) the reducing regime, where oxygen vacancies dominate, and (2) the oxidizing regime, where incorporation of additional oxygen leads to the formation of metal vacancies.

The formation of oxygen vacancies is desirable for transport of oxygen through the bulk, as required by SOFC cathodes. The oxygen vacancy formation energy appears to be a good predictor for oxygen ion conductivity. While experimental methods to approximate the oxygen vacancy formation energy (such as tracer diffusion or thermogravimetric analysis) have been applied to SOFC materials for decades, computational methods are not yet routinely applied. For our example material, LSM, large scatter is present in the computational literature, and such large variability suggests limited reliability of computational studies in predicting the oxygen vacancy formation energy in this material. The observed scatter can be partially resolved by the choice of supercell size, crystallographic space group, magnetization, the Hubbard U parameter, as well as the magnitude of correction to the total energy of the O_2 molecule.

Only few perovskite materials exhibit the formation of cation vacancies under practical temperatures and $p(O_2)$. In LSM, however, cation vacancies occur in measurable concentrations under SOFC operating conditions, so that they need to be considered in reliable computational models of the SOFC cathode. Due to the uncertainty in fitting defect models to experimental TGA data, significant scatter between different literature sources can be observed. This can be rationalized by density functional theory, which suggests very strong interactions between close-by cation vacancies. As a consequence, the defect distribution cannot be assumed as random, so that the approximation of ideal mixtures usually applied when analyzing

the experimental data breaks down. Thus, more exact approaches, such as Monte Carlo simulations or the quasichemical approximation, are called for to treat cation vacancies in perovskite oxides.

The surfaces play a major role for many catalytic applications of perovskite oxides. However, the surface structure and composition adapt dynamically to the external conditions, which often leads to significant deviations from the bulk properties in the boundary regions. One major problem that arises in 3–3 perovskites like LSM and LSC is surface polarity. Dipolar surfaces are unstable due to diverging surface energy, so that electronic and ionic reconstructions frequently occur at such interfaces. For example, electronic charge can be transferred from the LSM subsurface to the surface in LSM, and defects of opposing charge segregate to the surface to compensate the surface dipole moment.

An extreme case of surface reconstruction is observed at the (001) surface of perovskite oxides. These surfaces are experimentally observed to be terminated exclusively by AO or, more specifically, by SrO in Sr-doped materials. This cation exchange between surface and subsurface essentially eliminates the surface dipole, thus improving surface stability. Also, the B-cation is more strongly destabilized by removal of oxide ions due to its highly localized point charge. The AO termination thus restores the favorable BO_6 coordination.

The dissociation of O_2 is an important step in many electrochemical and thermochemical reactions. For example, in SOFC cathodes, O_2 needs to be split and reduced in order to be incorporated into the perovskite surface. The electronic states of O_2 are distinguished in the dissociation of O_2: the neutral triplet O_2, which has diradical character, superoxide O_2^-, and peroxide O_2^{2-}. O_2 can adsorb terminally or side-on, resulting in either one or two metal-O bonds being formed upon adsorption. O_2 adsorption is considered at the AO and the BO_2 termination of perovskite oxides.

The AO termination is the stable termination of most perovskite oxides under typical SOFC conditions, so that O_2 dissociation at this surface is the main pathway for oxygen splitting and reduction at SOFC cathodes. The dissociation of O_2 on the SrO surface of $SrTiO_3$ was shown to strongly depend on the presence of double oxygen vacancies near the surface. Their role is twofold. At a surface oxygen vacancy, the underlying Ti is exposed, which gives the O_2 molecule access to the d_{z^2} orbital of Ti, which transfers electron density to the antibonding π^* orbital of O_2, thus weakening its internal bond and inducing dissociation. The second role is in incorporating the dissociated oxygen into the surface, thus completing the O_2 reduction.

In comparison to the SrO termination on $SrTiO_3$, where O_2 dissociation occurs only at oxygen vacancies, the LaO termination of Sr-free materials displays a much higher activity in O_2 reduction, as exemplified on $La_2NiO_{4+\delta}$, despite very low oxygen vacancy concentration. Computational results show that the La centers at the LaO termination are able to adsorb and split oxygen without exposed transition metal cations. This is possible because La has occupied d and f electrons, which facilitates the interaction of La with molecular O_2 and allows for O-O bond cleavage.

References

Abbes L, Noura H (2015) Perovskite oxides $MRuO_3$ (M=Sr, Ca and Ba): Structural distortion, electronic and magnetic properties with GGA and GGA-modified Becke-Johnson approaches. Results Phys. 5:38–52

Agrafiotis C, Roeb M, Sattler C (2015) A review on solar thermal syngas production via redox pair-based water/carbon dioxide splitting thermochemical cycles. Renew Sust Energ Rev 42:254–285

Akbay T, Staykov A, Druce J, Tellez H, Ishihara T, Kilner JA (2016) The interaction of molecular oxygen on LaO terminated surfaces of La_2NiO_4. J Mater Chem A 4:13113–13124

Berenov A, MacManus-Driscoll J, Kilner J (1999) Oxygen tracer diffusion in undoped lanthanum manganites. Solid State Ionics 122:41–49

Blöchl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953–17979

Burriel M, Garcia G, Santiso J, Kilner JA, Chater RJ, Skinner SJ (2008) Anisotropic oxygen diffusion properties in epitaxial thin films of $La_2NiO_{4+\delta}$. J Mater Chem 18:416–422

Burriel M, Wilkins S, Hill JP, Munoz-Marquez MA, Brongersma HH, Kilner JA, Ryan MP, Skinner SJ (2014) Absence of Ni on the outer surface of Sr doped La_2NiO_4 single crystals. Energy Environ Sci 7:311–316

Cai M-Q, Zhang Y-J, Yang G-W, Yin Z, Zhang M-S, Hu W-Y, Wang Y-G (2006) Ab initio study of structural and electronic properties of $SrTiO_3$(001) oxygen-vacancy surfaces. J Chem Phys 124:174701

Calderón M, Brey L, Guinea F (1999) Surface electronic structure and magnetic properties of doped manganites. Phys Rev B 60:6698–6704

Chen H-T, Raghunath P, Lin M (2011) Computational investigation of O_2 reduction and diffusion on 25% Sr-doped $LaMnO_3$ cathodes in solid oxide fuel cells. Langmuir 27:6787–6793

Chen Y, Jung W, Cai Z, Kim J, Tuller H, Yildiz B (2012) Impact of Sr segregation on the electronic structure and oxygen reduction activity of $SrTi_{1-x}Fe_xO_3$ surfaces. Energy Environ Sci 5:7979–7988

Chen Y, Téllez H, Burriel M, Yang F, Tsvetkov N, Cai Z, McComb DW, Kilner JA, Yildiz B (2015) Segregated Chemistry and Structure on (001) and (100) Surfaces of $(La_{1-x}Sr_x)_2CoO_4$ Override the Crystal Anisotropy in Oxygen Exchange Kinetics. Chem Mater 27:5436–5450

Cui B, Song C, Li F, Wang G, Mao H, Peng J, Zeng F, Pan F (2014) Tuning the entanglement between orbital reconstruction and charge transfer at a film surface. Sci Rep 4:4206

Curnan M, Kitchin J (2014) Effects of concentration, crystal structure, magnetism, and electronic structure method on first-principles oxygen vacancy formation energy trends in perovskites. J Phys Chem C 118:28776–28790

da Silva FS, de Souza TM (2017) Novel materials for solid oxide fuel cell technologies: a literature review. Int J Hydrog Energy 42:26020–26036

De Souza RA, Kilner JA (1998) Oxygen transport in $La_{1-x}Sr_xMn_{1-y}Co_yO_{3\pm\delta}$ perovskites: Part I. Oxygen tracer diffusion. Solid State Ionics 106:175–187

Deacon-Smith DEE, Scanlon DO, Catlow CRA, Sokol AA, Woodley SM (2014) Interlayer cation exchange stabilizes polar perovskite surfaces. Adv Mater 26:7252–7256

Druce J, Tellez H, Burriel M, Sharp M, Fawcett L, Cook S, McPhail D, Ishihara T, Brongersma H, Kilner J (2014) Surface termination and subsurface restructuring of perovskite-based solid oxide electrode materials. Energy Environ Sci 7:3539–3599

Eglitis RI (2009) Ab initio calculations of SrTiO$_3$, BaTiO$_3$, PbTiO$_3$, CaTiO$_3$ and BaZrO$_3$ (001) and (011) surfaces. Integr Ferroelectr 108:11–20

Eglitis RI (2015) Comparative ab initio calculations of SrTiO$_3$ and CaTiO$_3$ polar (111) surfaces. Phys Stat Solid B 252:635–642

Goldschmidt VM (1926) Die Gesetze der Krystallochemie. Naturwissensch 14:477–485

Han J, Jalili H, Kuru Y, Cai Z, Yildiz B (2011) Strain effects on the surface chemistry of La$_{0.7}$Sr$_{0.3}$MnO$_3$. ECS Trans 35:2097–2104

Harrison W (2011) Origin of Sr segregation at La$_{1-x}$Sr$_x$MnO$_3$ surfaces. Phys Rev B 83:155437

Ishihara T (2009) Structure and properties of perovskite oxides. In: Ishihara T (ed) Perovskite oxide for solid oxide fuel cells. (Series: Fuel cells and hydrogen energy) Springer, New York. pp 1–16

Jiang S (2008) Development of lanthanum strontium manganite perovskite cathode materials of solid oxide fuel cells: a review. J Mater Sci 43:6799–6833

Kresse G, Furthmüller J (1996a) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15–50

Kresse G, Furthmüller J (1996b) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186

Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. Phys Rev B 47:558–561

Kuklja M, Kotomin E, Merkle R, Mastrikov Y, Maier J (2013) Combined theoretical and experimental analysis of processes determining cathode performance in solid oxide fuel cells. Phys Chem Chem Phys 15:5443–5471

Kuo J, Anderson H, Sparlin DJ (1989) Oxidation-reduction behavior of undoped and Sr-doped LaMnO$_3$ nonstoichiometry and defect structure. Sol State Chem 83:52–60

Kwon H, Lee W, Han J (2016) Suppressing cation segregation on lanthanum- based perovskite oxides to enhance the stability of solid oxide fuel cell cathodes. RSC Adv 6:69782

Lee W, Han J, Chen Y, Cai Z, Yildiz B (2013) Cation size mismatch and charge interactions drive dopant segregation at the surfaces of manganite perovskites. J Am Chem Soc 135:7909–7925

Lee Y-L, Morgan D (2012) Ab initio and empirical defect modeling of LaMnO$_{3\pm\delta}$ for solid oxide fuel cell cathodes. Phys Chem Chem Phys 14:290–302

Lee Y-L, Morgan D (2015) Ab initio defect energetics of perovskite (001) surfaces for solid oxide fuel cells: A comparative study of LaMnO$_3$ versus SrTiO$_3$ and LaAlO$_3$. Phys Rev B 91:195430

Lee Y-L, Kleis J, Rossmeisl J, Morgan D (2009) Ab initio energetics of LaBO$_3$(001) (B=Mn, Fe, Co, and Ni) for solid oxide fuel cell cathodes. Phys Rev B 80:224101

Lee Y-L, Gadre MJ, Shao-Horn Y, Morgan D (2015) Ab initio GGA+U study of oxygen evolution and oxygen reduction electrocatalysis on the (001) surfaces of lanthanum transition metal perovskites LaBO$_3$ (B = Cr, Mn, Fe, Co and Ni). Phys Chem Chem Phys 17:21643–21663

Li M, Zhang H, Cook SN, Li L, Kilner JA, Reaney IM, Sinclair DC (2015) Dramatic influence of A-site nonstoichiometry on the electrical conductivity and conduction mechanisms in the perovskite oxide Na$_{0.5}$Bi$_{0.5}$TiO$_3$. Chem Mater 27:629–634

Li Y, Zhang W, Zheng Y, Chen J, Yu B, Chen Y, Liu M (2017) Controlling cation segregation in perovskite-based electrodes for high electro-catalytic activity and durability. Chem Soc Rev 46:6345–6378

Mahato N, Banerjee A, Gupta A, Omar S, Balani K (2015) Progress in material selection for solid oxide fuel cell technology: a review. Prog Mater Sci 72:141–337

Mastrikov YA, Merkle R, Kotomin EA, Kuklja MM, Maier J (2018) Surface termination effects on the oxygen reduction reaction rate at fuel cell cathodes. J Mater Chem A 6:11929

Mebane D, Liu Y, Liu M (2008) Refinement of the bulk defect model for La$_x$Sr$_{1-x}$MnO$_{3\pm\delta}$. Solid State Ionics 178:1950–1957

Mizusaki J, Mori N, Takai H, Yonemura Y, Minamiue H, Tagawa H, Dokiya M, Inaba H, Naraya K, Sasamoto T, Hashimoto T (2000) Oxygen nonstoichiometry and defect equilibrium in the perovskite-type oxides La$_{1-x}$Sr$_x$MnO$_{3+d}$. Solid State Ionics 129:163–177

Nakagawa N, Hwang H, Muller D (2006) Why some interfaces cannot be sharp. Nat Mater 5: 204–209

Nowotny J, Rekas MJ (1998) Defect Chemistry of (La,Sr)MnO$_3$. Am Ceram Soc 81:67–80

Olsson E, Aparicio-Anglès X, de Leeuw N (2016) Ab initio study of vacancy formation in cubic LaMnO$_3$ and SmCoO$_3$ as cathode materials in solid oxide fuel cells. J Chem Phys 145:014703

Pesquera D, Herranz G, Barla A, Pellegrin E, Bondino F, Magnano E, Sanchez F, Fontcuberta J (2012) Surface symmetry-breaking and strain effects on orbital occupancy in transition metal perovskite epitaxial films. Nat Commun 3:1189

Petrov A, Zuev A, Tichinova I, Voronin V (2000) Crystal and defect structure of the mixed oxides LaMn$_{1-z}$Cu$_z$O$_{3\pm y}$ (0≤z≤0.4) Solid State Ionics 129:179–188

Piskunov S, Heifets E, Jacob T, Kotomin E, Ellis D, Spohr E (2008) Electronic structure and thermodynamic stability of LaMnO$_3$ and La$_{1-x}$Sr$_x$MnO$_3$ (001) surfaces: Ab initio calculations. Phys Rev B 78:121406

Pojani A, Finocchi F, Noguera C (1999) Polarity on the SrTiO$_3$ (111) and (110) surfaces. Surf Sci 442:179–198

Rao CNR, Dey SJ (2016) Generation of H$_2$ and CO by solar thermochemical splitting of H$_2$O and CO$_2$ by employing metal oxides. Sol State Chem 242:107–115

van Roosmalen J, Cordfunke E (1994) The defect chemistry of LaMnO$_{3\pm\delta}$. J Solid State Chem 110:113–117

Rupp GM, Opitz AK, Nenning A, Limbeck A, Fleig J (2017) Real-time impedance monitoring of oxygen reduction during surface modification of thin film cathodes . Nat Mater 16:640–645

Scheffe JR, Steinfeld A (2014) Oxygen exchange materials for solar thermochemical splitting of H$_2$O and CO$_2$: a review. Mater Today 7:341–348

Setvin M, Reticcioli M, Poelzleitner F, Hulva J, Schmid M, Boatner LA, Franchini C, Diebold U (2018) Polarity compensation mechanisms on the perovskite surface KTaO$_3$(001). Science 359:572–575

Sharma V, Mahapatra M, Singh P, Ramprasad R (2015) Cationic surface segregation in doped LaMnO$_3$. J Mater Sci 50:3051–3056

Skinner SJ, Kilner JA (2000) Oxygen diffusion and surface exchange in La$_{2-x}$Sr$_x$NiO$_{4+\delta}$. Solid State Ionics 135:709–712

Souza RD, Kilner J, Walker J (2000) A SIMS study of oxygen tracer diffusion and surface exchange in La$_{0.8}$Sr$_{0.2}$MnO$_{3+\delta}$. Mater Lett 43:43–52

Staykov A, Kamachi T, Ishihara T, Yoshizawa K (2008) Theoretical study of the direct synthesis of H$_2$O$_2$ on Pd and Pd/Au surfaces. J Phys Chem C 112:19501–19505

Staykov A, Téllez H, Akbay T, Druce J, Ishihara T, Kilner J (2015) Oxygen activation and dissociation on transition metal free perovskite surfaces. Chem Mater 27:8273–8281

Staykov A, Tellez H, Druce J, Wu J, Ishihara T, Kilner J (2018) Electronic properties and surface reactivity of SrO-terminated SrTiO$_3$ and SrO-terminated iron-doped SrTiO$_3$. Sci Technol Adv Mater 19:221–230

Takacs M, Hoes M, Caduff M, Cooper T, Scheffe J, Steinfeld A (2016) Oxygen nonstoichiometry, defect equilibria, and thermodynamic characterization of LaMnO$_3$ perovskites with Ca/Sr A-site and Al B-site doping. Acta Mater 103:700–710

Tarancon A, Skinner SJ, Chater RJ, Hernandez-Ramirez F, Kilner JA (2007) Layered perovskites as promising cathodes for intermediate temperature solid oxide fuel cells. J Mater Chem 17:3175–3181

Tasker PJ (1979) The stability of ionic crystal surfaces. J Phys C Solid State Phys 12:4977–4984

Téllez H, Druce J, Hong J-E, Ishihara T, Kilner JA (2015) Accurate and precise measurement of oxygen isotopic fractions and diffusion profiles by selective attenuation of secondary ions (SASI). Anal Chem 87:2907–2915

Tokura Y, Tomioka Y (1999) Colossal magnetoresistive manganites. J Magn Magn Mater 200:1–23

Tsvetkov N, Lu Q, Sun L, Crumlin EJ, Yildiz B (2016) Improved chemical and electrochemical stability of perovskite oxides with less reducible cations at the surface. Nat Mater 6:1010–1016

Wang L, Maxisch T, Ceder G (2006) Oxidation energies of transition metal oxides within the GGA+U framework. Phys Rev B 73:195107

Wang W, Jiang S (2006) A mechanistic study on the activation process of (La, Sr)MnO$_3$ electrodes of solid oxide fuel cells. Solid State Ionics 177:1361–1369

Wu W-Q, Chen D, Caruso RA, Cheng Y-B (2017) Recent progress in hybrid perovskite solar cells based on n-type materials. J Mater Chem A 5:10092–10109

Yagi S, Yamada I, Tsukasaki H, Seno A, Murakami M, Fujii H, Chen H, Umezawa N, Abe H, Nishiyama N, Mori S (2015) Covalency-reinforced oxygen evolution reaction catalyst. Nature Comm 6:8249

Yasuda I, Hishinuma M (1996) Electrical conductivity and chemical diffusion coefficient of strontium-doped lanthanum manganites. J Solid State Chem 123:382–390

Yasuda I, Ogasawara K, Hishinuma M, Kawada T, Dokiya M (1996) Oxygen tracer diffusion coefficient of (La, Sr)MnO$_{3\pm\delta}$. Solid State Ionics 86-88:1197–1201

Zenia H, Gehring G, Banach G, Temmerman W (2005) Electronic and magnetic properties of the (001) surface of hole-doped manganites. Phys Rev B 71:024416

Zhou J, Chen G, Wu K, Cheng Y (2013) Interaction of La$_2$NiO$_4$ (100) Surface with oxygen molecule: a first-principles study. J Phys Chem C 117:12991–12999

Zuev A, Tsvetkov D (2010) Oxygen nonstoichiometry, defect structure and defect-induced expansion of undoped perovskite LaMnO$_{3\pm\delta}$. Solid State Ionics 181:557–563

Part VIII
Surface Catalysis

A Decade of Computational Surface Catalysis

53

Karsten Reuter and Horia Metiu

Contents

Abstract

We briefly survey recent developments in surface catalysis modeling. The differentiated view on required level of accuracy established in the wake of multi-scale modeling approaches led to the emergence of high-throughput computational screening approaches. The large amounts of data created this way are now increasingly mined with machine learning techniques. We discuss status and challenges in this exciting mix of methodologies that describe catalytic systems from the electrons to the reactor. Next to the traditional focus on understanding and predicting catalytic activity, we argue that approaches to dynamical catalyst restructuring, to concomitant heat management, and to catalyst lifetime are important themes for the decade to come.

K. Reuter (✉)
Chair for Theoretical Chemistry and Catalysis Research Center, Technical University of Munich, Garching, Germany
e-mail: karsten.reuter@ch.tum.de

H. Metiu
Department of Chemistry and Biochemistry, University of California at Santa Barbara, Santa Barbara, CA, USA
e-mail: metiu@chem.ucsb.edu

1   Introduction

When the first edition of the *Handbook of Materials Modeling* appeared in 2005 (Yip 2005), many chapters featured surface catalysis as an important application area. However, compared to the situation we have reached now, a bit over a decade later, this research was much more disjunct, focused on individual methods at their corresponding length and time scales. Multi-scale modeling was in its early days but then took off to become a lead theme in catalysis modeling. In bottom-up approaches, electronic structure theorists tried to extend the predictive-quality power of first-principles techniques to longer length and time scales, advancing ab initio thermodynamics and microkinetic models to capture effects of the reactive environment or ensemble effects arising in the interplay of elementary processes in the catalytic cycle. In top-down approaches, chemical engineers aimed at introducing a more detailed chemistry into traditional continuum descriptions of the macroscopic heat and mass transport in reactor models. Today, the gap between both types of approaches is still not closed, but it has narrowed substantially, forging an interdisciplinary community that is used to tackle the multi-physics challenges of surface catalysis with multi-method approaches.

Perhaps the most important implication of this new paradigm was a more differentiated view on accuracy and required level of detail. For a long time, consistent chemical accuracy had been seen as a prerequisite for predictive quality in first-principles calculations. More coarse-grained models were instead frequently motivated by the perception that microscopic details (and concomitant errors in their description) would in some way average out. Increasingly established sensitivity analyses in multi-scale modeling approaches scrutinized these views and revealed that it is in fact often a few microscopic parameters that critically control targeted macroscopic functionalities like activity or selectivities (Campbell 2017; Döpking et al. 2018; Medford et al. 2014; Meskine et al. 2009; Sutton et al. 2016). This laid the foundation for high-throughput computational screening approaches that focus on these microscopic parameters (then called descriptors), compute them for a wide range of materials, and thereby allow to search vast materials spaces for promising catalysts (Andersen et al. 2016, 2017a; Andersson et al. 2006; Besenbacher et al. 1998; Greeley and Mavrikakis 2004; Greeley et al. 2006; Linic et al. 2004; Loffreda et al. 2009; 2014; Toulhoat and Raybaud 2003). The large amounts of data created this way are now increasingly mined with machine learning techniques. This provides further insight into correlations between materials properties and (catalytic) function and carries the predictive first-principles power over to the in silico design of improved systems (Jinnouchi and Asahi 2017; Li et al. 2017; Ma et al. 2015; Ulissi et al. 2017a, b).

The emerging methodological triangle of detailed multi-scale modeling, high-throughput screening, and data sciences converts contemporary computational catalysis research into a powerhouse of possibilities and exciting developments. Many of these are driven by the urgent need to switch to alternative energy sources and establish their conversion through sustainable energy technologies. Among others, this has led to a strong boost in electrocatalysis modeling. Tackling the challenges

connected with the electrified solid-liquid interface will likely continue to be a major theme for the next decade. Another such theme will be the complexity gap between structural models presently feasible with atomistic approaches and real high-surface area catalysts in complex reactor geometries, connected with a description of catalyst degradation. In the following we will briefly survey some of these developments and challenges that are then discussed in more detail in the following chapters in this section. What will already become clear from this short overview is that the next decade of catalysis modeling will be at least as fast paced as was the past. In this respect, it is tantalizing to imagine what could be the content of the surface catalysis section in a future third edition of the *Handbook of Materials Modeling*.

2   Methods in Surface Catalysis Modeling

With its inherent focus on support through computational spectroscopy and a description of the reactive surface chemistry, first-principles electronic structure theory has always been a natural basis for any modeling in catalysis research. Density functional theory (DFT) and its ever more sophisticated and powerful computer packages like VASP (Kresse and Furthmüller 1996a, b) and others are presently the de facto standard for extended systems. For metallic catalysts and in lack of any even conceptual avenues for systematic improvements, low-rung generalized-gradient approximation (GGA) functionals are still prevalent and probably will be for quite a while (Burke 2012; Cohen et al. 2012; Feibelman 2010). The concomitant uncertainty in key energetic parameters like reaction barriers will therefore continue to be a crucial issue, in particular, when considering that computed reaction barriers (and their errors) enter exponentially into kinetic rate constants. For gap systems like most oxides, simple +U corrections to GGA functionals are still a minimalistic state of the art in more applied studies (Himmetoglu et al. 2014; Kulik 2015). Admixture of exact exchange in hybrid, screened hybrid, or diagrammatic functionals offers an alternative route for systematic improvements for these systems that sees increasing usage also in surface catalysis research (Goerigk and Grimme 2011; Pacchioni 2008; Ren et al. 2012; Schimka et al. 2010). The possibility to efficiently account for dispersive interactions either through van der Waals (vdW) correction schemes or vdW functionals has been of unparalleled value for all types of systems and has certainly opened the door to also treat larger reaction intermediates (Klimeš and Michaelides 2012; Liu et al. 2014; Maurer et al. 2016), e.g., in the context of Fischer-Tropsch catalysis. Solid-state embedding approaches concentrating the quantum mechanical calculations to a finite core region around a considered active site also begin to enable more extensive use of correlated wave-function-based electronic structure theory calculations, not least providing highly accurate reference values (Boese and Sauer 2013; Kubas et al. 2016; Libisch et al. 2014; Richter et al. 2013; Stodt et al. 2013).

Next to such concurrent multi-scale modeling techniques, it is in particular hierarchical multi-scale modeling approaches that are now largely employed in surface catalysis research and that have led to qualitative advances in comparison to

the situation a decade ago. In bottom-up approaches, quantities calculated by first-principles techniques enter into more coarse-grained models that thereby allow to address longer time scales and larger length scales. A most widespread approach is ab initio thermodynamics, in which first-principles calculated total energies or vibrational frequencies are used to determine thermodynamic potentials like Gibbs free energies (Reuter 2016). Representing a surrounding gas or liquid environment in the form of chemical reservoirs to or from which species or charges are given or taken, this approach yields, for instance, (approximate) insight into the surface structure and composition of the catalyst surface in the reactive environment (in heterogeneous catalysis denoted as surface phase diagrams, in electrocatalysis denoted as Pourbaix diagrams) and of nanoparticle shape during reaction conditions (Garcia-Mota et al. 2015; Ouyang et al. 2013; Wang et al. 2013), or it provides a first assessment of thermodynamic barriers along an assumed reaction path, as, for instance, discussed in ▶ Chap. 56, "Supercell Models of Brønsted and Lewis Sites in Zeolites," by Li and Schneider in this section.

For surface electrocatalysis, the breakthrough for such models came when Nørskov and coworkers realized that the (otherwise computationally hard to determine) chemical potentials of solvated species can be related to their gas-phase chemical potentials at standard conditions (Nørskov et al. 2004, 2005). This computational hydrogen electrode (CHE) or equivalent references for other species then allows to quickly obtain estimates of overpotentials and the rate-limiting steps causing them. In the original formulation, solvation effects on the surface energetics were not treated at the electronic structure level, leading among others to deficiencies in the description of pH effects. Efficient treatment of solvation effects is therefore a major topic in current research, with implicit solvation models (Andreussi and Marzari 2014; Mathew et al. 2014; Ringe et al. 2016; Sinstein et al. 2017), static ice layers (Chan and Nørskov 2015; Rossmeisl et al. 2008; Skúlason et al. 2010), or full explicit dynamical simulations (Chen et al. 2013; Mattioli et al. 2013; Quaranta et al. 2017; Stecher et al. 2016), representing three major strands of activities as discussed in more detail in ▶ Chaps. 60, "Fundamental Atomic Insight in Electrocatalysis," ▶ 59, "Structure of Electrode-Electrolyte Interfaces, Modeling of Double Layer and Electrode Potential," and ▶ 61, "Electrocatalysis Beyond the Computational Hydrogen Electrode," by Bagger et al., Groß, and Oberhofer, respectively, in this part.

The explicit calculation of reaction barriers is a first step to account for kinetic effects (Sabbe et al. 2012). In heterogeneous catalysis, corresponding calculations are predominantly carried out using transition state search algorithms like the nudged elastic band (Henkelman et al. 2000). While nominally a "plug-and-play" functionality in almost all major computer packages, in practice every student working with it still struggles in practice with poor convergence and large computational costs. For reactions at solid-liquid interfaces, the situation is even worse, as in principle even more demanding dynamical or free energy simulations are required (Chen et al. 2013; Groß et al. 2014; Li et al. 2014; Mattioli et al. 2013; Pinto et al. 2014; Schnur and Groß 2011; Quaranta et al. 2017; Stecher et al. 2016). Further complications arise presently in this electrocatalytic

context, as standard first-principles supercell approaches for extended systems are for constant charge and not for constant potential (Bonnet et al. 2012; Chan and Nørskov 2015; Filhol and Bocquet 2007; Otani and Sugino 2006; Rossmeisl et al. 2008, 2013; Surendrala et al. 2018), as discussed in detail by Groß in ▶ Chap. 59, "Structure of Electrode-Electrolyte Interfaces, Modeling of Double Layer and Electrode Potential." Research toward numerically efficient approaches allowing the reliable (quantitative or at least semiquantitative but with controlled error bars) calculation of reaction barriers is therefore a hot topic that will likely even increase in importance in the next years. In this respect, the quantification of Brønsted-Evans-Polanyi (BEP) relationships pioneered within the more general context of so-called scaling relations by Nørskov and coworkers has been a true game changer in surface catalysis research in the last decade (Greeley 2016; Nørskov et al. 2008, 2014). Through BEP relationships approximate barriers can be derived from much less demanding thermochemical calculations, while other scaling relations allow to derive corresponding energetic parameters for more complex reaction intermediates from data for their constituting elements, or they allow to extrapolate the energetics over different catalyst materials; see ▶ Chap. 54, "Energy Trends in Adsorption at Surfaces," by Raman and Vojvodic for more details.

The increasing availability of such (approximate) first-principles reaction energetics for elementary processes has strongly fueled the development of microkinetic simulations that evaluate the interplay of these elementary processes in the reaction network and yield the intrinsic catalytic activities (and increasingly also selectivities) as a function of applied reaction conditions (Sabbe et al. 2012; Reuter 2016; Sutton and Vlachos 2015). While mean-field microkinetic models in the form of rate equations still dominate the field (Nørskov et al. 2014; van Santen and Neurock 2006), cf. ▶ Chap. 57, "Microkinetic Modeling of Surface Catalysis" by Vlachos and Wittreich, more accurate kinetic Monte Carlo (kMC) simulations that account for the spatial distribution of reaction intermediates at the catalyst surface are quickly catching up (Stamatakis and Vlachos 2012; Reuter 2013, 2016). Especially when exploiting scaling-relation-based larger energetic databases, either type of microkinetic simulations is successfully and increasingly used to also address more complex reaction networks, for instance, leading to synthetic fuels. kMC simulations are thereby often still challenged by networks featuring very fast processes (prototypically diffusion at metal catalysts). This starts to be addressed in emerging accelerated kMC approaches (Andersen et al. 2017b; Chatterjee and Voter 2010; Dybeck et al. 2017).

A particularly intriguing development within the realm of microkinetic simulations is the realization that through the use of scaling and BEP relations, the dimensionality of the problem can be dramatically reduced. While in principle any given microkinetic model depends on all rate constants (aka reaction barriers) of all elementary processes, the complexity of these dependencies decreases significantly when exploiting that groups of rate constants all scale with one another. When furthermore making (in parts strong) assumptions about the reaction mechanism and rate-limiting steps, the overall dependencies of a microkinetic model can be

minimized (Nørskov et al. 2014). Often the catalytic activity (or selectivity) can then be discussed as a function of the rate constants of just one or two elementary processes and through BEP and scaling relations as a function of the adsorption energies of just one or two elemental reaction intermediates like O or C. This strand of catalysis research has substantially increased our understanding of trends in the past decade and is nowadays largely employed to "screen" for better catalysts by simply computing corresponding descriptors like the O or C binding energies for a multitude of catalyst materials and seeing whether these energies fall into a range yielding optimum activity according to the simplified microkinetic model (Andersen et al. 2016, 2017a; Besenbacher et al. 1998; Greeley and Mavrikakis 2004; Greeley et al. 2006; Linic et al. 2004; Loffreda et al. 2009; Nørskov 2014; Toulhoat and Raybaud 2003). An exciting new development in this context is the fusing with data sciences, using machine learning techniques to identify better descriptors or to extend and refine the energetic databases through regression techniques (Jinnouchi and Asahi 2017; Li et al. 2017; Ma et al. 2015; Quaranta et al. 2017; Stecher et al. 2016; Ulissi et al. 2017a, b).

To also account for macroscopic heat and mass transport effects in a given reactor, microkinetic models are integrated into continuum fluid dynamical simulations. This is more established for mean-field rate equation models (Janardhanan and Deutschmann 2012; Maestri and Cuoci 2013) but has recently also been extended to kMC-based microkinetic models (Matera and Reuter 2009, 2010; Matera et al. 2014). Such coupled simulations help in particular to analyze data from increasingly employed in situ and *operando* measurements, which are often heat and mass transport limited and where then the normal intrinsic activity provided by a microkinetic model cannot be used to relate to the experimental data (Sabbe et al. 2012; Reuter 2016). To date, the integration of first-principles based microkinetic models (either mean-field or kMC) into fluid dynamical simulations at reactor level is, however, limited to simple model catalysts, in fact essentially to single crystals. The reactive chemistry in more complex high-surface area real catalysts is instead described through effective kinetic models as discussed in ▸ Chap. 58, "Computational Fluid Dynamics of Catalytic Reactors," by Hettel, Wörner, and Deutschmann. For such structured catalysts, explicit or effective treatments of porosity, the emergence of turbulence, or complex reactor approaches like fluidized beds are instead challenging topics in fluid dynamics research itself.

3   Open Challenges

Due to the impressive methodological advances briefly highlighted above, modeling and simulation makes substantial contributions to surface catalysis research at all levels from the active site to the reactor. In fact, modeling and simulation is by now firmly integrated as an indispensable part of essentially any project, even in industrial settings. Despite these advances, modeling is still severely challenged by the complexity of real catalytic systems. The development of suitable model systems that still contain the essential features at largely reduced complexity remains one of

the core issues and tasks for the years to come (Sauer and Freund 2015). At the level of the active sites, prevalent approaches often focus on singular high-symmetry sites at individual facets of crystalline nanoparticles. This does not adequately address the complexity in particular of compound materials, which offer multiple such sites – with often little insight available which of these sites dominate the catalytic activity (Reuter et al. 2017). Furthermore, point or extended defects can generally play a crucial role, which in semiconducting materials can additionally exist in multiple charge states. *Operando* studies increasingly report strong morphological changes of the active particles, which severely question the prevalent approach to base any type of active site models simply on structures derived from cutting the crystal lattice of the nominal bulk catalyst material (Reuter 2016). There is also increasing evidence of a more dynamical nature of operating catalysts, suggesting continuous structural changes, maybe even reminiscent of a near liquid state (Newton 2008; Reuter 2016; Schlögl 2015). At present, such dynamics cannot be captured by prevalent microkinetic models, in particular not by generally still lattice-based kMC models (Reuter 2013). That such dynamics has not been "seen" in corresponding modeling is therefore no proof against its existence but simply a shortcoming of our present methodological capabilities.

A second core issue is that (in particular first-principles-based academic) computational catalysis research is still largely focused on catalytic activity as singular measure for a good catalyst. This neglects that in practice a catalyst is part of a complex system and one cannot evaluate a catalyst without considering the whole system. The catalyst needs to contain cheap materials, it needs to be cheap to make, it needs to tolerate contaminants in the feed, it needs to last for a long time, and it needs to avoid elaborate separation of the products. In addition, it must allow for inexpensive heat management. While some computational studies start to consider material cost, for instance, searching for pareto-optimal solutions in high-throughput screening using cost as additional descriptor (Andersson et al. 2006; Nørskov et al. 2009), this does not yet address the actual costs of making a catalyst out of the material. Recall that the reason why, for example, ruthenium still prevails over iron as industrial ammonia synthesis catalyst are the high costs of making an iron catalyst out of cheap iron (Ertl 1980). Computational studies have hitherto also not really aimed at evaluating the effect of trace impurities in the stream at predictive-quality level, neither in detailed microkinetic models nor by considering an optimized interplay of catalyst and contaminant in computational screening. "Poisoning" by impurities is thereby not always detrimental. Sometimes it is possible to poison selectively the surface sites on which an undesired reaction takes place. In partial oxidation reactions, one often tries to poison the surface sites on which complete oxidation occurs. In either case, the absolute effect can be enormous. In ammonia fabrication, the catalytic reactor accounts for about 1% of the cost of a plant. The main expense is generating clean hydrogen that is devoid of CO or sulfur contaminants. A radical improvement for this process would be finding a catalyst that is immune to impurities in the feed. Such a catalyst would prevail in practice, even if the reaction rate is smaller than that of the existing catalyst. The point is that high rate is a good criterion but in many cases it is

not the dominant factor. Similar examples and arguments could be made about the importance of heat management and catalyst degradation. The last decade of computational catalysis modeling is primarily characterized by establishing a comprehensive trend understanding of catalytic activity in terms of underlying materials properties. Compared to this trend understanding, our present insight, for instance, into microscopic heat dissipation channels (Rittmeyer et al. 2018), their role in, as well as sintering or other aging processes in general is essentially nonexistent. In this respect, catalyst lifetime, catalyst tolerance to contaminants, and embracing a possibly dynamical state of catalysts are certainly worthwhile themes for the decade to come.

References

Andersen M, Medford AJ, Nørskov JK, Reuter K (2016) Analyzing the case for bifunctional catalysis. Angew Chem Int Ed 55:5210–5214

Andersen M, Medford AJ, Nørskov JK, Reuter K (2017a) Scaling-relation based analysis of bifunctional catalysis: the case for well-mixed bimetallic alloys. ACS Catal 7:3960–3967

Andersen M, Plaisance CP, Reuter K (2017b) Assessment of mean-field microkinetic models for CO methanation on stepped metal surfaces using accelerated kinetic Monte Carlo. J Chem Phys 147:152705

Andersson M, Bligaard T, Kustov A, Larsen KE, Greeley JP, Johannessen T, Christensen CH, Nørskov JK (2006) Toward computational screening in heterogeneous catalysis: Pareto-optimal methanation catalysts. J Catal 239:501–506

Andreussi O, Marzari N (2014) Electrostatics of solvated systems in periodic boundary conditions. Phys Rev B 90:24510

Besenbacher F, Chorkendorff I, Clausen BS, Hammer B, Molenbroek AM, Nørskov JK, Stensgaard I (1998) Design of a surface alloy catalyst for steam reforming. Science 279:1913–1915

Boese AD, Sauer J (2013) Accurate adsorption energies of small molecules on oxide surfaces: CO-MgO(001). Phys Chem Chem Phys 15:16481–16493

Bonnet N, Morishita T, Sugino O, Otani M (2012) First-principles molecular dynamics at a constant electrode potential. Phys Rev Lett 109:266101

Burke K (2012) Perspective on density functional theory. J Chem Phys 136:150901

Campbell CT (2017) The degree of rate control: a powerful tool for catalysis research. ACS Catal 7:2770–2779

Chan K, Nørskov JK (2015) Electrochemical barriers made simple. J Phys Chem Lett 6:2663–2668

Chatterjee A, Voter AF (2010) Accurate acceleration of kinetic Monte Carlo simulations through the modification of rate constants. J Chem Phys 132:194101

Chen J, Li YF, Sit P, Selloni A (2013) Chemical dynamics of the first proton-coupled electron transfer of water oxidation on TiO$_2$ anatase. J Am Chem Soc 135:18774–18777

Cohen AJ, Mori-Sanchez P, Yang WT (2012) Challenges for density functional theory. Chem Rev 112:289–320

Döpking S, Plaisance CP, Strobusch D, Reuter K, Scheurer C, Matera S (2018) Addressing global uncertainty and sensitivity in first-principles based microkinetic models by an adaptive sparse grid approach. J Chem Phys 148:034102

Dybeck EC, Plaisance CP, Neurock M (2017) Generalized temporal acceleration scheme for kinetic Monte Carlo simulations of surface catalytic processes by scaling the rates of fast reactions. J Chem Theory Comput 13:1525–1538

Ertl G (1980) Surface science and catalysis studies on the mechanism of ammonia synthesis: the P.H. Emmett award address. Catal Rev Sci Eng 21:201–223

Feibelman PJ (2010) DFT versus the "real world" (or, waiting for godft). Top Catal 53:417–422

Filhol JS, Bocquet ML (2007) Charge control of the water monolayer/pd interface. Chem Phys Lett 438:203–207

Garcia-Mota M, Rieger M, Reuter K (2015) Ab initio prediction of the equilibrium shape of supported Ag nanoparticles on α-Al₂O₃(0001). J Catal 321:1–6

Goerigk L, Grimme S (2011) A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. Phys Chem Chem Phys 13:6670–6688

Greeley J (2016) Theoretical heterogeneous catalysis: scaling relationships and computational catalyst design. Ann Rev Chem Biomol Eng 7:605–635

Greeley J, Mavrikakis M (2004) Alloy catalysts designed from first principles. Nat Mat 3:810–815

Greeley J, Jaramillo TF, Bonde J, Chorkendorff I, Nørskov JK (2006) Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. Nat Mat 5: 909–913

Groß A, Gossenberger F, Lin X, Naderian M, Sakong S, Roman T (2014) Water structures at metal electrodes studied by ab initio molecular dynamics simulations. J Electrochem Soc 161: E3015–E3020

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113:9901–9910

Himmetoglu B, Floris A, de Gironcoli S, Cococcioni M (2014) Hubbard-corrected DFT energy functionals: the LDA+U description of correlated systems. Int J Quantum Chem 114:14–49

Janardhanan VM, Deutschmann O (2012) Computational fluid dynamics of catalytic reactors. In: Deutschmann O (ed) Modelling and simulation of heterogeneous catalytic reactions: from the molecular process to the technical system. Wiley-VCH, Weinheim. pp 251–282

Jinnouchi R, Asahi R (2017) Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm. J Phys Chem Lett 8:4279–4283

Klimeš J, Michaelides A (2012) Perspective: advances and challenges in treating van der Waals dispersion forces in density functional theory. J Chem Phys 137:120901

Kresse G, Furthmüller J (1996a) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169

Kresse G, Furthmüller J (1996b) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15–50

Kubas A, Berger D, Oberhofer H, Maganas D, Reuter K, Neese F (2016) Surface adsorption energetics studied with "gold standard" wavefunction based ab initio methods: small molecule binding to TiO₂(110). J Phys Chem Lett 7:4207–4212

Kulik HJ (2015) Perspective: treating electron over-delocalization with the DFT+ U method. J Chem Phys 142:240901

Li P, Henkelman G, Keith JA, Johnson JK (2014) Elucidation of aqueous solvent-mediated hydrogen-transfer reactions by ab initio molecular dynamics and nudged elastic-band studies of NaBH₄ hydrolysis. J Phys Chem C 118:21385–21399

Li Z, Ma X, Xin H (2017) Feature engineering of machine-learning chemisorption models for catalyst design. Catal Today 280:232–238

Libisch F, Huang C, Carter EA (2014) Embedded correlated wavefunction schemes: theory and applications. Acc Chem Res 47:2768–2775

Linic S, Jankowiak J, Barteau MA (2004) Selectivity driven design of bimetallic ethylene epoxidation catalysts from first principles. J Catal 224:489–493

Liu W, Tkatchenko A, Scheffler M (2014) Modeling adsorption and reactions of organic molecules at metal surfaces. Acc Chem Res 47:3369–3377

Loffreda D, Delbecq F, Vigne F, Sautet P (2009) Fast prediction of selectivity in heterogeneous catalysis from extended Brønsted-Evans-Polanyi relations: a theoretical insight. Angew Chem Int Ed 48:8978–8980

Ma X, Li Z, Achenie LEK, Xin H (2015) Machine-learning-augmented chemisorption model for CO₂ electroreduction catalyst screening. J Phys Chem Lett 6:3528–3533

Maestri M, Cuoci A (2013) Coupling CFD with detailed microkinetic modeling in heterogeneous catalysis. Chem Eng Sci 96:106–117

Matera S, Reuter K (2009) First-principles approach to heat and mass transfer effects in model catalyst studies. Catal Lett 133:156–159

Matera S, Reuter K (2010) Transport limitations and bistability for in situ CO oxidation at RuO$_2$(110): first-principles based multi-scale modeling. Phys Rev B 82:085446

Matera S, Maestri M, Cuoci A, Reuter K (2014) Predictive-quality surface reaction chemistry in real reactor models: integrating first-principles kinetic Monte Carlo simulations into computational fluid dynamics. ACS Catal 4:4081–4092

Mathew K, Sundararaman R, Letchworth-Weaver K, Arias TA, Hennig RG (2014) Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways. J Chem Phys 140:084106

Mattioli G, Giannozzi P, Amore Bonapasta A, Guidoni L (2013) Reaction pathways for oxygen evolution promoted by cobalt catalyst. J Am Chem Soc 135:15353–15363

Maurer RJ, Ruiz VG, Camarillo-Cisneros J, Liu W, Ferri N, Reuter K, Tkatchenko A (2016) Adsorption structures and energetics of molecules on metal surfaces: bridging experiment and theory. Prog Surf Sci 91:72–100

Medford AJ, Wellendorff J, Vojvodic A, Studt F, Abild-Pedersen F, Jacobsen KW, Bligaard T, Nørskov JK (2014) Assessing the reliability of calculated catalytic ammonia synthesis rates. Science 345:197–200

Meskine H, Matera S, Scheffler M, Reuter K, Metiu H (2009) Examination of the concept of degree of rate control by first-principles kinetic Monte Carlo simulations. Surf Sci 603:1724–1730

Newton MA (2008) Dynamic adsorbate/reaction induced structural change of supported metal nanoparticles: heterogeneous catalysis and beyond. Chem Soc Rev 37:2644–2657

Nørskov JK, Rossmeisl J, Logadottir A, Lindqvist L, Kitchin JR, Bligaard T, Jonsson H (2004) Origin of the overpotential for oxygen reduction at a fuel-cell cathode. J Phys Chem B 108:17886–17892

Nørskov JK, Bligaard T, Logadottir A, Kitchin JR, Chen JG, Pandelov S, Stimming U (2005) Trends in the exchange current for hydrogen evolution. J Electrochem Soc 152:J23–J26

Nørskov JK, Bligaard T, Hvolbæk B, Abild-Pedersen F, Chorkendorff I, Christensen CH (2008) The nature of the active site in heterogeneous metal catalysis. Chem Soc Rev 37:2163–2171

Nørskov JK, Bligaard T, Rossmeisl J, Christensen CH (2009) Towards the computational design of solid catalysts. Nat Chem 1:37–46

Nørskov JK, Studt F, Abild-Pedersen F, Bligaard T (2014) Fundamental concepts in heterogeneous catalysis. Wiley, Hoboken

Otani M, Sugino O (2006) First-principles calculations of charged surfaces and interfaces: a plane-wave nonrepeated slab approach. Phys Rev B 73:115407

Ouyang R, Liu JX, Li WX (2013) Atomistic theory of Ostwald ripening and disintegration of supported metal particles under reaction conditions. J Am Chem Soc 135:1760–1771

Pacchioni G (2008) Modeling doped and defective oxides in catalysis with density functional theory methods: room for improvement. J Chem Phys 128:182505

Pinto LMC, Quaino P, Arce MD, Santos E, Schmickler W (2014) Electrochemical adsorption of OH on Pt(111) in alkaline solutions: combining DFT and molecular dynamics. Chem Phys Chem 15:2003–2009

Quaranta V, Hellström M, Behler J (2017) Proton-transfer mechanisms at the water–ZnO interface: the role of presolvation. J Phys Chem Lett 8:1476–1483

Ren X, Rinke P, Joas C, Scheffler M (2012) Random-phase approximation and its applications in computational chemistry and materials science. J Mater Sci 47:7447–7471

Reuter K (2013) First-principles kinetic Monte Carlo simulations for heterogeneous catalysis: concepts, status and frontiers. In: Deutschmann O (ed) Modelling and simulation of heterogeneous catalytic reactions: from the molecular process to the technical system. Wiley-VCH, Weinheim. pp 71–111

Reuter K (2016) Ab initio thermodynamics and first-principles microkinetics for surface catalysis. Catal Lett 146:541–563

Reuter K, Plaisance CP, Oberhofer H, Andersen M (2017) Perspective: on the active site model in computational catalyst screening. J Chem Phys 146:040901

Richter NA, Sicolo S, Levchenko SV, Sauer J, Scheffler M (2013) Concentration of vacancies at metal-oxide surfaces: case study of MgO(100). Phys Rev Lett 111:045502

Ringe S, Oberhofer H, Hille C, Matera S, Reuter K (2016) Function-space based solution scheme for the size-modified Poisson-Boltzmann equation in full-potential DFT. J Chem Theory Comput 12:4052–4066

Rittmeyer SP, Bukas VJ, Reuter K (2018) Energy dissipation at metal surfaces. Adv Phys X 3:1381574

Rossmeisl J, Skúlason E, Bjorketun MJ, Tripkovic V, Nørskov JK (2008) Modeling the electrified solid-liquid interface. Chem Phys Lett 466:68–71

Rossmeisl J, Chan K, Ahmed R, Tripkovic V, Bjorketun ME (2013) pH in atomic scale simulations of electrochemical interfaces. Phys Chem Chem Phys 15:10321–10325

Sabbe MK, Reyniers MF, Reuter K (2012) First-principles kinetic modeling in heterogeneous catalysis: An industrial perspective on best-practice, gaps and needs. Catal Sci Technol 2: 2010–2024

Sauer J, Freund HJ (2015) Models in catalysis. Cat Lett 145:109–125

Schimka L, Harl J, Stroppa A, Grüneis A, Marsman M, Mittendorfer F, Kresse G (2010) Accurate surface and adsorption energies from many-body perturbation theory. Nat Mater 9:741–744

Schlögl R (2015) Heterogeneous catalysis. Angew Chem Int Ed 54:3465–3520

Schnur S, Groß A (2011) Challenges in the first-principles description of reactions in electrocatalysis. Catal Today 165:129–137

Sinstein M, Scheurer C, Matera S, Blum V, Reuter K, Oberhofer H (2017) Efficient implicit solvation method for full potential dft. J Chem Theor Comput 13:5582–5603

Skúlason E, Tripkovic V, Björketun ME, Gudmundsdottir S, Karlberg G, Rossmeisl J, Bligaard T, Jónsson H, Nørskov JK (2010) Modeling the electrochemical hydrogen oxidation and evolution reactions on the basis of density functional theory calculations. J Phys Chem C 114: 18182–18197

Stamatakis M, Vlachos DG (2012) Unraveling the complexity of catalytic reactions via kinetic Monte Carlo simulation: current status and frontiers. ACS Catal 2:2648–2663

Stecher T, Reuter K, Oberhofer H (2016) First-principles free-energy barriers for photo-electrochemical surface reactions: proton abstraction at $TiO_2(110)$. Phys Rev Lett 117:276001

Stodt D, Noei H, Hättig C, Wang Y (2013) A combined experimental and computational study on the adsorption and reactions of NO on rutile TiO_2. Phys Chem Chem Phys 15:466–472

Surendrala S, Todorova M, Finnis MW, Neugebauer, J (2018) First-principles approach to model electrochemical reactions: Understanding the fundamental mechanisms behind Mg corrosion. Phys Rev Lett 120:246801

Sutton JE, Vlachos DG (2015) Building large microkinetic models with first-principles[U+05F3] accuracy at reduced computational cost. Chem Eng Sci 121:190–199

Sutton JE, Guo W, Katsoulakis MA, Vlachos DG (2016) Effects of correlated parameters and uncertainty in electronic-structure-based chemical kinetic modelling. Nat Chem 8:331–337

Toulhoat H, Raybaud P (2003) Kinetic interpretation of catalytic activity patterns based on theoretical chemical descriptors. J Catal 216:63–72

Ulissi ZW, Medford AJ, Bligaard T, Nørskov JK (2017a) To address surface reaction network complexity using scaling relations machine learning and DFT calculations. Nat Commun 8:14621

Ulissi ZW, Tang MT, Xiao J, Liu X, Torelli DA, Karamad M, Cummins K, Hahn C, Lewis NS, Jaramillo TF, Chan K, Nørskov JK (2017b) Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO_2 reduction. ACS Catal 7:6600–6608

van Santen RA, Neurock M (2006) Molecular heterogeneous catalysis: a conceptual and computational approach. Wiley-VCH, Weinheim

Wang T, Jelic J, Rosenthal D, Reuter K (2013) Exploring pretreatment-morphology relationships: ab initio Wulff construction for RuO_2 nanoparticles under oxidizing conditions. Chem Cat Chem 5:3398–3403

Yip S (ed) (2005) Handbook of materials modeling. Springer, Dordrecht

Energy Trends in Adsorption at Surfaces

54

Abhinav S. Raman and Aleksandra Vojvodic

Contents

Abstract

Adsorption is a process of vital importance in surface science and heterogeneous catalysis, with early studies dating back to the days of Irving Langmuir. Over the last decades, the advent of first-principles methods coupled with increasing computing power has enabled generation of large data sets of the energetics describing the interaction between adsorbates with different solid surfaces. This has aided in the conceptualization of energy trends through scaling relations, originally obtained by empirical means, and primarily used for the ease of computational burden in chemical transformations involving adsorption. However, the use of scaling relations has now evolved beyond its original intended purpose, with their use in identifying descriptors for reactivity and activity, as well as extension to more complex adsorbates and surfaces beyond transition metals. This chapter traces the history of scaling relations from their inception, through their extension to enable the understanding of chemical transformations and aid in the design of better catalysts, ending with the challenges and an outlook.

A. S. Raman · A. Vojvodic (✉)
Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA, USA
e-mail: sanraman@seas.upenn.edu; alevoj@seas.upenn.edu

1321

1 Introduction

Adsorption is the phenomenon of adhesion of atoms, molecules, radicals, or ions from a gas or liquid onto a surface of a solid material, a nanoparticle, or a larger molecule. In this chapter, the focus is on the adsorption on solid surfaces. Adsorption is at the heart of any gas-surface and liquid-surface interaction found in, e.g., catalytic processes, growth, etc. and has been a vivid research area for decades. Studies of the interaction of atoms and molecules with surfaces go back to Irving Langmuir, considered to be one of the founders of surface science (Langmuir 1961). Over the last two decades, we have seen rapid changes in the fields of surface science and heterogeneous catalysis to a point where we have reached an understanding of the chemistry on surfaces at a level where we can start designing them for their chemical properties in a controlled and systematic way (Medford et al. 2015). The advent of density functional theory (DFT) coupled with increasing computational power has made it possible to gain a deeper understanding of surface processes while also providing the ability to computationally screen among existing, modified, and entirely new materials for complex heterogeneous catalytic processes involving many adsorption intermediates (Nørskov et al. 2009, 2011; Xu 2009; Christensen and Nørskov 2008). At the same time, first-principles atomistic thermodynamics-based approaches have made it possible to determine the relative stabilities of different surface facets and terminations under relevant experimental conditions, improving the accuracy and predictability power of computational methods (Bollinger et al. 2003; Reuter and Scheffler 2001, 2003; Reuter et al. 2004; Nørskov et al. 2006; Rossmeisl et al. 2006; Hansen et al. 2008; Inoğlu and Kitchin 2009; Bendavid and Carter 2013; Holby et al. 2009). An effective method to quantitatively calculate and ultimately predict adsorption energies for the many possible adsorbates that might be involved in chemical processes is highly desirable. With such a suite of data on hand, it allows for comparative studies of different surfaces, enabling the identification of trends in the adsorption energetics that ultimately can be captured by simpler adsorption models, rather than through time-consuming heavy first-principles calculations. Correlations known as energy scaling relations of the adsorption energies between chemically related species have become one of the most popular methods in achieving this. Initially, these relations emerged as a result of empirically fitting large enough computationally derived data sets and were then rationalized and explained for certain groups of materials based on electronic structure arguments, which we will return to in a later subchapter. These scaling relations serve multiple purposes, the most obvious being the reduction in the computational cost associated with the calculation of adsorption energies for a large number of reaction intermediates, when potentially the overall surface reactivity trend can be gauged from a single or a few adsorbate(s), which is discussed in further detail later in this chapter. This way, these scaling relations can aid in the reduction of the dimensionality of a chemical transformation problem, providing insights into the identification of effective descriptors for surface reactivity (Medford et al. 2015; Nørskov et al. 2014; Montemore and Medlin 2014; Greeley 2016). A plethora of papers on adsorption and reactivity, a few

reviews on scaling trends (Montemore and Medlin 2014; Greeley 2016), and a book (Nørskov et al. 2014) have been written on these topics. Therefore, we will here focus on introducing the concepts of adsorption and scaling trends, discussing their applicability, putting them in a larger perspective, and providing an outlook on what the future might bring.

2 Origin of Adsorption Scaling Relations

The origin of adsorption scaling relations for chemisorbed adsorbates can be traced back to the seminal paper by Abild-Pedersen et al. (2007), which in a systematic way established a linear correlation between the adsorption energies of simple hydrides (AH_n) and the corresponding central atom (A) on transition metal surfaces, as given by Eq. 1.

$$\Delta E_{AH_n} = \gamma_s(n)\Delta E_A + \xi,\qquad(1)$$

with the slope $\gamma_s(n)$ given by:

$$\gamma_s(n) = \frac{N-n}{N}\qquad(2)$$

where N is the maximum valency of the central atom A and $n \leq N$ are the number of bonds saturated by bonding to other atoms. For example, C, N, and O would have N equal to 4, 3, and 2, respectively, with CH_3 and CH_2 having n of 3 and 2, respectively. Conceptually, this can be rationalized by taking the N levels to be degenerate and linearly independent, thereby providing an equal contribution to the overall adsorption energy when the central atom A binds to a surface. Removing the contribution from the number of bonds saturated by other atoms (n) provides the number of bonds that the central atom can form with the surface (i.e., dangling bonds). Figure 1 shows the linear adsorbate scaling relationship for the CH_n, NH_n, OH_n, and SH_n species on different surfaces of unary elemental transition metals, respectively.

As is evident from Fig. 1, the slope or the scaling parameter $\gamma_s(n)$ in Eq. 1 is independent of the surface facet, while the intercept ξ is surface-facet dependent. This essentially arises from the fact that the adsorbate, at least to a first-order approximation, will form bonds with the surface commensurate with its degree of unsaturation, irrespective of the surface. On the other hand, the difference in the exact local coordination of the adsorption site at the surface affects the overall adsorption energy, resulting in the difference in the intercept ξ. While further modifications to account for more complex adsorbates, coverage effects, etc. have all been built on the scaling relationship given by Eq. 1, it should be noted that some scaling parameters can be determined from bond-order conservation models such as the UBI-QEP methods (Shustorovich 1986; Shustorovich and Bell 1988; van Santen 2010), which predate the use of DFT for adsorption energy calculations.
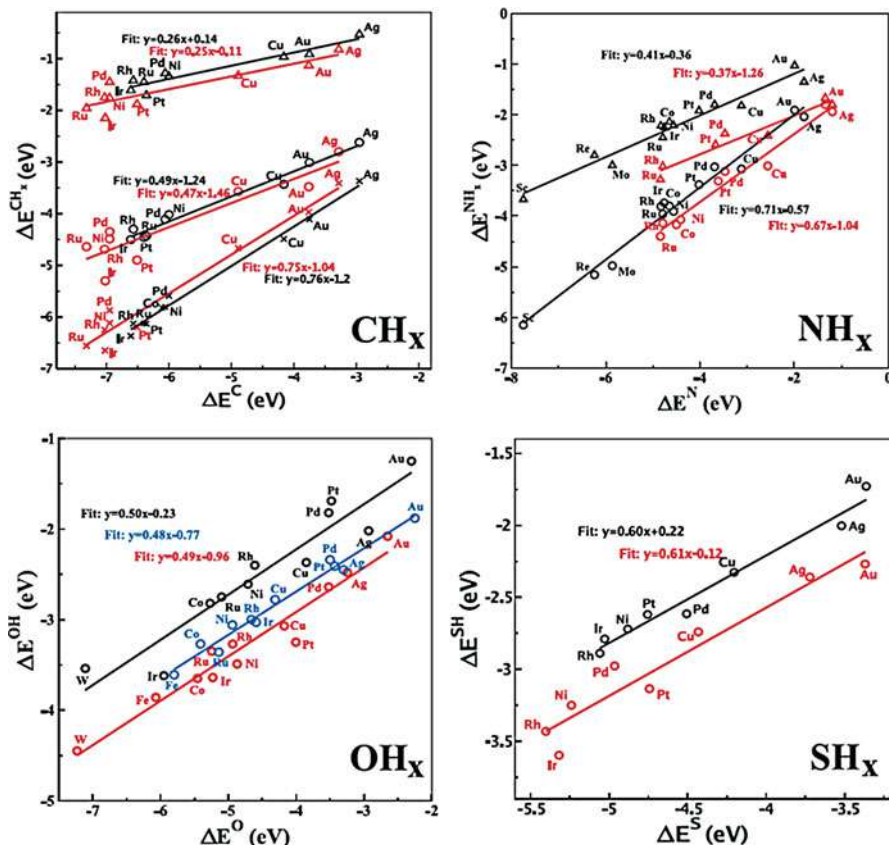
Fig. 1 Linear adsorption energy scaling relations for simple hydrides AH$_x$ on close-packed (black) and stepped (red) surfaces of different transition metals. (Adapted from Fig. 1 in Abild-Pedersen et al. (2007) based on DFT calculated data)

The simple scaling relationship scheme presented in Eq. 1 has been extended to account for more complex adsorbates. This can be done by first realizing that the adsorption energy of an AH_n species on a given metal can be rewritten in terms of a comparison of the central atom A on that given metal and a reference transition metal, as given below (Abild-Pedersen et al. 2007; Jones et al. 2008):

$$\Delta E_{AH_n;M_2} = \Delta E_{AH_n;M_1} + \gamma_s(n)\left(\Delta E_{A;M_2} - \Delta E_{A;M_1}\right) \tag{3}$$

where M_1 and M_2 represent the two different transition metals. Equation 3 provides the basis for a simple generalization to complex adsorbates, by simply summing over all the adsorbate atoms that bind to the surface, as given below (Abild-Pedersen et al. 2007; Jones et al. 2008):

$$\Delta E_{R;M_2} = \Delta E_{R;M_1} + \sum_i \gamma_i \left(\Delta E_{A_i;M_2} - \Delta E_{A_i;M_1} \right) \tag{4}$$

where R refers to any generic adsorbate and γ_i is redefined as:

$$\gamma_i = \frac{N - n_{intra}}{N} \tag{5}$$

where n_{intra} refers to the total number of intra-adsorbate bonds attached to the central atom A. This generalization has, for example, been used for unsaturated hydrocarbons (CH_x-CH_2), with x = (0,1,2) to successfully predict adsorption energetics on both close-packed and stepped surfaces of transition metals as a function of the adsorption energy of atomic carbon (Jones et al. 2011), as shown in Fig. 2. One finds that a single linear scaling is not enough; rather a piecewise scaling demarcates two separate independent regions, one corresponding to those transition metals that are reactive enough (including non-noble metals and some Cu surfaces) to alter the intramolecular bond of the adsorbate versus those that are not (including the metals Au, Ag, and certain Cu surfaces).

It should be noted that other bond-order-based approaches have been used for multicenter adsorbates, especially those belonging to the class of ethanol derivatives ($CH_xCH_yO_z$) (Ferrin et al. 2009; Salciccioli et al. 2010; Liu and Greeley 2013; Liu et al. 2015) with varying degrees of success on different transition metal surfaces. While scaling concepts based on the usual bond-order conservation approach can be used for such adsorbates, care must be taken when applying the principle to noble metals and potentially other systems with complex surfaces and adsorbates.
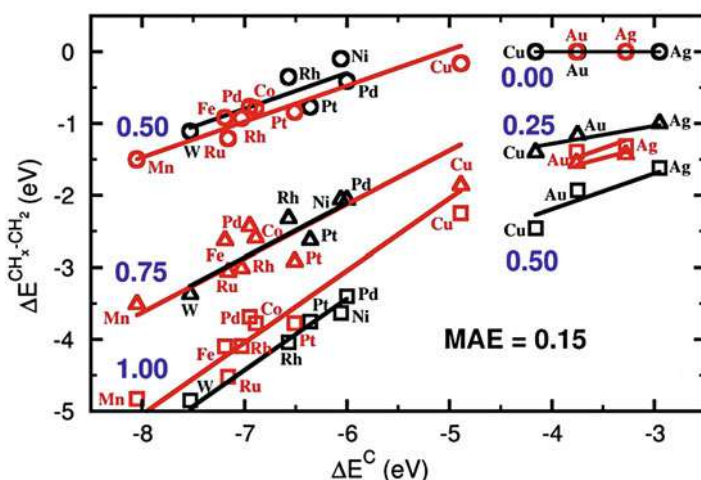
Fig. 2 Calculated adsorption energies for CH_x-CH_2 hydrocarbons, with x = (0,1,2) as a function of the adsorption energy of atomic carbon on close-packed (black) and stepped (red) surfaces of transition metals. (Adapted from Jones et al. (2011))

3   Adsorption Energies as Descriptors for Reactivity/Activity

While it is important to generalize the scaling relations to more complex adsorbates, it should also be realized that the true power of scaling relations goes beyond just easing computational effort. Understanding and quantifying surface reactivity has been the main goal of heterogeneous catalysis in order to provide a systematic design paradigm for superior catalysts, in contrast to the conventional trial and error approaches. To this end, the Sabatier principle (Sabatier 1913; Che 2013) provides a conceptual framework by suggesting that the optimal performing catalyst should bind the adsorbates with some intermediate strength, i.e., neither too weakly (activation limited) nor too strongly (desorption limited). This therefore establishes the typical "volcano-type" relation between rate (activity) and the bond strength, as shown in Fig. 3.

    While this provides a theoretical framework, it is not quantitative, as the physical quantity characterizing the "bond strength" is not described, and the value of the "bond strength" that describes the optimal catalyst is unknown. Therefore, to use the Sabatier principle in a predictive manner, descriptors characterizing a heterogeneous catalytic reaction are needed. Since adsorption plays a crucial role in heterogeneous catalysis, the adsorption energies of the relevant intermediates on a given surface of interest serve as descriptors that suitably characterize the reactivity and potentially the activity. To illustrate this, the ammonia synthesis activity on different transition metals is shown in Fig. 4 with two different descriptors: the bulk oxide heat of formation (Fig. 4a) and the calculated nitrogen adsorption energy (Fig. 4b).

    It is evident that the activity "volcano" is well defined when the descriptor is the nitrogen adsorption energy, since it provides a better description of the interaction between the adsorbate and the transition metal surface. While the adsorption of nitrogen is a key step in ammonia synthesis, there are several other elementary steps with multiple reaction intermediates whose adsorption energetics are vital for the overall reaction. What then distinguishes the adsorption of nitrogen as a single descriptor for ammonia synthesis? To illustrate this, the transition state energy for

Fig. 3 Representation of the Sabatier principle. (Adapted from Medford et al. (2015))
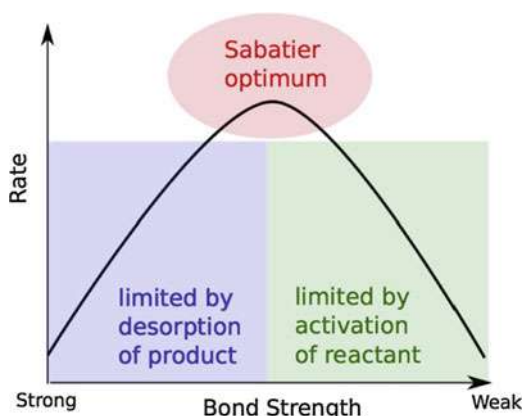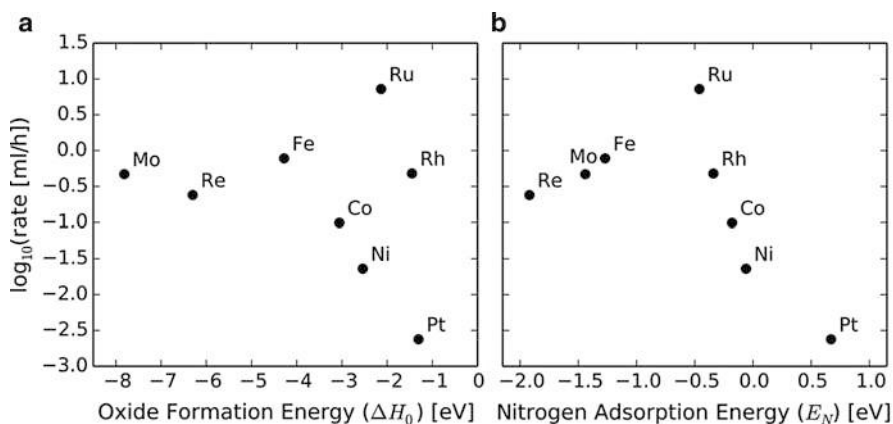
Fig. 4 Measured ammonia synthesis activity as a function of two suggested descriptors for the reaction: oxide heat of formation and nitrogen adsorption energy, of which the adsorption energy provides the most satisfactory explanation of the trends by ordering the elemental metal catalysts onto a volcano curve. (Adapted from Medford et al. (2015))

Fig. 5 Calculated transition state energy for nitrogen dissociation vs nitrogen adsorption energy on stepped transition metal surfaces. (Adapted from Medford et al. (2015))
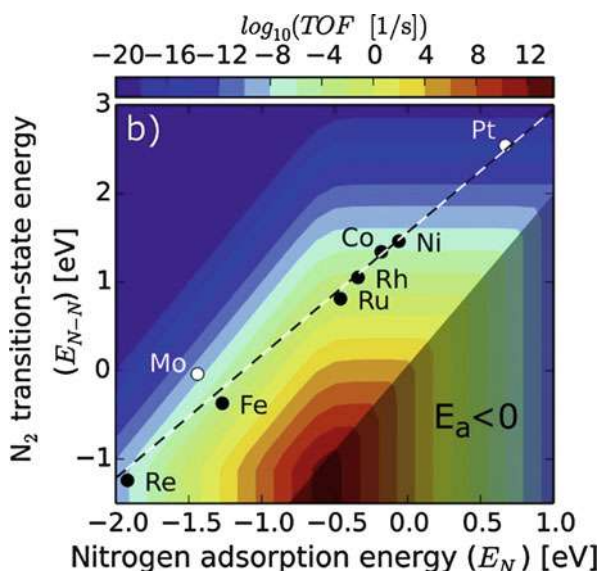
dissociation of a nitrogen molecule is plotted against nitrogen adsorption energy and the rate of reaction computed using a mean-field kinetic model in Fig. 5. It is well established from both experiments and theory that the N_2 dissociation step is rate determining for certain elemental metal active catalysts (Schlögl 2003; Ertl 1980; Logadottir et al. 2001). According to the findings presented in Fig. 5, the transition state energy for the dissociation step scales linearly with the nitrogen adsorption

energy. This explains why the nitrogen adsorption energy would serve as a unique descriptor for ammonia synthesis.

More fundamentally, it is also clear that variations in transition state energies and adsorption energies are governed by the same physics, resulting in such correlations between the two (Plessow and Abild-Pedersen 2015). This enables the development of transition state scaling relations that effectively map the space of adsorption energies onto transition state energies, aiding in the prediction of surface activity (Logadottir et al. 2001; Plessow and Abild-Pedersen 2015; Nørskov et al. 2002; Liu and Hu 2001a, b; Michaelides et al. 2003; Bligaard et al. 2004). Even more remarkable is the similarity of the transition state scaling relationships for chemically and geometrically related molecules undergoing a similar mechanistic reaction step. For example, the transition state energies of simple diatomic molecules show a similar correlation with respect to their dissociative chemisorption energies, providing at least to a first-order approximation an adsorbate-independent way of estimating transition state energies and activation barriers (Nørskov et al. 2002). It should be noted that this correlation sometimes referred to as "universality" of the transition state scaling relations as shown in Fig. 6 is only valid for subsets of molecules on very similar surfaces and adsorption sites. Finally, it should be noted that the popular Brønsted-Evans-Polanyi (BEP) relations (Brønsted 1928; Evans

Fig. 6 Transition state scaling relations for dissociation of simple diatomic molecules on flat and stepped metal surfaces. (Adapted from Nørskov et al. (2002, 2014))
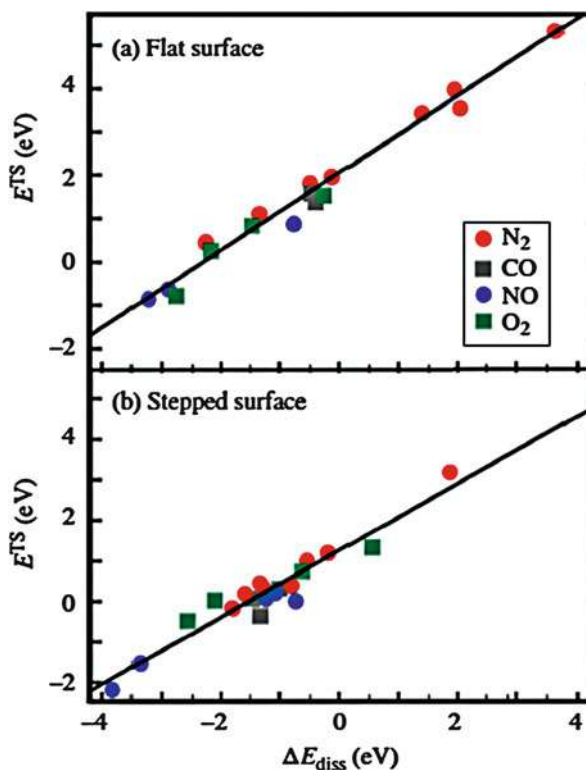
and Polanyi 1938) which have been used extensively for understanding trends in chemical reactions through linear correlations between the activation and reaction energies are simply a subset of the transition state scaling relations. This therefore establishes the importance of trends in adsorption energetics in quantifying surface activity.

4   Extensions to Other Systems Beyond Transition Metals

Considerable progress in the description of adsorption and chemical reactions of more complex systems has been made including scaling relations and extensions of similar ideas for systems beyond pure elemental transition metal surfaces. Adsorbate scaling relations have been established for metal alloys including near-surface metal alloys (Calle-Vallejo et al. 2012; Greeley et al. 2009). In addition to metal alloys, scaling relations have also been found for even more complex solid systems most prevalently for transition metal compounds, including certain surfaces of transition metal oxides, nitrides, carbides, sulfides, etc. (Fernández et al. 2008; Vojvodic et al. 2009, 2011; Rossmeisl et al. 2005; Man et al. 2011; Viñes et al. 2013). One of the earliest papers in this field by Fernandez et al. (2008) showed that linear scaling relations exist between OH, NH_x, and SH and the corresponding central atom, for adsorption on oxide, nitride, and sulfide surfaces, respectively. Further, Vojvodic et al. (2009) demonstrated scaling for NH_x intermediates and BEP relations for transition metal carbides, i.e., scaling relations for intermediates with a central atom having a different chemical component compared to the transition metal compound itself. When it comes to oxides, they have received a lot of attention, for example, with the aim of modeling the electrochemical oxygen evolution reaction, most notably for single metal oxides such as MO_2 oxides and binary metal oxides such as perovskites (Rossmeisl et al. 2005; Man et al. 2011). Versatile BEP relations were found for a range of transition metal oxides including rutiles and perovskites (Vojvodic et al. 2011), where the BEP relation depends on the type of oxide, the active site, and the dissociating molecule. A comparison between the different BEP relations for transition metal carbides and oxides with metals has also been performed (Viñes et al. 2013).

Challenges in modeling these complex solid surfaces arise, on one hand, due to complications in describing the structure and stoichiometries of these surfaces and, on the other hand, from possible limitations in our current description of the exchange-correlation in standard density functional theory (Bengone et al. 2000). Although there are numerous examples of excellent agreement between DFT calculations and experiments, our current description of the exchange and correlation may be insufficient when studying the surface chemistry of complex systems.

Up until now we have only discussed adsorption in the chemisorption limit; however, physisorption is another important class of adsorption. In many cases, physisorption arises from weak interactions due to the relatively long-ranged van der Waals forces between the adsorbate and the surface. These interactions can have

a strong dependence on the spatial extent of the adsorbate, implying that adsorbate correlations normalized to the corresponding active components can be established for comparable systems. This has been observed experimentally, for example, the physisorption of linear alkanes on a Pt (Hong et al. 2016) surface (Tait et al. 2006). Further, with the advent of exchange-correlation functionals that capture dispersion interactions (Dion et al. 2004; Grimme 2004, 2006; Tuma and Sauer 2006; Grimme et al. 2007, 2010; Román-Pérez and Soler 2009; Svelle et al. 2009; Hansen et al. 2010; Lee et al. 2010; Wellendorff et al. 2012; Berland et al. 2015), it is now feasible to computationally model physisorption in systems involving significant contributions from dispersion interactions. For example, the experiment by Tait et al. (2006) was rationalized by Londero et al. (2012). Physisorption in more complex systems such as zeolites, which constitute a large class of materials commonly employed in heterogeneous catalysis and surface science (Brogaard et al. 2012; Bukowski and Greeley 2016; Göltl et al. 2017; Liu et al. 2017), has also been illustrated, for example, see correlation for the physisorption enthalpy of n-alkanes on a zeolite H-ZSM-22 shown in Fig. 7.

In contrast to heterogeneous catalysts with active sites located at continuous surfaces, the active site of a homogeneous catalyst is often comprised of a transition metal center consisting of one or multiple metal atoms embedded in a metal-ligand complex. Homogeneous systems offer large chemical tunability through, e.g., ligand control and steric interactions, to promote or prevent specific chemical reactions. Early work by Bryndza et al. (1987) identified linear correlations between the ligand to transition metal complex interaction and the bond strength of a ligand to a hydrogen atom. Schock and Marks (1988) further investigated this through bond order and electronegativity aspects. More recently, examples of scaling relations for homogeneous systems (Wodrich et al. 2016; Wang et al. 2016) showing that
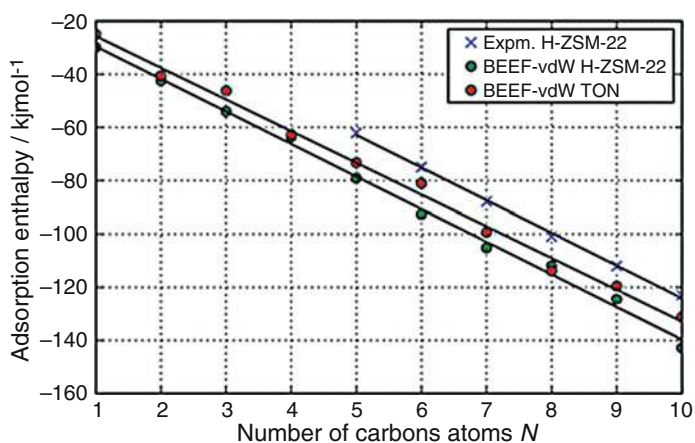
Fig. 7 Physisorption enthalpy of n-alkanes in a H-ZSM-22 zeolite. (Adapted from Brogaard et al. (2012))

thermodynamic descriptors and kinetic relations can be applicable beyond solid surfaces have been demonstrated.

5   Electronic Structure-Based Adsorption Models

The canonical model for understanding chemisorption on metal surfaces can be traced back to the Newns-Anderson model (Newns 1969; Grimley 1971, 1967a, b; Anderson 1961), which employs a simplified Hamiltonian that accounts for the dominant mechanisms involved in bond formation between the adsorbate and the metal surface. By studying the evolution in the electronic structure of the individual components and the combined system as the adsorbate approaches the metal surface, one can capture the interaction between the adsorbate and the metal through the adatom-projected density of states. This essentially leads to two adsorption limits known as the weak and strong chemisorption limits, classified based on the relative strength between the average coupling matrix element between adsorbate and surface, V, and the bandwidth of the surface states, W. When $V \ll W$, as would be the case when the adsorbate interacts with the sp band of the metal surface, the weak chemisorption limit is reached, and when $V \gg W$, as would be the case when the adsorbate interacts with the d-band of the metal surface, the strong chemisorption limit is reached. In the case of adsorption on a transition metal surface, which consists of a broad sp band and a narrow d-band, both these limits are encountered. The solution to the Newns-Anderson Hamiltonian in such a scenario is schematically illustrated in Fig. 8.

As can be seen in Fig. 8, the interaction between the sp band and an adsorbate results in a broadening of the adatom energy level, with a downward shift in the energy relative to the free adsorbate. The subsequent interaction between the shifted adatom state and the d-band splits the adatom energy level into a broadened
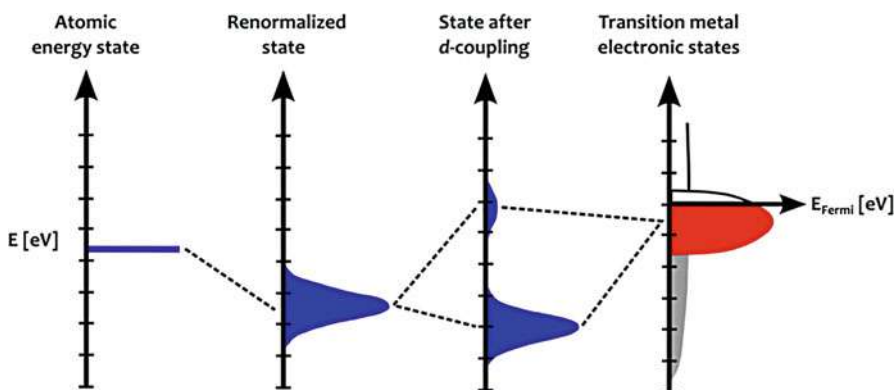
Fig. 8 Illustration of the changes in the adatom-projected density of states upon adsorption on a transition metal surface. (Adapted from Nørskov et al. (2014))

downward-shifted bonding state that is occupied, since it is positioned lower in energy than the lowest occupied metal state, and a broadened upward shifted antibonding state, which may be unoccupied or partly (or fully) occupied based on its position relative to the Fermi level. The region between them corresponds to what is sometimes referred to as a weak nonbonding state. Clearly, the strongest bond between the adsorbate and the metal surface occurs when all the bonding states are filled and all the antibonding states are empty. This provides a simplified fundamental understanding of the energetics involved in the interaction between an adsorbate and a transition metal surface.

6   Adsorption Model for Metals: *d*-Band Model

Since transition metals are vital to most of heterogeneous catalysis, the role played by the *d*-electrons must be carefully elucidated. To this end, the *d*-band model for adsorption on transition metal surfaces, proposed by Hammer and Nørskov (1995a, b), aims to systematically explain the energy trends seen in transition metals, building on the Newns-Anderson model discussed previously. The adsorption energy according to the *d*-band model is split into a contribution from the *sp* states and the *d*-states assuming the following expression:

$$E_{ads} = n_{deg} E_{sp-hyb} + n_{deg} E_{d-hyb} \tag{6}$$

where n_{deg} is the degeneracy of the adsorbate level and E_{sp-hyb} and E_{d-hyb} are the energy contributions resulting from the interaction between the adatom and the surface *sp* and *d*-bands, respectively. The energy contributions resulting from the interaction between the *sp* band and the adsorbate are shown to be very similar for the different transition metals, resulting in the adsorption energy trends being determined by the interaction with the *d*-band of the metal. Therefore, the *d*-band center, filling, and the coupling matrix element between the adsorbate and the *d*-band of the metal serve as important parameters that determine the strength of the bond formed between the adsorbate and the transition metal surface. To illustrate this, the adsorption of atomic oxygen on 4*d* transition metals, as a function of the *d*-band centers, is shown in Fig. 10. Moving from right to left in the periodic table, the *d*-band center moves up in energy, resulting in the antibonding states of the adsorbate becoming empty (see Fig. 9). This yields a stronger bond between the adsorbate and the transition metals toward the left of the periodic table.

Generalized reactivity descriptors beyond the *d*-band center, i.e., the average energy of electronic *d*-states projected onto a surface metal atom, have been introduced. For example, the upper edge of the *d*-band was identified as an improved reactivity descriptor for pristine transition metals (Vojvodic et al. 2014a) taking into account the effects of the spread in energy on the position of adsorbate-metal antibonding states. In addition, the *d*-band shape of a metal site was shown to have an important role in determining the local surface reactivity of transition metal alloys (Xin et al. 2014). Numerous studies have over the last decades contributed
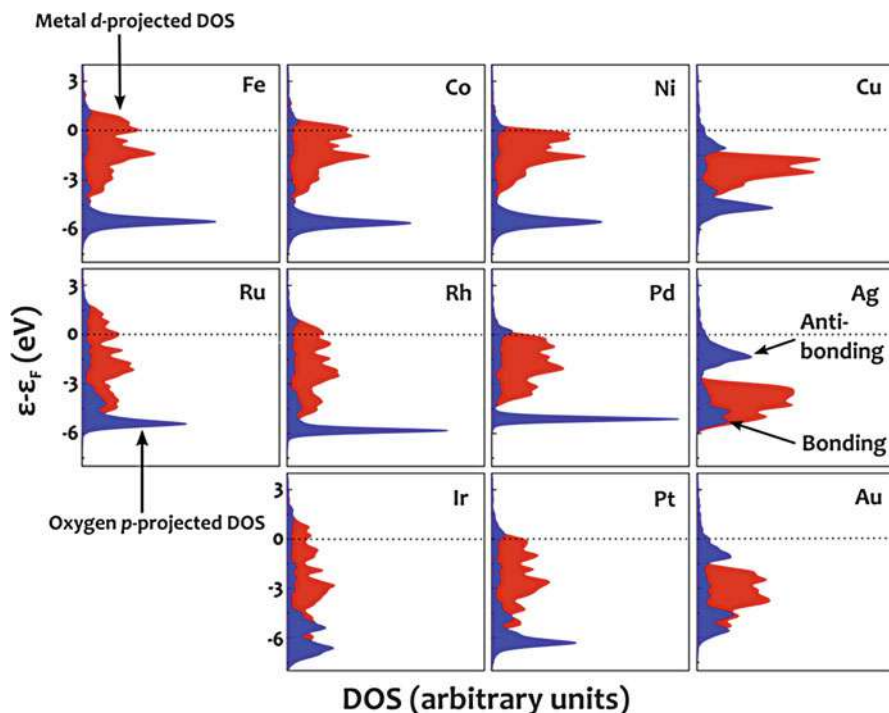
Fig. 9 Calculated density of states projected onto the *d*-states of the surface atoms of the transition metals and the *2p*-projected density of states of an adsorbed oxygen atom. (Adapted from Nørskov et al. (2014))

to understand the interaction between metal states including the *d*-states and an adsorbate state (Gajdoš et al. 2004; Hammer 2006; Roudgar and Groß 2003; Greeley and Mavrikakis 2004; Kitchin et al. 2004; Nikolla et al. 2009; Marković and Ross 2002; Tripa et al. 1999; Mills et al. 2003; Zhang et al. 2005; Kibler et al. 2005; Behm 1998; Schnur and Groß 2010; Menning and Chen 2009; Groß 2008; Liu et al. 2004; Inderwildi et al. 2007).

7   Outlook

Even though the existence of energy scaling relations which together with activity maps and the *d*-band model has enabled a quantitative understanding and a faster discovery of new catalyst materials (Medford et al. 2015; Nørskov et al. 2011, 2014), recently it has been discovered (Vojvodic and Nørskov 2015; Koper 2013; Seh et al. 2017) that their existence also sets limitations. Under the assumption that each group of reaction intermediates binds to the same type of active site for a given type of material facet, it has been found for several catalytic reactions that the activity
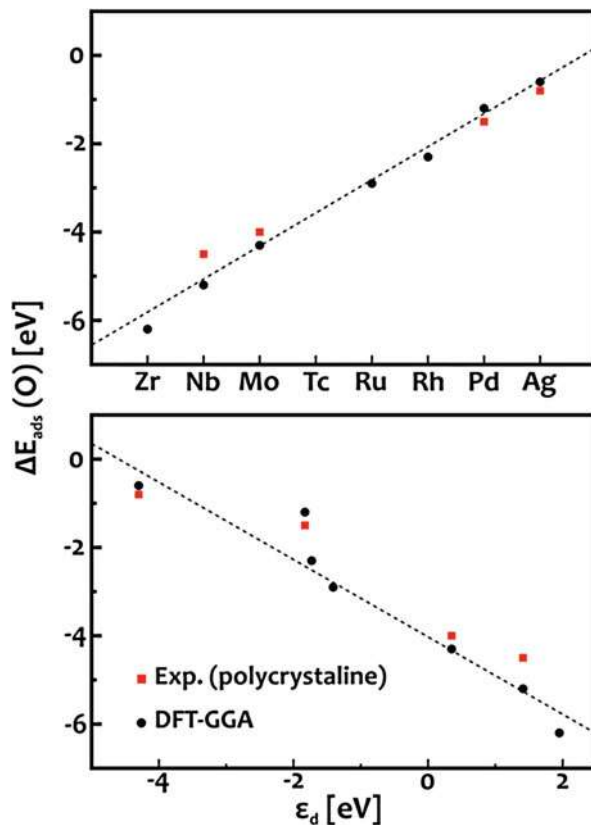
of the surface has an upper theoretical limit, for example, the thermal ammonia synthesis (Vojvodic et al. 2014b), electrochemical ammonia synthesis (Montoya et al. 2015), oxygen evolution reaction, oxygen reduction reaction (Koper 2013; Viswanathan et al. 2012; Nørskov et al. 2004), CO_2 reduction reaction (Peterson and Nørskov 2012; Hansen et al. 2013), and catalytic hydrogenation of carbon monoxide (Cheng and Hu 2008) to name a few. It has been argued that for us to overcome these limitations, we need to find systems where the scaling relations are very different from the systems that we have discovered so far or find systems where the scaling relations do not apply. Several schemes that might be used to circumvent the scaling relations have been suggested; see, for example, a summary in Vojvodic and Nørskov (2015), where both tuning of properties internal to the material, as well as external means are discussed. One of the ideas is to either change how different intermediates are binding to the surfaces, in that way changing the adsorbate scaling relations, while others suggest how one could change the transition state relative to the thermodynamic intermediates, in that way possibly decoupling the kinetics and thermodynamics or finding more favorable BEP relations. So far, only a couple of these strategies have been computationally explored, for example, bifunctional

catalysts through alloying, doping, or defects (Hong et al. 2016; Halck et al. 2014) and 3D active sites as can be found in a confined space (Doyle et al. 2015, 2017). In other words, many possible routes for tuning the surface chemistry remain unexplored, while others might already have been observed but not identified as separate phenomena. There is a long way to go for us to model and design more complicated catalytic systems comprised of multiple components and more intricate details of the material and the environment it is interacting with, for example, understanding the solid-liquid interfaces is a necessity for us to gain a full picture of electrode-electrolyte interactions present in electrochemical systems. Besides this more applied direction for research to progress, there is the more fundamental one that we should be concerned with.

Despite considerable progress, a theoretical understanding of the reactivity of transition metal oxides is less established than that of pure transition metals. As a matter of fact, a complete underlying electronic structure theory is still lacking for complex solid surfaces beyond metals. Vojvodic et al. (2009) showed that the electronic structure descriptor to be considered for an adsorption model for transition metal compounds, as was illustrated for carbides, is more intricate than the one for metals. This approach was applied to a doped oxide system as well (García-Mota et al. 2013) and was found to be able to describe the adsorption trends. Other descriptors have been suggested as well, including the simple one being the number of d-electrons of a system (Bockris and Otagawa 1984; Mun et al. 2005; Calle-Vallejo et al. 2013), the e_g filling of the d-electrons of the bulk system if such are present (Suntivich et al. 2011), e_g or t_{2g} filling of the active surface metal site for a specific surface facet (Vojvodic and Nørskov 2011; Montoya et al. 2018), d-band center, and the p-band center of O in the bulk (Lee et al. 2016) system if present (Hong et al. 2015) or of the surface O near the active site (Montoya et al. 2018), all with different level of success. In addition, coordination-based descriptors have emerged in the literature for complex metal systems including nanoparticles (Jiang et al. 2009; Calle-Vallejo et al. 2014, 2015a, b; Wang et al. 2018; Ma and Xin 2017). Also, recently relationships between adsorbate binding energies and the stability of individual metal surface sites have been considered as descriptors (Roling and Abild-Pedersen 2017). The key thing for us to show is how transferable these are, not only between different materials but also between different surfaces. Ultimately, the goal is to identify suitable descriptor(s), potentially electronic structure-based descriptor(s), that can be used to rationalize the different adsorption trends on these surfaces and help in the endeavor of designing new materials.

References

Abild-Pedersen F et al (2007) Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. Phys Rev Lett 99:016105

Anderson PW (1961) Localized magnetic states in metals. Phys Rev 124:41–53

Behm RJ (1998) Spatially resolved chemistry on bimetallic surfaces. Acta Phys Pol A 93:259–272

Bendavid LI, Carter EA (2013) First-principles predictions of the structure, stability, and photocat-
    alytic potential of Cu_2O surfaces. J Phys Chem B 117:15750–15760
Bengone O, Alouani M, Blöchl P, Hugel J (2000) Implementation of the projector augmented-wave
    LDA+U method: application to the electronic structure of NiO. Phys Rev B 62:16392–16401
Berland K et al (2015) van der Waals forces in density functional theory: a review of the vdW-DF
    method. Rep Prog Phys 78:066501
Bligaard T et al (2004) The Brønsted–Evans–Polanyi relation and the volcano curve in heteroge-
    neous catalysis. J Catal 224:206–217
Bockris JO, Otagawa T (1984) The electrocatalysis of oxygen evolution on perovskites.
    J Electrochem Soc 131:290–302
Bollinger MV, Jacobsen KW, Nørskov JK (2003) Atomic and electronic structure of MoS_2
    nanoparticles. Phys Rev B 67:085410
Brogaard RY, Moses PG, Nørskov JK (2012) Modeling van der Waals interactions in zeolites with
    periodic DFT: physisorption of *n*-alkanes in ZSM-22. Catal Lett 142:1057–1060
Brønsted JN (1928) Acid and basic catalysis. Chem Rev 5:231–338
Bryndza HE, Fong LK, Paciello RA, Tam W, Bercaw JE (1987) Relative metal-hydrogen,
    -oxygen, -nitrogen, and -carbon bond strengths for organoruthenium and organoplatinum
    compounds; equilibrium studies of $Cp^*(PMe_3)_2RuX$ and $(DPPE)MePtX$ systems. J Am Chem
    Soc 109:1444–1456
Bukowski BC, Greeley J (2016) Scaling relationships for molecular adsorption and dissociation in
    Lewis acid zeolites. J Phys Chem C 120:6714–6722
Calle-Vallejo F, Martínez JI, García-Lastra JM, Rossmeisl J, Koper MTM (2012) Physical and
    chemical nature of the scaling relations between adsorption energies of atoms on metal surfaces.
    Phys Rev Lett 108:116103
Calle-Vallejo F et al (2013) Number of outer electrons as descriptor for adsorption processes on
    transition metals and their oxides. Chem Sci 4:1245–1249
Calle-Vallejo F, Martínez JI, García-Lastra JM, Sautet P, Loffreda D (2014) Fast prediction
    of adsorption properties for platinum nanocatalysts with generalized coordination numbers.
    Angew Chem Int Ed 53:8316–8319
Calle-Vallejo F, Loffreda D, Koper MTM, Sautet P (2015a) Introducing structural sensitivity into
    adsorption–energy scaling relations by means of coordination numbers. Nat Chem 7:403–410
Calle-Vallejo F et al (2015b) Finding optimal surface sites on heterogeneous catalysts by counting
    nearest neighbors. Science 350:185–189
Che M (2013) Nobel prize in chemistry 1912 to Sabatier: organic chemistry or catalysis? Catal
    Today 218–219:162–171
Cheng J, Hu P (2008) Utilization of the three-dimensional volcano surface to understand the
    chemistry of multiphase systems in heterogeneous catalysis. J Am Chem Soc 130:10868–10869
Christensen CH, Nørskov JK (2008) A molecular view of heterogeneous catalysis. J Chem Phys
    128:182503
Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density
    functional for general geometries. Phys Rev Lett 92:246401
Doyle AD, Montoya JH, Vojvodic A (2015) Improving oxygen electrochemistry through
    nanoscopic confinement. ChemCatChem 7:738–742
Doyle AD, Bajdich M, Vojvodic A (2017) Theoretical insights to bulk activity towards oxygen
    evolution in oxyhydroxides. Catal Lett 147:1533–1539
Ertl G (1980) Surface science and catalysis – studies on the mechanism of ammonia synthesis: the
    P. H. Emmett award address. Catal Rev Sci Eng 21:201–223
Evans MG, Polanyi M (1938) Inertia and driving force of chemical reactions. Trans Faraday Soc
    34:11–24
Fernández EM et al (2008) Scaling relationships for adsorption energies on transition metal oxide,
    sulfide, and nitride surfaces. Angew Chem Int Ed 47:4683–4686
Ferrin P et al (2009) Modeling ethanol decomposition on transition metals: a combined application
    of scaling and Brønsted−Evans−Polanyi relations. J Am Chem Soc 131:5809–5815

Gajdoš M, Eichler A, Hafner J (2004) CO adsorption on close-packed transition and noble metal surfaces: trends from ab initio calculations. J Phys Condens Matter 16:1141

García-Mota M, Vojvodic A, Abild-Pedersen F, Nørskov JK (2013) Electronic origin of the surface reactivity of transition-metal-doped TiO_2(110). J Phys Chem C 117:460–465

Göltl F, Müller P, Uchupalanun P, Sautet P, Hermans I (2017) Developing a descriptor-based approach for CO and NO adsorption strength to transition metal sites in zeolites. Chem Mater 29:6434–6444

Greeley J (2016) Theoretical heterogeneous catalysis: scaling relationships and computational catalyst design. Annu Rev Chem Biomol Eng 7:605–635

Greeley J, Mavrikakis M (2004) Alloy catalysts designed from first principles. Nat Mater 3: 810–815

Greeley J et al (2009) Alloys of platinum and early transition metals as oxygen reduction electrocatalysts. Nat Chem 1:552–556

Grimley TB (1967a) The indirect interaction between atoms or molecules adsorbed on metals. Proc Phys Soc 90:751–764

Grimley TB (1967b) The electron density in a metal near a chemisorbed atom or molecule. Proc Phys Soc 92:776

Grimley TB (1971) Electronic structure of adsorbed atoms and molecules. J Vac Sci Technol 8: 31–38

Grimme S (2004) Accurate description of van der Waals complexes by density functional theory including empirical corrections. J Comput Chem 25:1463–1473

Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem 27:1787–1799

Grimme S, Antony J, Schwabe T, Mück-Lichtenfeld C (2007) Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules. Org Biomol Chem 5:741–758

Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132:154104

Groß A (2008) Adsorption at nanostructured surfaces from first principles. J Comput Theor Nanosci 5:894–922

Halck NB, Petrykin V, Krtil P, Rossmeisl J (2014) Beyond the volcano limitations in electrocatalysis – oxygen evolution reaction. Phys Chem Chem Phys 16:13682–13688

Hammer B (2006) Special sites at noble and late transition metal catalysts. Top Catal 37:3–16

Hammer B, Norskov JK (1995a) Electronic factors determining the reactivity of metal surfaces. Surf Sci 343:211–220

Hammer B, Norskov JK (1995b) Why gold is the noblest of all the metals. Nature 376:238–240

Hammer B, Nørskov JK (2000) Theoretical surface science and catalysis – calculations and concepts. Adv Catal 45:71–129

Hansen HA, Rossmeisl J, Nørskov JK (2008) Surface Pourbaix diagrams and oxygen reduction activity of Pt, Ag and Ni(111) surfaces studied by DFT. Phys Chem Chem Phys 10:3722–3730

Hansen N, Kerber T, Sauer J, Bell AT, Keil FJ (2010) Quantum chemical modeling of benzene ethylation over H-ZSM-5 approaching chemical accuracy: a hybrid MP2: DFT study. J Am Chem Soc 132:11525–11538

Hansen HA, Varley JB, Peterson AA, Nørskov JK (2013) Understanding trends in the electrocatalytic activity of metals and enzymes for CO_2 reduction to CO. J Phys Chem Lett 4:388–392

Holby EF, Sheng W, Shao-Horn Y, Morgan D (2009) Pt nanoparticle stability in PEM fuel cells: influence of particle size distribution and crossover hydrogen. Energy Environ Sci 2: 865–871

Hong WT et al (2015) Toward the rational design of non-precious transition metal oxides for oxygen electrocatalysis. Energy Environ Sci 8:1404–1427

Hong X, Chan K, Tsai C, Nørskov JK (2016) How doped MoS_2 breaks transition-metal scaling relations for CO_2 electrochemical reduction. ACS Catal 6:4428–4437

Inderwildi OR, Jenkins SJ, King DA (2007) When adding an unreactive metal enhances catalytic activity: NO_x decomposition over silver–rhodium bimetallic surfaces. Surf Sci 601:L103–L108

Inoğlu N, Kitchin JR (2009) Atomistic thermodynamics study of the adsorption and the effects of water–gas shift reactants on Cu catalysts under reaction conditions. J Catal 261: 188–194

Jiang T et al (2009) Trends in CO oxidation rates for metal nanoparticles and close-packed, stepped, and kinked surfaces. J Phys Chem C 113:10548–10553

Jones G, Bligaard T, Abild-Pedersen F, Nørskov JK (2008) Using scaling relations to understand trends in the catalytic activity of transition metals. J Phys Condens Matter 20:064239

Jones G, Studt F, Abild-Pedersen F, Nørskov JK, Bligaard T (2011) Scaling relationships for adsorption energies of C_2 hydrocarbons on transition metal surfaces. Chem Eng Sci 66:6318–6323

Kibler LA, El-Aziz AM, Hoyer R, Kolb DM (2005) Tuning reaction rates by lateral strain in a palladium monolayer. Angew Chem Int Ed 44:2080–2084

Kitchin JR, Nørskov JK, Barteau MA, Chen JG (2004) Modification of the surface electronic and chemical properties of Pt(111) by subsurface 3d transition metals. J Chem Phys 120:10240–10246

Koper MTM (2013) Theory of multiple proton–electron transfer reactions and its implications for electrocatalysis. Chem Sci 4:2710–2723

Langmuir I (1961) In: Suits CG (ed) The collected works of Irving Langmuir. Pergamon Press Ltd, New York

Lee K, Murray ÉD, Kong L, Lundqvist BI, Langreth DC (2010) Higher-accuracy van der Waals density functional. Phys Rev B 82:081101

Lee Y-L et al (2016) Kinetics of oxygen surface exchange on epitaxial Ruddlesden–Popper phases and correlations to first-principles descriptors. J Phys Chem Lett 7:244–249

Liu B, Greeley J (2013) A density functional theory analysis of trends in glycerol decomposition on close-packed transition metal surfaces. Phys Chem Chem Phys 15:6475–6485

Liu B, Zhou M, Chan MKY, Greeley JP (2015) Understanding polyol decomposition on bimetallic Pt–Mo catalysts – a DFT study of glycerol. ACS Catal 5:4942–4950

Liu C, Tranca I, van Santen RA, Hensen EJM, Pidko EA (2017) Scaling relations for acidity and reactivity of zeolites. J Phys Chem C 121:23520–23530

Liu Z-P, Hu P (2001a) General trends in the barriers of catalytic reactions on transition metal surfaces. J Chem Phys 115:4977–4980

Liu Z-P, Hu P (2001b) General trends in CO dissociation on transition metal surfaces. J Chem Phys 114:8244–8247

Liu Z-P, Jenkins SJ, King DA (2004) Car exhaust catalysis from first principles: selective NO reduction under excess O_2 conditions on Ir. J Am Chem Soc 126:10746–10756

Logadottir A et al (2001) The Brønsted–Evans–Polanyi relation and the volcano plot for ammonia synthesis over transition metal catalysts. J Catal 197:229–231

Londero E et al (2012) Desorption of n-alkanes from graphene: a van der Waals density functional study. J Phys Condens Matter 24:424212

Ma X, Xin H (2017) Orbitalwise coordination number for predicting adsorption properties of metal nanocatalysts. Phys Rev Lett 118:036101

Man IC et al (2011) Universality in oxygen evolution electrocatalysis on oxide surfaces. ChemCatChem 3:1159–1165

Marković NM, Ross PN (2002) Surface science studies of model fuel cell electrocatalysts. Surf Sci Rep 45:117–229

Medford AJ et al (2015) From the Sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. J Catal 328:36–42

Menning CA, Chen JG (2009) General trend for adsorbate-induced segregation of subsurface metal atoms in bimetallic surfaces. J Chem Phys 130:174709

Michaelides A et al (2003) Identification of general linear relationships between activation energies and enthalpy changes for dissociation reactions at surfaces. J Am Chem Soc 125:3704–3705

Mills G, Gordon MS, Metiu H (2003) Oxygen adsorption on Au clusters and a rough Au(111) surface: the role of surface flatness, electron confinement, excess electrons, and band gap. J Chem Phys 118:4198–4205

Montemore MM, Medlin JW (2014) Scaling relations between adsorption energies for computational screening and design of catalysts. Cat Sci Technol 4:3748–3761

Montoya JH, Tsai C, Vojvodic A, Nørskov JK (2015) The challenge of electrochemical ammonia synthesis: a new perspective on the role of nitrogen scaling relations. ChemSusChem 8:2180–2186

Montoya JH, Doyle AD, Nørskov JK, Vojvodic A (2018) Trends in adsorption of electrocatalytic water splitting intermediates on cubic ABO_3 oxides. Phys Chem Chem Phys 20:3813–3818

Mun BS et al (2005) A study of electronic structures of Pt_3M (M=Ti,V,Cr,Fe,Co,Ni) polycrystalline alloys with valence-band photoemission spectroscopy. J Chem Phys 123:204717

Newns DM (1969) Self-consistent model of hydrogen chemisorption. Phys Rev 178:1123–1135

Nikolla E, Schwank J, Linic S (2009) Measuring and relating the electronic structures of nonmodel supported catalytic materials to their performance. J Am Chem Soc 131:2747–2754

Nørskov JK et al (2002) Universality in heterogeneous catalysis. J Catal 209:275–278

Nørskov JK et al (2004) Origin of the overpotential for oxygen reduction at a fuel-cell cathode. J Phys Chem B 108:17886–17892

Nørskov JK, Scheffler M, Toulhoat H (2006) Density functional theory in surface science and heterogeneous catalysis. MRS Bull 31:669–674

Nørskov JK, Bligaard T, Rossmeisl J, Christensen CH (2009) Towards the computational design of solid catalysts. Nat Chem 1:37–46

Nørskov JK, Abild-Pedersen F, Studt F, Bligaard T (2011) Density functional theory in surface chemistry and catalysis. Proc Natl Acad Sci U S A 108:937–943

Nørskov JK, Studt F, Abild-Pedersen F, Bligaard T (2014) Fundamental concepts in heterogeneous catalysis, John Wiley & Sons, Hoboken, NJ, USA

Peterson AA, Nørskov JK (2012) Activity descriptors for CO_2 electroreduction to methane on transition-metal catalysts. J Phys Chem Lett 3:251–258

Plessow PN, Abild-Pedersen F (2015) Examining the linearity of transition state scaling relations. J Phys Chem C 119:10448–10453

Reuter K, Scheffler M (2001) Composition, structure, and stability of $RuO2$ (110) as a function of oxygen pressure. Phys Rev B 65:035406

Reuter K, Scheffler M (2003) First-principles atomistic thermodynamics for oxidation catalysis: surface phase diagrams and catalytically interesting regions. Phys Rev Lett 90:046103

Reuter K, Frenkel D, Scheffler M (2004) The steady state of heterogeneous catalysis, studied by first-principles statistical mechanics. Phys Rev Lett 93:116105

Roling LT, Abild-Pedersen F (2017) Structure-sensitive scaling relations: adsorption energies from surface site stability. ChemCatChem 10:1643–1650

Román-Pérez G, Soler JM (2009) Efficient implementation of a van der Waals density functional: application to double-wall carbon nanotubes. Phys Rev Lett 103:096102

Rossmeisl J, Logadottir A, Nørskov JK (2005) Electrolysis of water on (oxidized) metal surfaces. Chem Phys 319:178–184

Rossmeisl J, Nørskov JK, Taylor CD, Janik MJ, Neurock M (2006) Calculated phase diagrams for the electrochemical oxidation and reduction of water over Pt(111). J Phys Chem B 110:21833–21839

Roudgar A, Groß A (2003) Local reactivity of thin Pd overlayers on Au single crystals. J Electroanal Chem 548:121–130

Sabatier P (1913) (1854–1941) A. du texte. La catalyse en chimie organique, par Paul Sabatier, C. Béranger

Salciccioli M, Chen Y, Vlachos DG (2010) Density functional theory-derived group additivity and linear scaling methods for prediction of oxygenate stability on metal catalysts: adsorption of open-ring alcohol and polyol dehydrogenation intermediates on Pt-based metals. J Phys Chem C 114:20155–20166

Schlögl R (2003) Catalytic synthesis of ammonia – a "never-ending story"? Angew Chem Int Ed 42:2004–2008

Schnur S, Groß A (2010) Strain and coordination effects in the adsorption properties of early transition metals: a density-functional theory study. Phys Rev B 81:033402

Schock LE, Marks TJ (1988) Organometallic thermochemistry. Metal hydrocarbyl, hydride, halide, carbonyl, amide, and alkoxide bond enthalpy relationships and their implications in pentamethylcyclopentadienyl and cyclopentadienyl complexes of zirconium and hafnium. J Am Chem Soc 110:7701–7715

Seh ZW et al (2017) Combining theory and experiment in electrocatalysis: insights into materials design. Science 355:eaad4998

Shustorovich E (1986) Bond making and breaking on transition-metal surfaces: theoretical projections based on bond-order conservation. Surf Sci 176:L863–L872

Shustorovich E, Bell AT (1988) Analysis of CO hydrogenation pathways using the bond-order-conservation method. J Catal 113:341–352

Suntivich J, May KJ, Gasteiger HA, Goodenough JB, Shao-Horn YA (2011) Perovskite oxide optimized for oxygen evolution catalysis from molecular orbital principles. Science 334:1383–1385

Svelle S, Tuma C, Rozanska X, Kerber T, Sauer J (2009) Quantum chemical modeling of zeolite-catalyzed methylation reactions: toward chemical accuracy for barriers. J Am Chem Soc 131:816–825

Tait SL, Dohnálek Z, Campbell CT, Kay BD (2006) n-alkanes on Pt(111) and on C(0001)/Pt(111): chain length dependence of kinetic desorption parameters. J Chem Phys 125:234308

Tripa CE, Zubkov TS, Yates JT, Mavrikakis M, Nørskov JK (1999) Molecular N_2 chemisorption – specific adsorption on step defect sites on Pt surfaces. J Chem Phys 111:8651–8658

Tuma C, Sauer J (2006) Treating dispersion effects in extended systems by hybrid MP2: DFT calculations – protonation of isobutene in zeolite ferrierite. Phys Chem Chem Phys 8:3955–3965

van Santen RA (2010) On Shustorovich's bond-order conservation method as applied to chemisorption. Recueil Trav Chim Paysâ-Bas 109:59–63

Viñes F, Vojvodic A, Abild-Pedersen F, Illas F (2013) Brønsted–Evans–Polanyi relationship for transition metal carbide and transition metal oxide surfaces. J Phys Chem C 117:4168–4171

Viswanathan V, Hansen HA, Rossmeisl J, Nørskov JK (2012) Universality in oxygen reduction electrocatalysis on metal surfaces. ACS Catal 2:1654–1660

Vojvodic A, Nørskov JK (2011) Optimizing perovskites for the water-splitting reaction. Science 334:1355–1356

Vojvodic A, Nørskov JK (2015) New design paradigm for heterogeneous catalysts. Natl Sci Rev 2:140–149

Vojvodic A, Hellman A, Ruberto C, Lundqvist BI (2009) From electronic structure to catalytic activity: a single descriptor for adsorption and reactivity on transition-metal carbides. Phys Rev Lett 103:146103

Vojvodic A et al (2011) On the behavior of Brønsted-Evans-Polanyi relations for transition metal oxides. J Chem Phys 134:244509

Vojvodic A, Nørskov JK, Abild-Pedersen F (2014a) Electronic structure effects in transition metal surface chemistry. Top Catal 57:25–32

Vojvodic A et al (2014b) Exploring the limits: a low-pressure, low-temperature Haber–Bosch process. Chem Phys Lett 598:108–112

Wang S, Omidvar N, Marx E, Xin H (2018) Coordination numbers for unraveling intrinsic size effects in gold-catalyzed CO oxidation. Phys Chem Chem Phys 20:6055–6059

Wang Y et al (2016) Scaling relationships for binding energies of transition metal complexes. Catal Lett 146:304–308

Wellendorff J et al (2012) Density functionals for surface science: exchange-correlation model development with Bayesian error estimation. Phys Rev B 85:235149

Wodrich MD, Busch M, Corminboeuf C (2016) Accessing and predicting the kinetic profiles of homogeneous catalysts from volcano plots. Chem Sci 7:5723–5735

Xin H, Vojvodic A, Voss J, Nørskov JK, Abild-Pedersen F (2014) Effects of *d*-band shape on the surface reactivity of transition-metal alloys. Phys Rev B 89:115114

Xu Y (2009) Recent advances in heterogeneous catalysis enabled by first-principles methods. In: Spivey JJ, Dooley KM (eds) Catalysis. R Soc Chem Lond 21:131–153

Zhang J et al (2005) Mixed-metal Pt monolayer electrocatalysts for enhanced oxygen reduction kinetics. J Am Chem Soc 127:12480–12481

Oxide Catalysts

55

Horia Metiu, Vishal Agarwal, and Henrik H. Kristoffersen

Contents

Abstract

This chapter reviews some aspects of computational work on the catalytic chemistry of oxides. The difficulties of using density functional theory in calculations are explained. Different ways of structural or chemical modifications aimed at improving catalytic activity are reviewed. The reaction mechanism of partial oxidation reaction catalyzed by oxides is discussed. The focus is on

H. Metiu (✉)
Department of Chemistry and Biochemistry, University of California at Santa Barbara, Santa Barbara, CA, USA
e-mail: metiu@chem.ucsb.edu

V. Agarwal
Department of Chemical Engineering, Indian Institute of Technology, Kanpur, Uttar Pradesh, India

H. H. Kristoffersen
DTU Energy, Technical University of Denmark, Kgs. Lyngby, Denmark

qualitative design rules rather than on obtaining highly accurate computational results.

1 Introduction

The number of publications on oxide catalysts is enormous and we can only attempt to provide a summary of those properties that seem to be common to classes of oxide catalysts and/or for classes of reactions. The main subject of this section is what computations teach us about catalysis by simple oxides having the formula M_xO_y. We do not discuss compounds having two cations such as perovskites, spinels, vanadates, etc. Most zeolites are oxides of Si and Al with peculiar, porous structures. We do not examine them here since their catalytic chemistry is different from that of ordinary oxides. Within these self-imposed limitations, we will discuss pure oxides, doped oxides, submonolayers of oxides supported on oxides, and oxide submonolayers supported on metals. Among reactions, we will mention oxidative dehydrogenation and methane oxidative coupling. The topic is so vast that more is left unsaid than is presented.

Oxide catalysts are used industrially for the synthesis of sulfuric acid, the conversion of hydrogen sulfide to sulfur, of methanol to formaldehyde, of o-xylene to phthalic anhydride, of ethylbenzene to styrene, and of propane to propylene.

2 Mars–van Krevelen Mechanism

In the majority of reactions catalyzed by oxides, the gaseous feed contains a reductant and an oxidant. Hydrocarbons, alcohols, CO, H_2, etc. are reductants, and O_2 N_2O or CO_2 are oxidants. There is extensive evidence that oxidation reactions catalyzed by oxides take place through a Mars–van Krevelen mechanism. This postulates that the reductant in the feed reacts with the surface oxygen atoms, removing them and creating oxygen vacancies on the surface; the oxidant reoxidizes the surface. For example, in the reaction $CO(g) + 1/2\,O_2(g) \rightarrow CO_2(g)$, CO reduces the oxide surface to form CO_2 and oxygen vacancies; one oxygen atom in the CO_2 molecule originates from the surface of the oxide. The role of the gas-phase oxygen is to reoxidize the surface. In a slightly more complicated reaction, such as $CH_3OH + \frac{1}{2}\,O_2(g) \rightarrow CH_2O + H_2O$ the oxygen in the water molecule is taken from the oxide surface and, after water desorbs (thus removing an oxygen atom from the surface), the gaseous oxygen reoxidizes the surface. The oxygen in the aldehyde is the oxygen atom contained in the methanol feed.

Transient kinetics experiments, experiments using $^{18}O_2$, and Density Functional Theory (DFT) calculations, all support the Mars–van Krevelen mechanism. However, it is bad practice to assume uncritically that all oxides work according to a Mars–van Krevelen mechanism, for all reactions, under all conditions. In some cases, the oxide might activate the adsorbed oxygen and this oxidizes the reductant.

According to the Mars–van Krevelen mechanism, the reductant removes oxygen from the surface of the oxide and the oxidant puts it back. Because of this, at steady state, there is a constant concentration of oxygen vacancies at the surface. The catalyst is the partially reduced oxide, not the stoichiometric oxide. For a given reductant, the surface of the catalyst is less reduced if the oxidant is O_2 and more reduced if the oxidant is CO_2. In many experiments using oxygen, it has been observed that the steady-state reaction rate is independent of the partial pressure of the oxygen. This indicates that the reoxidation of the surface by gaseous oxygen is much faster than the other reaction steps. This independence of the rate on the oxygen partial pressure does not mean that the reaction will proceed without gaseous oxygen. If the feed does not contain oxygen, the oxide will be reduced until the degree of reduction is so high that the oxide will stop providing oxygen and the oxidation reaction stops. At high temperature, this reduction may go all the way to the formation of the metal, an undesirable outcome unless one is a metallurgist.

3   Irreducible and Reducible Oxides

The Mars–van Krevelen mechanism suggests a classification of the oxide into those that provide oxygen easily (reducible oxides) and those do not (irreducible oxides). Examples of irreducible oxides are alkali oxides, alkali earth oxides, La_2O_3, Sm_2O_3, Y_2O_3, Sc_2O_3, ZrO_2, Ta_2O_5, etc. The term "irreducible" indicates that the oxide is difficult to reduce. It is reasonable to assume that an oxide that is difficult to reduce is not a good oxidant. This is why irreducible oxides are used as catalysts for partial-oxidation reactions at high temperature. An example is the oxidative methane coupling, $2 CH_4 + 1/2 O_2 \rightarrow C_2H_4 + H_2O$, which takes place at $\sim 800\,°C$ (Zavyalova et al. 2011). At this temperature, a reducible oxide will combust methane to CO_2 and water, or to CO and H_2, depending on the working conditions. A characteristic common to all irreducible oxides is that their cation does not make stable oxides having different cation-to-oxygen ratios. For example, in the case of La_2O_3, La does not make stable LaO or LaO_2 oxides. However, as in many cases in chemistry, the rules have exceptions. The fact that a cation forms only one oxide is not necessarily an indication that the oxide is difficult to reduce: for example, NiO and ZnO do not form an oxide containing Ni^+ or Zn^+ but they can be reduced much more easily than, for example, MgO.

Reducible oxides, such as TiO_2, CeO_2, V_2O_5, MoO_3, Mn_3O_4, WO_3, CrO_2, are grouped together because they are better oxidants than the other oxides. In other words, it is easier to partially reduce them by making oxygen vacancies on their surface. The elements in these oxides are able to make oxides having a lower oxygen content (per cation), such as Ti_2O_3, Ce_2O_3, V_2O_3, VO_2, MnO_2, Mn_2O_3, MoO_2, WO_2, Cr_2O_3. Some of them form a large number of crystalline oxides of intermediate stoichiometry. For example, vanadium forms V_nO_{2n+1} ($n = 3, 4, 5$) and V_nO_{2n-1} ($n = 4, 5, 6, 7, 8, 9$). Similar compounds are formed by Mo, Ce, and Ti. The existence of such compounds suggests that it is easy to induce these cations to release oxygen and lower their valence.

4 Oxides Whose Cations Have an Incomplete d- or f-Electronic Shell

There is another important classification of oxides, based on whether the d-shell or the f-shell of the cation is filled. The cations in rows 1, 2, and 11–16 in the periodic table have filled d- or f-shells. The rest of the cations do not, and this causes difficulties: calculations using the generalized gradient approximation of DFT (GGA-DFT) make substantial, qualitative errors. A well-studied example of an oxide having a cation with an incomplete d-shell is TiO_2; equally well studied is CeO_2, in which Ce has an incomplete f-shell. A number of things go wrong when GGA-DFT is used to calculate the properties of these oxides. (1) The d-band (or the f-band) is much too wide (i.e., the energy of the d- or f-orbitals changes substantially when the wave vector is changed) whereas experiments show them to be narrow. The narrow band detected by experiments indicates the d (or the f) electrons are localized on the cations. GGA predicts the opposite: electrons located in d orbitals are delocalized. (2) GGA does not accurately describe processes in which one or two electrons are donated to the oxide. Such donation takes place, for example, when radicals such as H or CH_3 bind to the oxide, or when an oxygen vacancy is formed. Experiments indicate that the donated electrons are localized on one cation while GGA predicts that they are delocalized. (3) In most processes in which one or two electrons are donated to the surface, GGA-DFT predicts that the energy of the orbital in which such electrons are located is at the bottom of the valence band. However, ultra-violet photoelectron spectroscopy (UPS) experiments show that their energy is in the band gap. No such difficulties are encountered when GGA is used for oxides with a complete d- or f-shell (e.g., MgO).

This tendency of GGA to delocalize electrons is attributed to a flaw in the DFT method. The DFT energy contains a term in which each electron interacts with itself. This energy is diminished if the electron distribution is diffuse rather than localized. Hence, the variational procedure will lower the energy by spreading out the electron distribution.

To correct the behavior of GGA, when one calculates the properties of such oxides, one introduces in the DFT energy a term U designed to increase the energy when more than one d or f electron is present on a cation. As a result, the d or f electrons are localized on the cations, as they should be. The theory containing this term, called GGA + U, is an improvization that does not have a firm theoretical foundation and does not provide a reliable prescription for the magnitude of U. U is determined to fit some property of interest. Chemist may pick U to give correct energy of formation of the oxide. People interested in photochemistry at the surface pick U to obtain the experimentally measured band gap. There are also recipes for calculating U from the atomic properties of the cation, but they are approximate. In some cases, people adjust the value of U, in GGA + U, to give results close to the results of hybrid functionals (e.g., B3LYP or HSE), which are deemed to be more accurate. In general, it is prudent, when using GGA + U, to perform calculations for several values of U in order to confirm that at least the qualitative conclusions of the study are independent of the magnitude of U.

The hybrid density functionals (e.g., B3LYP, HSE) give more reliable results than GGA + U. These methods add a fractional amount of Hartree–Fock exchange energy to the ordinary exchange term in the Kohn–Sham equation. This diminishes, to some extent, the self-interaction error that plagues GGA-DFT. The fraction of Hartree–Fock exchange has been determined by fitting a data set. The method works well for molecules and the consensus is that hybrid functionals are more reliable than GGA + U, when applied to oxides. Unfortunately, hybrid functional calculations require substantially more computer time than GGA + U. Since the study of a catalytic process requires many calculations, on systems having many electrons, the hybrid functionals are not used in "production runs." It is more common to use them to validate (or not) some essential step in a reaction mechanism predicted by GGA + U.

5   Oxygen Vacancy Formation

The Mars–van Krevelen mechanism assumes that the reductant in the feed reacts with surface oxygen, making oxygen vacancies. One would expect that oxides for which the energy of oxygen vacancy formation is the lowest are the best oxidation catalysts. Such oxides are likely to be the better oxidants, but they are not necessarily the best partial oxidation catalysts. For example, one would not try to use HgO as a catalyst; it reacts vigorously with reductants but it does not reoxidize easily. Nevertheless, barring a few exceptions, the energy of vacancy formation is not a bad descriptor for a preliminary screening of an oxide catalyst for partial or total oxidation.

It is very difficult to obtain reliable experimental information about the energy of oxygen vacancy formation at the surface of an oxide. Our search of the experimental literature, for the energy of oxygen-vacancy formation for CeO_2, found that numerous values between 0.5 and 4.5 eV have been proposed. There are several reasons for the discrepancies between various measurements. (1) The energy of vacancy formation seems to depend on crystallite size. Therefore, different methods of oxide preparation give different results for the energy of vacancy formation. (2) The energy of oxygen vacancy formation ought to depend strongly on vacancy concentration. As one removes a few oxygen atoms, one expects that the cations will bind more tightly the remaining oxygen. (3) The energy of vacancy formation is affected by the presence of small dopant concentrations. Even research grade oxides have 99.99% purity. In the semiconductor industry, the presence of 0.01% dopants is considered unacceptable since it could dramatically change some properties (e.g., the electrical conductivity). Calculations show that the presence of dopants whose valence is lower than that of the cations they substitute will lower substantially the energy of oxygen-vacancy formation.

Given the paucity of conclusive experimental data, the calculation of the energy of oxygen-vacancy formation plays an important role in our qualitative understanding of the properties of oxide catalysts. While the absolute values of the calculated energy are unreliable, it is likely that they can be used to compare

the oxidative power of different oxides. The calculated energy of oxygen-vacancy formation, when done properly, is consistent with the oxidizing activity observed in experiments.

When an oxygen atom is removed from the oxide surface, it leaves behind (formally) two electrons that used to be tied in the bonds between the oxide and the oxygen atom being removed. The energy of vacancy formation depends on the strength of metal–oxygen chemical bond and on the fate of these "stranded electrons." If the oxide is reducible, these two electrons will reduce two cations. For example, in the case of TiO_2 the removal of an oxygen atom creates two Ti^{3+} ions. We are using here what inorganic chemists call "formal charges," a term that facilitates communication but should not be taken literally. The calculated Bader charge on a Ti atom, in TiO_2, is not 4+ nor is the charge on oxygen equal to -2. A more precise statement is that when an oxygen vacancy is formed the Bader charge on two Ti ions increases by almost one electron. The Ti^{3+} ion, together with the distortion of the oxygen atom surrounding it, is called a *polaron*. Polaron formation lowers the energy of the system (as compared to the case in which the electrons are stuck in the vacancy). It is expected that in all reducible oxides, the formation of an oxygen vacancy, at the surface or in the bulk, is accompanied by the formation of two polarons. This phenomenon has been documented for TiO_2, CeO_2, MoO_3, V_2O_5, and WO_3.

The energy of the polarons formed when an oxygen vacancy is created depends on the nature of the oxide, the position of the polarons with respect to the vacancy site, and the positions of the polaron with respect to each other. This means that to determine the lowest energy for oxygen-vacancy formation, one must study all possible locations of the two polarons.

Polarons do not form spontaneously when an oxygen atom is removed to form an oxygen vacancy. If the oxygen atom is removed and the energy is minimized, two Ti^{3+} ions are formed at the vacancy site. To form a polaron on a specific Ti ion, one must push the nearest oxygen atoms away from the cation and optimize the energy with these oxygen atoms held fixed. When a minimum energy is reached, one releases the oxygen atoms that were held in place during the first optimization, and one optimizes the geometry again. This will form a polaron at the desired location. If DFT + U is used, polarons are generally stable if U is larger than 2.5 eV (or thereabouts, depending on the oxide). The larger the value of U, the more stable is the polaron.

When DFT + U is used, the energy of the orbital involved in the polarons is in the band gap, in agreement with photoelectron experiments. Stoichiometric, pure TiO_2 has a large band gap and does not conduct electricity. However, if oxygen vacancies are created the material becomes an n-conductor. It is widely accepted that in the case of TiO_2 the electric conductivity of the oxide is due to polaron hopping from site to site. The energy barrier of site-to-site hopping has been calculated (Deskins and Dupuis 2007) to be ~ 0.25 eV. This is consistent with conductivity experiments. The band gap of TiO_2 is ~ 2 eV and the states in the gap are below the conduction band by ~ 0.7 eV. Therefore, the electrical conductivity is not due to thermal promotion of an electron from a state in the gap to the conduction band.

The only possible explanation for the conductivity by n-carrier is the mobility of the polarons. This is also true if electrons are injected in the material by a cathode: the electron will form a Ti^{3+} with a polaronic distortion and the electric conductivity is due to polaron migration. The polarons caused by oxygen-vacancy formation have been detected by electron paramagnetic resonance (EPR) spectroscopy.

Experiments in Moskovits's group (Zhang et al. 2004) measured the conductivity of a SnO_2 nanowire that was reduced by CO and reoxidized by O_2. The measurements have shown that reduction increases the electrical conductivity of the wire and oxidation reduces it, as expected if polarons are formed during reduction. We are not aware of a computational study of oxygen vacancies in this oxide, which differs from TiO_2 because Sn^{3+} is unlikely to be formed.

Why is this relevant to surface chemistry? Polaron formation explains why it is much easier to form oxygen vacancies on reducible oxides. This in turn explains why such oxides are better oxidants than the irreducible ones. Furthermore, the polaron are strong Lewis bases (electron donor). Their presence facilitates the adsorption of Lewis acids, such as Au atoms, O_2, Cl_2, Br_2, and I_2.

Polarons are also invoked (in the case of reducible oxides) when H adsorption or CH_3 adsorption is studied. The binding energy of these species to reducible oxides is increased by the fact that they donate an electron and form a polaron.

The formation of oxygen vacancies on irreducible oxides does not result in polaron formation. The "stranded electrons" remain at the vacancy site where they act as strong Lewis bases, which bind Lewis acids strongly.

6 Two-Step Oxidation–Reduction

There has been quite a bit of experimental work studying the possibility of a two-step process for conversion of methane to syngas. It turns out that at high temperature many oxides react with methane to produce syngas and a reduced oxide. In a second step, the oxide can be regenerated by reacting with water, to produce H_2 and a fully oxidized oxide. An alternative second step will oxidize the reduced oxide with CO_2 to produce a stoichiometric oxide and CO. The degree of reduction must be monitored carefully to remove only a few layers of oxygen atoms from the surface of each oxide nanoparticle, so that the surface does not become purely metallic. This partial reduction will avoid coarsening and the cracking of the solid particles due to the stress created by a large change of density when cycling between a stoichiometric oxide and a strongly reduced one.

7 The Lewis Acid–Base Rule

The catalytic chemistry of oxides is very complex and it is therefore useful to have some design rules that give general guidance regarding how one might modify an oxide surface to change its catalytic chemistry. Useful in this regard is the Lewis acid–base rule that is stated below. The interaction energy between two species

adsorbed on an oxide surface is increased substantially if one species can function as a Lewis acid (electron acceptor) and the other as a Lewis base (electron donor). A consequence of this rule is that two species that have several options for binding sites (e.g., could bind to the cation or to the oxygen) would prefer the binding configuration that allows one to be a Lewis acid and the other a Lewis base. For example, if they are alone on an oxide surface, H or CH_3 each prefers to bind to an oxygen atom. However, when they are coadsorbed, the energy is lower if H binds to O and donates an electron to a CH_3 bound to the cation. In this case, H is a Lewis base and CH_3 is a Lewis acid. This acid–base rule seems to be general: it has been documented by numerous calculations for the dissociative adsorption of CH_4, H_2, HX and X_2 (X = Cl or Br), on the surface of a variety of oxides (e.g., MgO, La_2O_3, CaO, TiO_2, CeO_2).

8   Doped Oxides

This acid–base rule suggested a number of strategies for modifying oxide surface to make them better oxidants. One of them is doping (Metiu et al. 2012; McFarland and Metiu 2013; Nilius and Freund 2015): replacing some of the cations near or at the surface of the oxide with cations having a lower valence (e.g., MgO doped with Li). This creates a deficit of electrons in the oxide, making the surface an electron acceptor (Lewis acid), which enhances the ability of the surface to bind Lewis bases such as H and CH_3. The chemistry of the doped oxide does not involve the dopant directly. By not supplying the oxide with enough electrons, the dopant weakens the bond of one of the surface oxygen atoms to the oxide, making the atom more reactive. The use of a higher-valence dopant (e.g., MgO doped with La) has the opposite effect. Formally, the dopant is capable of donating three electrons but the oxide will only use two of them (because we replace a divalent cation in oxide with a trivalent element). The unwanted electron is stuck on to the dopant, making it a good Lewis base. According to the acid–base rule, this site will adsorb well a Lewis acid such as oxygen. In this example, the oxygen becomes negatively charged and is chemically more active than neutral O_2.

It is difficult to prove by experiments with powder catalysts that these concepts work as expected, mainly because one is not certain where the dopants are located (in the surface layer, in the second layer, in the bulk?). It is also not known whether the dopant is oxidized during reaction to form an isolated oxide cluster (the oxide of the dopant). Resolving this issue is made difficult, among other reasons, by the fact that the chemistry of a submonolayer of oxidized dopant will not behave chemically like the bulk oxide of the dopant. In a few experiments, electron microscopy has detected isolated dopants but the mechanism through which they act as catalyst cannot be verified experimentally. Most experiments find that the use of lower valence doping does make the oxide a better oxidant. Careful experiments in Freund's laboratory, working in ultra-high vacuum, with well-defined systems characterized by scanning tunneling microscopy (STM) are generally in agreement with the acid–base rules.

9   Oxides Supported on Other Oxides

Several recent articles (Wachs 2013, 2005; Stacchiola et al. 2013; Carrero et al. 2014; Freund et al. 2014) have reviewed catalysis by systems consisting of an oxide submonolayer supported on another oxide. Any oxide can be used as a submonolayer if it "wets" the support. Broadly speaking the strategy is the same as in doping: create a new bonding scheme in the system, with the goal of changing reactivity. Of course, not every new bonding is beneficial and a central goal of research is to find the ones that are. Experience indicates that the catalytic chemistry is more interesting if the submonolayer is a reducible oxide (e.g., vanadium or molybdenum oxide) but this is not an absolute rule. We examine, as an example, systems in which the submonolayer is V_xO_y. We use this symbol because we are not certain about the stoichiometry of the submonolayer. An exception is provided by experiments (Price et al. 2014) in which mass-selected V_xO_y clusters are deposited, in ultra-high vacuum, on $TiO_2(110)$ surface, and are examined in situ, by scanning tunneling microscopy (STM). Thermal desorption experiments were performed on these systems to determine the catalytic conversion of methanol to formaldehyde for different V_xO_y stoichiometry. Experiments on well-defined systems (Freund et al. 2014; Nilius and Freund 2015) are particularly important when comparison with computations is desired.

The qualitative idea in the design of these submonolayer systems is to force some oxygen atoms to have an unusual bonding. For example, in V_xO_y supported on TiO_2, some oxygen atoms form -Ti-O-V- groups (the lines indicate chemical bonds). The reactivity of these "bridging oxygen atoms" should be different from that of the oxygen atoms in TiO_2 or in V_2O_5. Experiments show that they are. By studying the oxidative dehydrogenation of propane to propylene (Carrero et al. 2014), or of methanol to formaldehyde (Wachs 2013), the following conclusions were drawn. (1) The V_xO_y submonolayer is more active than V_2O_5 powder or TiO_2 powder. The catalytic activity decreases when the amount of vanadium used in the preparation of the catalyst is sufficient to build a second vanadium oxide layer. Moreover, the selectivity changes: pure V_2O_5 is a good total combustion catalyst, while supported V_xO_y is effective for propane or methanol oxidative dehydrogenation. (2) The effect of the support is surprisingly strong. The turnover frequency on VO_x supported on CeO_2 or on TiO_2 is roughly three orders of magnitude higher than that on VO_x supported on silica or alumina. (3) One can prepare a catalyst having a very low concentration of vanadium, in which case one has isolated VO_4 clusters on the surface. Adding more vanadium leads to the formation of clusters having several V atoms connected by -V-O-V- bridges. The oxygen atoms in these bridges do not seem to have high reactivity. (4) Under working conditions, at atmospheric pressure, the catalyst is covered with reactant, intermediate, and product molecules, in addition to having an unknown and probably inhomogeneous morphology. Possible adsorbed species are C_3H_8, C_3H_7, C_3H_6, hydroxyls, oxygen vacancies, and all the intermediates of the combustion of carbon-containing species. This complexity, in surface composition and morphology, makes it difficult to obtain a detailed understanding of these catalysts (oxide submonolayer on oxide).

An important task is to develop a qualitative understanding of the V_xO_y interaction with the support.

10   Inverse Catalysts

Many catalysts used in industry or research consist of metal nanoparticles supported on an oxide. In many cases, the oxide support confers mechanical stability, inhibits coarsening, helps create a large surface for the metal catalyst, etc. In some cases, the oxide support affects the performance substantially. For example, the best support for the Ag catalyst for ethylene epoxidation is α-alumina or SiC. Other supports, such as SiO_2, MgO, TiO_2, Y_2O_3, ZrO_2, have been tried and found inadequate. Even Ag supported on γ-Al_2O_3 has poor performance because γ-alumina is more acid than α-alumina. In some cases, the oxide support covers the metal particle with a thin oxide layer. Often it seems that catalysis takes place at the border between the oxide support and the metal nanoparticle. It is natural therefore to try to create such a border differently, by using an oxide nanoparticle supported on a metal. Catalysts designed in this way are called inverse catalysts and their properties have been studied and reviewed in several publications (Hornes et al. 2010; Rodriguez and Hrbek 2010; Senanayake et al. 2010; Freund et al. 2014; Pan et al. 2014; Rodriguez et al. 2017). Inverse here means opposite to the more common catalysts that consist of metal particles on an oxide support.

11   Summary and Outlook

Several facts make computational studies of real catalysts, oxides or otherwise, difficult. In spite of great advances in in situ electron microscopy and the spectroscopy of catalysts under working conditions, we do not know in detail the morphology (steps, kink, facets) and the composition of the surface of working catalysts. At atmospheric pressure the surface is covered with reactants, products, and intermediates, which compete for surface sites and often interact strongly with each other. For example, the adsorption of O_2 is enhanced substantially by the presence of hydroxyls. While the hydroxyls might not participate in the mechanism of an oxidation reaction, they might influence it through their effect on oxygen adsorption. In many cases the support of the catalyst is important and it is possible that the border between the oxide catalyst and the support is catalytically active. Modeling such a border is difficult. The models we use to study catalysts are almost always a simplified version of reality. The divorce from reality is amplified by the fact that almost all practical catalysts contain small amounts of promoters of the desired reaction and inhibitors of unwanted reactions. Such small additives make a very substantial difference in a catalyst's performance.

A study of catalysis requires the examination of systems with many atoms, which could have a great variety of positions. Because of this, computational efficiency is an essential requirement. That requirement is the reason GGA-DFT + U is the

workhorse in the field. Hybrid functionals require substantially more computer power and are mostly used to recalculate some of the critical steps in a mechanism derived by using GGA calculations. The validation of these calculations by experiment is difficult: there are few groups that can perform experiments with well-defined, single-crystal, oxide surfaces in ultra-high vacuum. Careful measurements of adsorption isotherms, for real catalysts, would be helpful but they are out of fashion. Validation of the DFT calculations by using high-accuracy quantum chemistry methods is, in principle, possible; however, modeling catalysts requires dealing with a large number of electrons and a large number of geometries, and demands too much computer time. Modeling a catalyst by a very small cluster, for which high-quality calculations are possible, is often not satisfactory. Progress in the near future is likely to come from the development of new functionals designed specifically for dealing with oxides. Another promising possibility is the development of embedded cluster methods, to accurately treat a critical group of atoms and more poorly (by DFT or classical potentials) treat the rest.

One might think that the morphology and the composition of the surface could be determined by finding the structure having the lowest energy. However, catalysis is performed at steady state and the morphology and the composition of the surface under working conditions is controlled by kinetics, not by thermodynamics. Kinetic Monte Carlo simulations could be attempted but the number of rate events is very large and the calculated rate constants are likely to have substantial errors. DFT calculations of activation energies are not accurate and unfortunately the Arrhenius formula amplifies the errors.

To summarize: if we intend to study real catalysts, we are forced to use an inaccurate computational method (DFT), in order to examine inaccurate models of systems that experiments do not describe in detail. One might have to give up, temporarily, the dream of performing a complete ab initio description of a catalytic process and accept that computation is one of the tools we have along with spectroscopy and kinetics. This frees the computational chemist to focus on those questions of the catalytic process that the experiments cannot settle. A great strength of the theory is that it has a complete control of morphology and composition. Rather than try to imitate real catalysts, theory should create, on the computer, new structures and new compositions to find the ones that have very promising properties.

References

Carrero CA, Schloegl R, Wachs IE, Schomaecker R (2014) Critical literature review of the kinetics for the oxidative dehydrogenation of propane over well-defined supported vanadium oxide catalysts. ACS Catal 4:3357–3380

Deskins NA, Dupuis M (2007) Electron transport via polaron hopping in bulk TiO_2: a density functional theory characterization. Phys Rev B 75:195212

Freund H-J, Shaikhutdinov S, Nilius N (2014) Model studies on heterogeneous catalysts at the atomic scale. Top Catal 57:822–832

Hornes A, Hungria AB, Bera P, Camara AL, Fernandez-Garcia M, Martinez-Arias A, Barrio L, Estrella M, Zhou G, Fonseca JJ, Hanson JC, Rodriguez JA (2010) Inverse CeO_2/CuO catalyst as an alternative to classical direct configurations for preferential oxidation of Co in hydrogen-rich stream. J Am Chem Soc 132:34–35

McFarland EW, Metiu H (2013) Catalysis by doped oxides. Chem Rev 113:4391–4427

Metiu H, Chrétien S, Hu Z, Li B, Sun X (2012) Chemistry of Lewis acid-base pairs on oxide surfaces. J Phys Chem C 116:10439–10450

Nilius N, Freund HJ (2015) Activating nonreducible oxides via doping. Acc Chem Res 48:1532–1539

Pan Q, Liu BH, McBriarty ME, Martynova Y, Groot IMN, Wang S, Bedzyk MJ, Shaikhutdinov S, Freund HJ (2014) Reactivity of ultra-thin ZnO films supported by Ag(111) and Cu(111): a comparison to ZnO/Pt(111). Catal Lett 144:648–655

Price SP, Tong X, Ridge C, Neilson HL, Buffon JW, Robins J, Metiu H, Bowers MT, Buratto SK (2014) Catalytic oxidation of methanol to formaldehyde by mass-selected vanadium oxide clusters supported on a TiO_2(110) surface. J Phys Chem A 118:8309–8313

Rodriguez JA, Hrbek J (2010) Inverse oxide/metal catalysts: a versatile approach for activity tests and mechanistic studies. Surf Sci 604:241–244

Rodriguez JA, Grinter DC, Liu Z, Palomino RM, Senanayake SD (2017) Ceria-based model catalysts: fundamental studies on the importance of the metal-ceria interface in co oxidation, the water-gas shift, CO_2 hydrogenation, and methane and alcohol reforming. Chem Soc Rev 46:1824–1841

Senanayake SD, Stacchiola D, Evans J, Estrella M, Barrio L, Perez M, Hrbek J, Rodriguez JA (2010) Probing the reaction intermediates for the water-gas shift over inverse CeO_x/Au(111) catalysts. J Catal 271:392–400

Stacchiola DJ, Senanayake SD, Liu P, Rodriguez JA (2013) Fundamental studies of well-defined surfaces of mixed-metal oxides: special properties of MO_x/TiO_2(110) {M = V, Ru, Ce, or W}. Chem Rev 113:4373–4390

Wachs IE (2005) Recent conceptual advances in the catalysis science of mixed metal oxide catalytic materials. Catal Today 100:79–94

Wachs IE (2013) Catalysis science of supported vanadium oxide catalysts. Dalton Trans 42:11762–11769

Zavyalova U, Holena M, Schlögl R, Baerns M (2011) Statistical analysis of past catalytic data on oxidative methane coupling for new insights into the composition of high-performance catalysts. ChemCatChem 3:1935–1947

Zhang Y, Kolmakov A, Chrétien S, Metiu H, Moskovits M (2004) Control of catalytic reactions at the surface of a metal oxide nanowire by manipulating electron density inside it. Nano Lett 4:403–407

Supercell Models of Brønsted and Lewis Sites in Zeolites

56

Sichi Li and William F. Schneider

Contents

Abstract

Zeolites are three-dimensional, crystalline silicate-based materials of interest for catalysis and separations. Computational models of zeolites must capture their three-dimensional structure, the intrinsic microscopic heterogeneity introduced by heteroatom substitutions that underlie their interesting chemical behavior, and the dynamic nature of reactive sites within the pores of molecular dimensions. Here we describe the use of supercell density functional theory (DFT) models for tackling these problems, focusing primarily on Brønsted acidic and Cu-exchanged chabazite (CHA) zeolites and their chemistry related to the catalytic

S. Li · W. F. Schneider (✉)
Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA
e-mail: sli12@nd.edu; wschneider@nd.edu

chemistry of nitrogen oxides (NO_x). We describe considerations important in model construction, applications of *ab initio* molecular dynamics to structure annealing and accurate computations of reaction and activation free energies, first-principles thermodynamics approaches for predicting site compositions at realistic conditions, and approaches for incorporating heteroatom distributions into material models.

1 Introduction

Zeolites are a general class of crystalline silicate-based materials (Auerbach et al. 2003). The presence of micropores and voids of fixed, molecular dimension and the ability to tailor through chemical functionalization make them broadly useful as catalysts ranging from fluid catalytic cracking (Corma 1989) to hydrocarbon oligomerization (Nicholas 2017) to selective catalytic reduction (SCR) of NO_x with NH_3 (Paolucci et al. 2016a). The primary building unit of a zeolite is an SiO_4 tetrahedron (or T-site), which can be linked together at the corners to create a variety of secondary building units, including rings and cages of various sizes, that are further organized into distinct three-dimensional, periodic structures. The topology of a zeolite determines the sizes and dimensionalities of pores that permit molecular access into the interior of the zeolite as well as the sizes and interconnectivities of any cages. The identity of a zeolite is further nuanced by its chemical composition, including the types, concentrations, and locations of any framework or extra-framework heteroatoms. Depending on the synthesis method and treatment history, heteroatoms can be distributed in different ways, e.g., substituting framework Si (Lopez et al. 1993; Matsubayashi et al. 1996), cationically charge-compensating anionic framework blocks (Sherry 1966; Colella 1996), or even forming aggregated metal clusters or metal oxides inside the zeolite cages (Melson et al. 1983). Substitution of framework Si by Al is the most common chemical modification of a zeolite. Due to the 3+ oxidation state of Al, each substitution of a 4+ Si introduces a negatively charged site that can in principle host an extra-framework cation, e.g., a proton, alkali metal, or transition metals. These substituents introduce distinct chemical functionality into the zeolite, such as Lewis acidity or basicity, Brønsted acidity, and/or redox activity, all relevant to catalysis by enhancement of adsorption and/or bond activation of reactant(s) and intermediate(s). How to accurately model the electronic properties of different cations in zeolites from first-principles has been a popular topic in the computational catalysis community for many years (Sauer et al. 1980; Trout et al. 1996; Schneider et al. 1998; Gonzales et al. 1998; Sierka and Sauer 2001; Tuma and Sauer 2006; Paolucci et al. 2016b).

The two common approaches to represent a zeolite within an atomistic simulation are cluster and supercell models. The basic idea of cluster model is to excise the relevant portion of the zeolite framework (a cluster) out of the three-dimensional structure. These cluster representations are amenable to treatment with standard local-basis-set quantum mechanical (QM) methods (Kramer et al. 1991). To recover the effect of zeolite confinement on chemistry, larger models are often created in

which the central cluster containing the chemically relevant active site is treated with QM and a surrounding silicate cage is treated at some lower level of theory (Sherwood et al. 1997; Sierka and Sauer 2005). Such hybrid quantum mechanics-molecular mechanics (QM/MM) methods potentially offer computational efficiency combined with the accuracy of general QM methods. These models take some care to construct and are not well suited to problems in which the site of interest is not well known or well localized. A more natural approach to modeling a zeolite is to apply a supercell approach in which periodic boundary conditions (PBC) are used to naturally recover the dimensionality of a zeolite structure (Rozanska et al. 2003; Hafner 2008). The periodic nature of a supercell model is well suited to plane-wave density functional theory (DFT) treatment and also enables ready implementation of ab initio molecular dynamics (AIMD) simulations to study dynamics behavior of the material (Hass et al. 1998). In this contribution, several common static DFT and state-of-the-art dynamics approaches as well as systematic strategies to model exchanged cation properties in Al-substituted zeolites using supercell model are described. Emphases are placed on methodologies used to probe Brønsted acidity, reactivity of exchange protons, Lewis acidity, siting preference, and redox thermodynamics of exchanged metal cations.

2   Brønsted Acid Site Modeling

A proton is the simplest form of an exchanged cation present in a zeolite, typically residing at one of the tetrahedrally coordinated oxygen atoms around the Al center. Exchanged protons contribute to the Brønsted acidity of a zeolite, which is vital to a broad spectrum of acid-catalyzed reactions. The Brønsted acid strength can influence catalytic activity and is determined by the local environment and accessibility of the protons, as determined by framework topology, type of T-site, different framework oxygen adjacent to each T-site, etc. For example, the MFI zeolite framework contains 12 symmetry-distinct T-sites with 4 first-shell oxygen atom for each, yielding in principle 48 different types of Brønsted acid sites in total, which can all be captured by an orthorhombic supercell as shown in Fig. 1. Experimentally observed Brønsted acid properties are generally an ensemble-average among all possible sites. In order to build the structure-property relationship of Brønsted acids, it is more straightforward and unambiguous to simulate each site explicitly through first-principles modeling than deconvolution from experimental observations.

2.1   Deprotonation Energy

One fundamental electronic property of a Brønsted acid site is the interaction strength between the proton and the anionic framework block, which can be quantified by a deprotonation energy (DPE), defined as the energy to move a proton from a Brønsted acid site into a noninteracting environment. DPE is therefore a

Fig. 1 Orthorhombic MFI zeolite supercell with 12 structural-unique T-sites highlighted and indexed. Four possible Brønsted acid site structures associated with single T-site Al are illustrated. (Adapted from Hernandez-Tamargo et al. 2016)
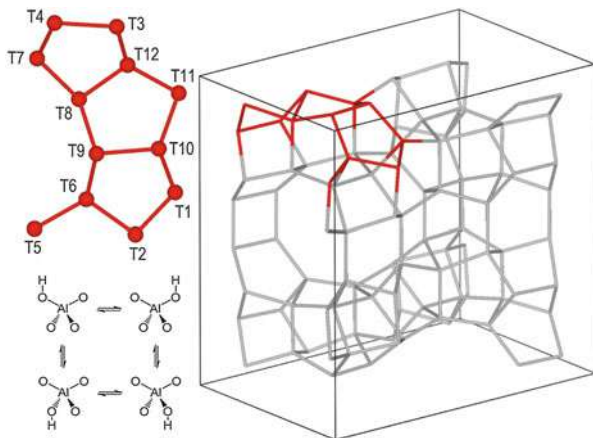
direct descriptor of the acid strength of a Brønsted acid site. Here we adopt the common "Z" shorthand as a symbol for an Al-substituted tetrahedral site in a zeolite. DPE formally corresponds to the energy of the reaction

$$DPE = E_{Z^-} + E_{H^+} - E_{ZH} \qquad (1)$$

where E_{Z^-}, E_{H^+}, and E_{ZH} are DFT-computed 0-K energies of the anionic zeolite framework after deprotonation, an isolated proton, and neutral zeolite framework with a Brønsted acid site, respectively. Comprehensive studies of DPE for a variety of zeolite frameworks have been reported by several authors (Jones and Iglesia 2015; Rybicki and Sauer 2015). It is worthy to note that within the PBC, supercell model, calculations on the charged anionic zeolite framework or gaseous proton suffer from electrostatic interactions between the periodic images that can lead to erroneous DFT energies without appropriate corrections. One should use caution while trying to compare Brønsted acid strength using PBC DFT-computed DPE values across zeolite framework topologies because the degree of periodic image interaction artifact depends on supercell size as well as void fraction of the cell (Jones and Iglesia 2015).

2.2 Probe Molecule Binding

An alternative approach to quantify Brønsted acid strength is to calculate binding strength of a basic probe molecule, commonly NH_3, to the acid site. It is noteworthy that, unlike deprotonation energy, probe molecule binding strength is not determined only by the acid strength but also the local environment of the acid site through van der Waals interaction and/or hydrogen bonding with the zeolite framework (Jones and Iglesia 2015). Consequently, the net binding strength is the summation
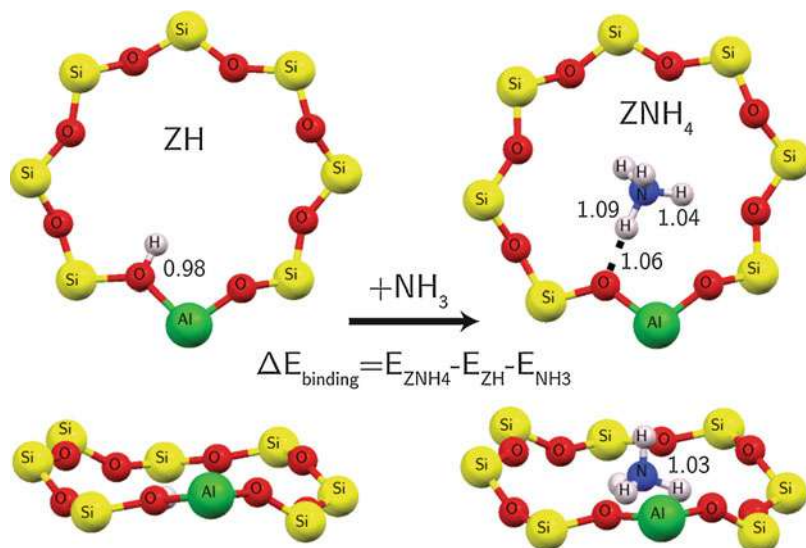
Fig. 2 DFT-optimized structures of ZH and ZNH_4. Only eight-membered rings and adsorbates are illustrated. Distances are given in Å. Equation of NH_3 binding energy calculation is inserted (Li et al. 2017)

of acid-base interaction and other weak interactions that are sensitive to framework pore and ring structure surrounding the Al substituted T-site.

Figure 2 presents an example in which NH_3 is used to probe the Brønsted acid strength of chabazite (CHA) zeolite (Li et al. 2017). The CHA lattice contains only one type of symmetry-distinct T-site, and the entire structure can be representing within a 12-T-site supercell. In a first step, DFT calculations were performed to search for the lowest energy structures of each species, including searches over the four distinct first-shell oxygen atoms adjacent to a substituted Al where a proton can locate. To identify the most favorable proton position, DFT geometry optimization calculations were performed for ZH structures with various proton positions. When NH_3 is placed close to a Brønsted acid, the location of energy minima becomes more challenging, as NH_3 extracts the H^+ from the framework to create an NH_4^+ that exhibits multiple local minima that are close in energy. In such situations, AIMD simulations at elevated temperatures are useful for sampling the potential energy surface. Low energy structures visited in the AIMD trajectory can be further relaxed to their ground states to collect local and, ideally, the global minimum. The binding energy can be calculated as the reaction energy of moving an adsorbed NH_3 to a noninteracting gas-phase environment, formulated by the equation shown in Fig. 2, where E_{ZNH_4}, E_{ZH}, and E_{NH_3} are DFT-computed energies of a zeolitic Brønsted acid with, without NH_3 adsorbate, and isolated gaseous phase NH_3, respectively. Binding energies are typically found to be in the range of $130 \, kJ \, mol^{-1}$.

2.3   AIMD Potentials of Mean Force

Raw DFT-computed probe molecule binding energies include only the 0-K electronic contributions to binding. While typically good reporters of chemical reactivity, comparisons with experimental observations, such as observed NH_3 temperature programmed desorption profiles, are facilitated by the incorporation of finite temperature effects, including the temperature-dependent entropic and enthalpic contributions to the binding free energy. The most common model to incorporate finite temperature effects is the harmonic oscillator (HO) approximation, appropriate for species that exhibit little conformational flexibility. The HO approximation breaks down when adsorbate motions contain significant anharmonicities or are approximately free translators and/or rotators. Improper treatment of these translational and rotational entropies can lead to underestimation of binding free energies and thus of adsorbate coverages and desorption temperatures (Campbell and Sellers 2012; Paolucci et al. 2014). One approach for incorporating these finite temperature, dynamical effects within a supercell model is AIMD coupled with the potential of mean force (PMF) algorithm, a MD-based umbrella sampling method (Wong and York 2012; den Otter 2013). In the PMF simulation, a relevant reaction coordinate is first defined. For adsorbate binding, the reaction coordinate is most naturally the interatomic distance between adsorbate and adsorbent. MD simulations are performed along the one-dimensional grid of the defined reaction coordinate. For any points on the grid that are away from the equilibrium state, constraint forces are present along the reaction coordinate. The Helmholtz free energy to desorb to the center of the cage is determined by integrating the ensemble-averaged constraint forces along the reaction coordinate from the equilibrium to the desorbed state (Li et al. 2018):

$$\Delta A(T) = -\int_{r_{eq}}^{r_{max}} \langle F(r, T) \rangle \, dr \tag{2}$$

Taking NH_3 binding to a Brønsted acid site in CHA zeolite again as an example (Li et al. 2017), a reaction coordinate was defined as the interatomic separation between the proton and the NH_3 nitrogen center (d_{H-N}), as shown in Fig. 3. Grid points are selected to be sufficiently close to ensure a smooth integration of Eq. (2.3) and sufficiently far to keep computational expense manageable. Twenty discrete values of d_{H-N} were chosen in the range from 0.9 to 4.5 Å. At each distance, 40 ps of AIMD were performed at 473 K using the CPMD code, and the constraint forces averaged over the second half of trajectories. As the AIMD simulations are rather computational demanding but the simulation quality greatly depends on the sampling time, to estimate the statistical errors, multiple replicate simulations with different initial atomic coordinates were performed at each d_{H-N}. The free energy surface (FES) at 473 K was constructed through integration of constraint forces along d_{H-N}. As suggested by the location of the minimum in the FES, the equilibrium $d_{H-N} \approx 1.0$ Å, and at distances less than or greater than this separation,
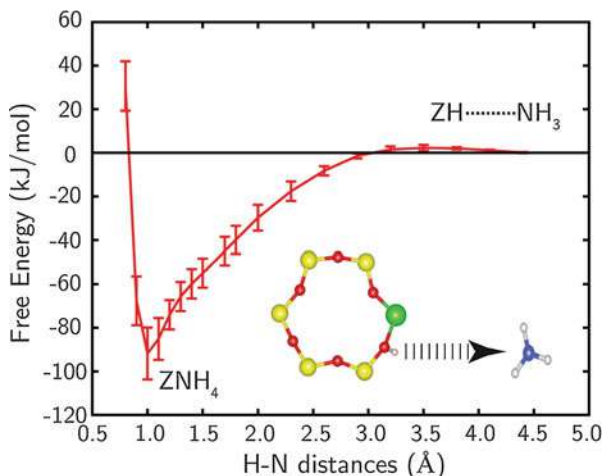
Fig. 3 Free energy profiles for chemisorption of NH_3 to a Brønsted acid site in CHA from PMF simulations (error bars included). Color scheme: O, red; H, pink; Si, yellow; N, blue; Al, green (Li et al. 2017)

the constraint forces are repulsive or attractive, respectively. The FES plateaus as d_{H-N} approaches 4.5 Å, suggesting the NH_3 becomes noninteracting with the Brønsted acid at this separation. To assure that the probe is not perturbed by the framework, the rotational and vibrational dynamics at the end of the constraint path can be compared with those of the free gas. The NH_3 binding free energy at the simulated temperature is the difference between the minimum value and the flat region value in the FES plus the free energy associated with moving the adsorbate from the cage center to outside the zeolite, which is readily estimated (Li et al. 2018).

2.4   Metadynamics Simulations of Brønsted Acid Reactions

The strengths of a Brønsted acid and its local framework environment are both potential factors governing their influence on the rate of an acid-catalyzed reaction. To explore relationship between the nature of a Brønsted acid site and its catalytic function, the intermediates(s) and transition state(s) along the path need to be explicitly identified. The nudged elastic band (NEB) algorithm (Jónsson et al. 1998) and the further refined climbing image NEB (Henkelman et al. 2000) are the most common tools used to locate transition states and minimum energy paths for reactions within the supercell model. As with the geometry optimizations discussed above, these static NEB methods report on the 0-K electronic energy of a particular reaction path and when combined with the harmonic approximation can be used to extract harmonic transition state theory rate constants for use in further microkinetic modeling. However, these static methods are less satisfactory for reactions that

have loosely bound, anharmonic reactant, transition, or product states. A PMF approach is an appropriate alternative when a single reaction coordinate can be clearly identified, but for more complex concerted reactions, the more flexible metadynamics approach is more suitable (Iannuzzi et al. 2003; Laio et al. 2005).

The underlying principle of metadynamics is to supplement the intrinsic DFT potential energy surfaces with a series of Gaussian-shaped biasing potentials accumulated during the course of the AIMD trajectory. These biasing potentials, if introduced as small perturbations through the course of a dynamics trajectory, can smoothly drive the system away from one equilibrium state toward some other region of configuration space of interest. These biasing potentials are applied along predefined collective variables (CV) that can be as simple as a single distance or as complex as a linear combination of many internal degrees of freedom, contributing to the overall flexibility of the method. As with a PMF, metadynamics accelerates the sampling of higher free energy parts of the PES and in particular accelerates the crossing of energy barriers. The collection of biasing potentials accumulated over the course of the simulation reports on the underlying free energy surface, including both the enthalpic and entropic parts of the reaction. The FES of the reaction can be recovered by summing up the biasing potentials added in the defined grid of CV space. To ensure that a particular metadynamics simulation proceeds along a productive direction and to enhance simulation efficiency, CVs should be chosen that are directly relevant to the bond breaking and formation events associated with the reactants and products. Several applications of metadynamics for constructing FES within a supercell model have been reported (Brüssel et al. 2011; Moors et al. 2013; Erichsen et al. 2015).

Here we use the reaction between a Brønsted acid site and two nitrogen oxides (NO_x), a reaction relevant to the Brønsted-acid-catalyzed fast selective catalytic reduction (SCR) of NO_x with NH_3, to illustrate a practical application of metadynamics for a catalytic process (Li et al. 2017). The reactants and products are shown in Fig. 4b. The desired reaction coordinate involves proton (H^+) transfer from a framework oxygen (O_f) to one oxygen of an approaching NO_2 molecule (O_{NO_2}) and concerted adsorption of a NO molecule to an O_f through its N center (N_{NO}). One complication in modeling such a process is that NO at the product stage can potentially coordinate with any one of O_f adjacent to Al and, to obtain a smooth FES and avoid bias, must be included on an equal footing into the CVs. For this purpose one can define a generalized coordination number:

$$\sum_{j \neq i}^{n_{\text{list}}} \frac{1 - \left(\frac{d_{ij}}{d_0}\right)^p}{1 - \left(\frac{d_{ij}}{d_0}\right)^q}, \tag{3}$$

Here d_{ij} is the distance between atoms i and j, and d_0 is a reference distance corresponding to the approximate bond distance between atoms, and p and q are parameters that determine the balance between sampling bound and unbound regions of the FES. In the simulation illustrated in Fig. 4a, CV1 was set to the
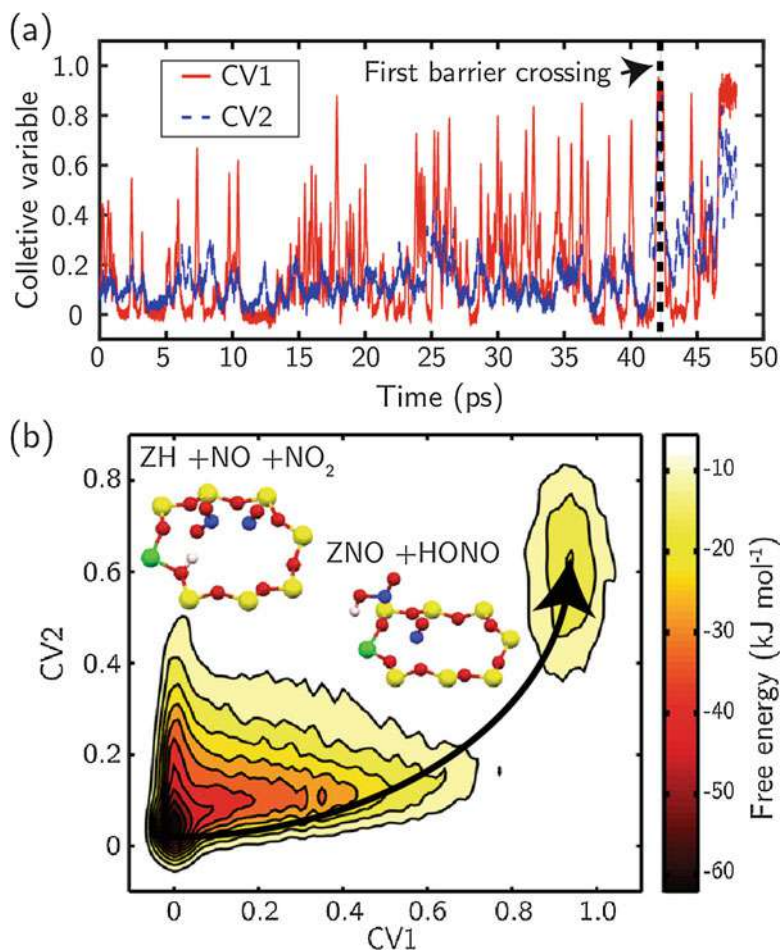
Fig. 4 (**a**) Time evolution of CVs in single run of metadynamics simulation for reaction between NO_x and a Brønsted acid site in CHA at 473 K. (**b**) Free energy surface projected on the two auxiliary collective variables (CV1 and CV2) from metadynamics simulation. Structures of reactants and products are superimposed. Color scheme: O, red; H, pink; Si, yellow; N, blue; Al, green (Li et al. 2017)

$H-O_{NO2}$ CN and CV2 to the O_f-N_{NO} CN, averaged over all four O_f. CV trajectories during the course of the metadynamics simulation are plotted in Fig. 4a. Near the beginning of the simulation, both CVs fluctuate around 0.1, corresponding to a region of configuration space in which $H-O_{NO_2}$ and O_f-N_{NO} are unbound and a system in the ZH, NO, NO_2 reactant well. As time progresses, the metadynamics biasing algorithm pushes the system to explore regions of higher CV space. Near 42 ps, both CV1 and CV2 attain values associated with the product state, suggestive of a first crossing of the reaction barrier and formation of ZNO and nitrous acid

HONO. The simulation can be terminated at this point if the purpose is solely to quantify the activation free energy, which corresponds to the depth of the reactant basin in the constructed FES. If the goal is to simulate the full FES including reactant and product basins, simulations are continued until the CV values return to the reactant region so that product free energy basin is well sampled and filled, as shown in Fig. 4b. Metadynamics can in principle be applied to any reactions of interest and is particularly powerful for those reactions with only limited knowledge on the reaction pathways, because CVs are generally loosely defined and the system still retains a high degree of freedom that allows the reaction to proceed through the most favorable pathway on the fly. On the other hand, really free energies depend on careful selection of the parameters that control the rate of addition and size of the biasing potentials.

2.5   Interaction Between Brønsted Sites

We have discussed to this point the properties of an isolated Brønsted site within a zeolite, taking CHA as an example. In real zeolite materials, Al T-sites are generally not ordered, and the silicon to aluminum ratio (SAR) is typically such that there is always some probability for two Al-containing T-sites to be in sufficient proximity to influence chemistry (Paolucci et al. 2017). The distribution of Al within a zeolite framework can in principle be dependent on the types of symmetry-distinct T-sites as well as any interactions between the sites that prevail at the time of synthesis. "Löwenstein's rule," for instance, is the empirical observation that two Al are generally not found in adjacent T-sites. This observation suggests that there may be some energetic interaction between T-sites. This idea has been explored in supercell simulations of paired Brønsted sites in CHA zeolite (Fletcher et al. 2017).

The size of the zeolite supercell determines the maximum unique Al−Al pair combinations that can be represented. Figure 5 shows a representative 36-T-site supercell for CHA. This supercell nominally contains 35 distinct Al−Al pair combinations and, considering that each T-site has 4 distinct associated framework O, 560 Brønsted acid pair configurations. Because of PBC considerations, the actual number of unique combinations is reduced. Figure 5 shows DFT-computed energies for all possible combinations within the 36-T-site supercell. Results are plotted against Al−Al separation and normalized to the energy at infinite Al separation:

$$2\,HSi_{35}AlO_{72} \longrightarrow H_2Si_{34}Al_2O_{72} + Si_{36}O_{72} \tag{4}$$
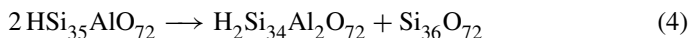
The broad distribution of energies at any separation reflects the strong dependence of relative energy on proton location. At any given separation, energies span the infinite separation baseline. Further, this range broadens with decreasing Al−Al separation. Energy is minimized at the first-nearest-neighbor separation, suggesting that an Al−O−Al configuration is particularly low in energy when protons are the charge-compensating species and thus that this energy landscape cannot account for Löwenstein's rule. Further, the results suggest that Brønsted acid strength

Fig. 5 Relative energy of Brønsted acid site pair against framework aluminum separation for H-SSZ-13 (Fletcher et al. 2017)
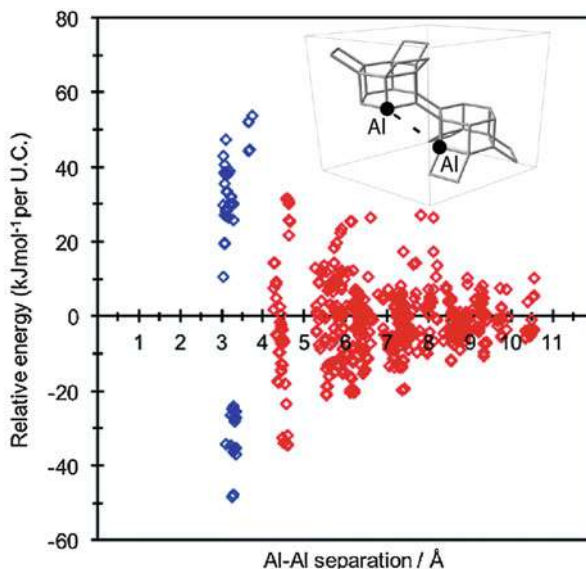
is sensitive to the proximity of an acid site to a neighbor. In general, these findings highlight the importance of considering the influence of heterogenous site distributions in zeolite material simulation.

3 Cationic Metal Site Modeling

3.1 Metal Cations Anchored to Isolated Al T-Sites

Metals that can exist as monovalent cations, e.g., alkali metals and some transition metals, are similar to protons in their ability to charge-compensate isolated Al T-sites in zeolites. The same methodologies discussed above can be applied to study the electronic and chemical properties of these cationic metal sites, which are typically Lewis acidic (Thang et al. 2014). For metals that exhibit more than one oxidation state, redox characteristics become relevant. For instance, Cu exists in both 1+ and 2+ oxidation states and thus can in principle charge-compensate one or two Al within a zeolite. Even in the 2+ state, the presence of additional charge-compensating ligands (OH^-) will alter the net charge of the Cu complex and thus its interaction with a zeolite host. Redox cycling between oxidation states has been demonstrated to be key to the performance of Cu-exchanged SSZ-13 zeolites in the standard SCR of NO_x by ammonia (Paolucci et al. 2014, 2017; Gunter et al. 2015). Thus, first-principles models of metal exchange in zeolites become significantly more rich and challenging than that of monovalent Brønsted sites.

3.2   Polyvalent Metal Cations Anchored to Paired Al T-Sites

Because of its relevance to SCR, Cu exchange in CHA zeolites has received
considerable computational attention. Figure 6 shows a representation of the CHA
cage structure, illustrating the various rings and voids within the zeolite that can host
a metal ion (Paolucci et al. 2016b). All T-sites within this structure are equivalent.
DFT calculations report that an isolated Cu^I ion charge-compensating a single Al
T-site seeks a twofold coordinate site, near one O_f adjacent to Al and one one-
site removed. Two Al are required to charge-compensate a ligand-free Cu^{II}, and
the structure and stability thus becomes a function of the proximity of the two Al.
Figure 7 reports DFT-supercell-computed structures of a nominally Cu^{II} ion in a
variety of Al pair sites present in CHA, including a second-nearest-neighbor (2NN)
Al pair in a four-membered ring (4MR), third-nearest-neighbor (3NN) pairs in six
(6MR)- and eight-membered rings (8MR), and fourth-nearest-neighbor (4NN) pairs
in an 8MR (Verma et al. 2014). All structures were optimized within a 12-T-site
rhombohedral CHA supercell and some repeated in a doubled, 24-T-site supercell to
test the sensitivity of computed energies on supercell size. Bader charge analysis on
Cu confirms that all Cu are in the same (formally 2+) oxidation state. The structural
preference for Cu^{II} for higher coordination is evident in the computed structures,
particularly in structure (b).

   Computed DFT energies are summarized in Table 1. Relative energies span
approximately 1.5 eV and are generally insensitive to supercell size, and among
the six configurations explored, configuration (b), corresponding to a Cu^{II} in
the 6MR, is substantially lower in energy than other combinations. This energy
preference likely reflects the ability of the 6MR to comfortably accommodate the
fourfold coordination preference and preferred $Cu^{II}-O$ distances with minimal
lattice strain. These energies convolute the energy of interaction of the Cu^{II} ion
with the zeolite framework and the energy difference between different $Al-Al$
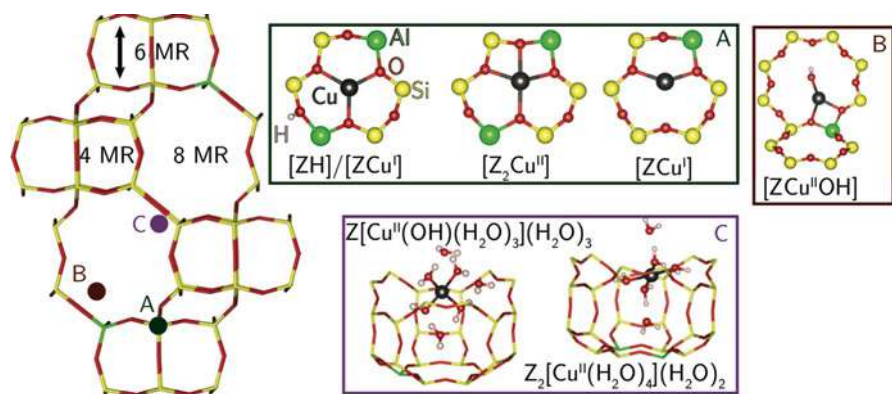
Fig. 6 Side view of the CHA cage structure and computed structures of Cu^I and Cu^{II} ions
attached to the zeolite framework and solvated by H_2O (Paolucci et al. 2016b)
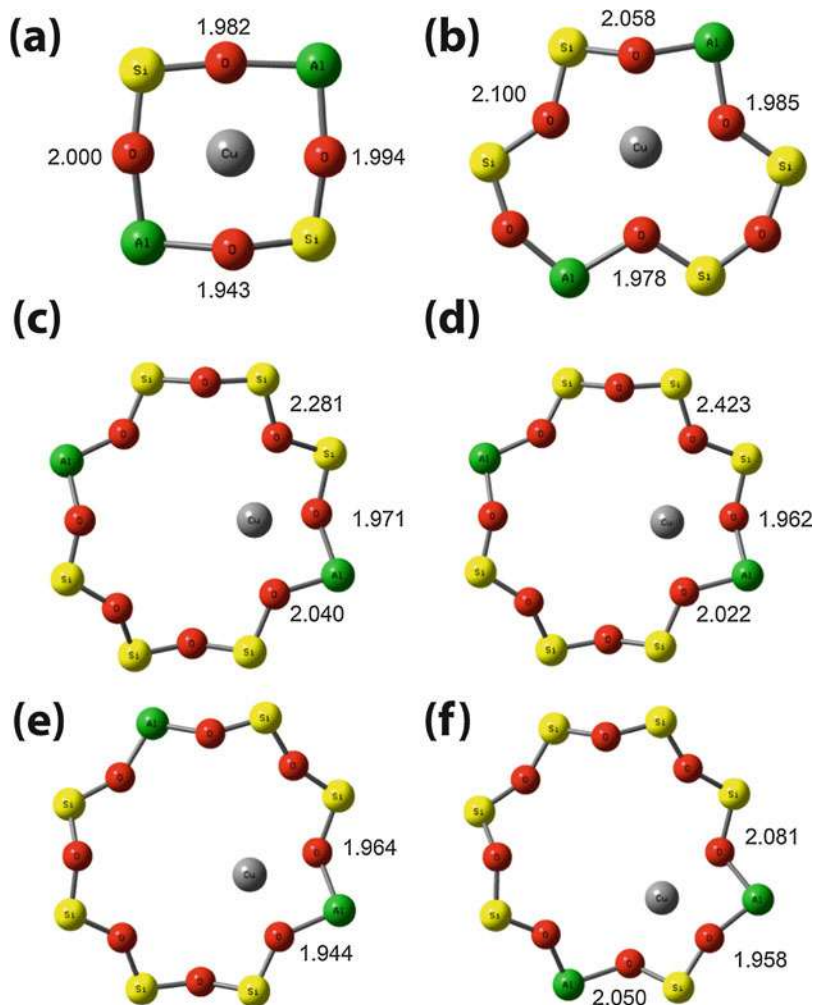
Fig. 7 Relaxed structures of an isolated exchanged Cu^{II} ion in CHA (Verma et al. 2014). Distances between O_f and Cu indicated in Å. (**a**) $2 \times 1 \times 1$ supercell 2NN Al pair in 4MR. (**b**) $2 \times 1 \times 1$ supercell 3NN Al pair in 6MR. (**c**) $2 \times 1 \times 1$ supercell 4NN Al pair in 8MR. (**d**) $1 \times 1 \times 1$ supercell 4NN Al pair in 8MR. (**e**) $1 \times 1 \times 1$ supercell 3NN Al pair in 8MR. (**f**) $1 \times 1 \times 1$ supercell 2NN Al pair in 8MR. Color scheme: O, red; Si, yellow; Cu, gray; Al, green

pair combinations. These two factors are not directly separable, and one can only make comparisons on the basis of some balanced reaction involving other charge-compensating species. One can define, for instance, an exchange energy between different sites as

$$Z_2H_{2\,config1} + Z_2Cu_{config2} \longrightarrow Z_2H_{2\,config2} + Z_2Cu_{config1} \qquad (5)$$
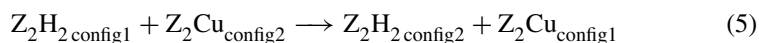
Table 1 Raw and relative energies of isolated Cu^{II} ions at different Al pair locations in CHA. All energies in eV (Verma et al. 2014)

Al-pair sites	From $2 \times 1 \times 1$ supercell		From $1 \times 1 \times 1$ supercell	
	Raw energy	Relative energy	Raw energy	Relative energy
2NN in 4MR	-574.622	-0.37	-284.155	-0.37
3NN in 6MR	-575.751	-1.51	-285.203	-1.42
4NN in 8MR	-574.246	0	-283.783	0
3NN in 8MR	N/A	N/A	-283.550	0.23
2NN in 4MR	N/A	N/A	-284.030	-0.25

which convolves the energies shown in Table 1 with those reported in Fig. 5. Exact results depend on site, but it is evident that the 1.5 eV advantage of the 6MR site over others for Cu^{II} overwhelms the $\ll 0.4$ eV difference between various Z_2H_2 configurations. Thus, on the basis of this exchange reaction, Cu^{II} will preferentially populate 2Al 6MR sites over other sites, when exchange occurs against the Brønsted form of the zeolite.

3.3   Cu Exchange Siting Between 1Al and 2Al Sites in CHA

As a redox-active ion, Cu^{II} can be present in the zeolite in more than one chemical form. Figure 6 shows the computed structure of ZCuOH, a formally Cu^{II} ion charge-compensated by both a single Al T-site plus an extra-lattice OH^-. DFT calculations show that the addition of this OH^- ligand to an isolated Cu^I moves the Cu center from the center of a 6MR up into an adjacent 8MR (Fig. 6), illustrating how Cu location is both a function of the Al site and additional Cu coordination.

In a typical Cu ion exchange experiment, Cu^{II} is introduced to the zeolite as an aqueous solution, Cu is absorbed, and the material is then heated to drive off H_2O. In principle every Al T-site could host a Cu^{II} ion as ZCuOH, leading to a Cu/Al ratio of 1. Ionic Cu uptake in Cu-CHA is typically observed to saturate at Cu/Al ratios lower than unity, however, likely reflecting the competition between 1Al (ZCuOH) and 2Al (Z_2Cu) sites for Cu^{II}. The relative energies of these two types of sites can only be compared by constructing a stoichiometric reaction that connects the two. A variety of choices are possible, but one that is consistent with the conditions of Cu exchange is

$$Z_2Cu^{II}H_2O + ZH \longrightarrow Z_2H_2 + ZCu^{II}OH \qquad (6)$$
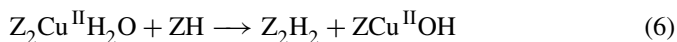
where H_2O Lewis-bound to Z_2Cu is used to balance the overall reaction and the same Al pair site is implied on both sides of the reaction. The Z_2CuH_2O structure and energy can be evaluated using a computational approach analogous to that described above for NH_3 on a Brønsted site, including AIMD annealing followed

by a relaxation. The H_2O binding energy to a 6MR Z_2Cu evaluated this way is computed to be approximately 0.8 eV (Paolucci et al. 2016b). The overall energy of Eq. (6) can then be computed either in separate single Al and paired Al supercells or in one, larger, supercell that contains both types of sites of interest. Tests indicate that for Cu in CHA, both approaches give similar results and for two Al 3NN in a 6MR give an energy of about 0.6 eV, reflecting an energetic preference for populating these Z_2Cu sites over ZCuOH sites.

Predicting the equilibrium fraction of Z_2Cu and ZCuOH sites in a particular zeolite requires knowledge both of these exchange preferences and the distribution of Al with the framework. This Al distribution is in general a function of the Si to Al ratio (Si/Al) and the zeolite synthesis conditions and cannot be predicted, for instance, from the relative energies of Brønsted sites (Fig. 5). Because CHA contains only a single T-site type, a simple model for the Al distribution is that Al are distributed randomly on the CHA lattice subject to the empirically observed constraint that Al avoid 1NN sites (Löwenstein's rule). This model can be exercised in numerical simulations to predict the distribution of Al as a function of Si/Al. Combining this distribution with a rule derived from Eq. (6) that exchanged Cu^{II} populates 6MR 2Al sites preferentially before populating residual isolated Al sites as ZCuOH leads to the composition phase diagram shown in Fig. 8 (Paolucci et al. 2016b). The white line represents the saturation limit of Z_2Cu sites, and the colored range the relative $ZuOH/Z_2Cu$ fraction. A variety of chemical probes can be used to compare these predictions with observation, including residual Brønsted site density, spectroscopic observation of the ZCuO−H intensity, and chemical titrations using probes sensitive only to paired Al. Paolucci et al. found that these predictions matched experimental observations for Cu-CHA samples of a wide range of Si/Al and Cu/Al ratios prepared under the same conditions.
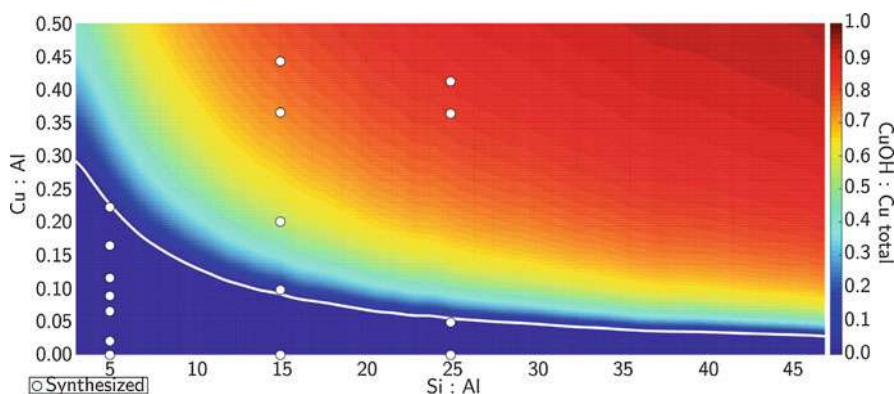
Fig. 8 Predicted Cu site compositional phase diagram versus Si/Al and Cu/Al ratios. Color scale indicates predicted fraction of $ZCu^{II}OH$. White line demarcates transition from Z_2Cu^{II}-only region to mixed $Z_2Cu^{II}/Cu^{II}OH$ region. White circles indicated compositions of synthesized Cu-SSZ-13 samples (Paolucci et al. 2016b)

These results illustrate the potential richness of metal ion exchange in zeolites introduced by polyvalency, extra-lattice ligands, and Al distributions. Cu exchange in CHA is a relatively simple example in which these factors can all be accounted for through DFT evaluation of energies combined with stochastics. These problems become more challenging for lower symmetry zeolites and for metal ions that exhibit more complex exchange chemistry.

3.4   Thermodynamics of Cu Speciation in CHA

Heterogeneous catalytic reactions at metal surfaces are understood in terms of the adsorption, dissociation, rearrangement, and desorption of species at a relatively constant metal (or metal oxide) surface. Metal-exchanged zeolites, while technically heterogeneous catalysts at the macro level, exhibit many of the characteristics of homogeneous catalysts at the atomic scale. As shown in panel C of Fig. 6, isolated Z_2Cu or ZCuOH sites can adsorb multiple ligands (H_2O, in the example shown) and in the process become detached from the zeolite framework itself and move into the void space of the zeolite cage (Kwak et al. 2014; Psofogiannakis et al. 2015; Paolucci et al. 2016b). In the process, Cu centers evolve from a rather static into a dynamic state. Figure 9 compares the volume of space visited during 90 ps of AIMD at 298 K of "dry" and fully hydrated Cu^{II} ions in the CHA cage (Paolucci et al. 2016b); hydration substantially increases this space even at this ambient temperature.
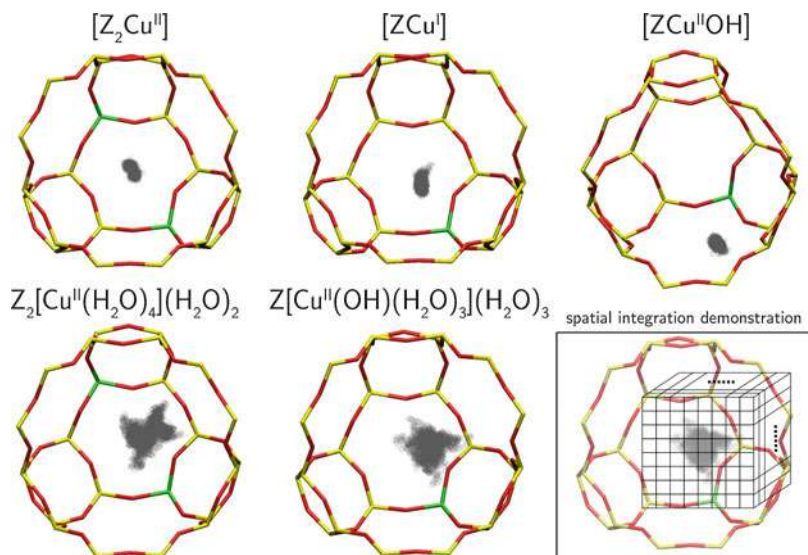
Fig. 9 Volume of space visited by the Cu center during 90 ps of 298 K AIMD as a function of the Cu oxidation and coordination states (Paolucci et al. 2016b)

To compare with and rationalize experimental observations, it is important to be able to predict the equilibrium state of the Cu sites as a function of external conditions, including temperature and pressures of various gases. A first-principles thermodynamic approach is well suited to this problem. In this approach, the various states of a catalyst site are written in terms of formation reactions with respect to some base state of the site and some combination of relevant reactive species. Under hydrothermal conditions, for instance, this formation reaction can be written as

$$Z_*\text{Cu} + \frac{x}{2}\left(H_2O - \frac{1}{2}O_2\right) + \frac{y}{2}O_2 \xrightarrow{\Delta G_f} Z_*\text{CuH}_xO_y \tag{7}$$
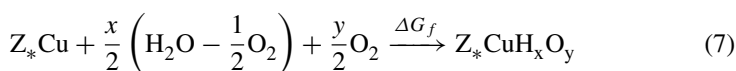
where $Z_*\text{Cu}$ represents either a single or paired Al site and H_2O and O_2 are used as reference species for the various ligands that may cover the Cu ion or associate with the lattice itself. Application of this strategy presents two significant challenges: first, to generate a comprehensive list of product species and second, to relate DFT-computed energies to finite temperature free energies. Because the number of candidate species involved is potentially quite large, the AIMD-based free energy approaches of Sects. 2.3 and 2.4 become prohibitively expensive. In such cases, chemical intuition combined with judicious AIMD-based annealings are currently the best alternative. Paolucci et al. used this approach to identify the structures and compute the energies of more than a dozen different candidate combinations of H_xO_y ligands on both 1Al and 6MR 2Al sites in Cu-CHA (Paolucci et al. 2016b). Further, they used correlations derived from PMF simulations on a limited number of ligands (Paolucci et al. 2014; Li et al. 2018) to estimate the entropic contributions to the free energy. These PMF simulations suggest that adsorbates retain approximately $1/2$ of their gas-phase entropy upon adsorption at the highly dynamic Cu sites.

Figure 10 shows the composition phase diagram as a function of O_2 pressure and temperature at a fixed 0.02 bar H_2O for Cu exchanged near single and paired Al sites. These conditions determine the chemical potentials needed to evaluate the free energy of reaction (7):

$$\Delta G_f^{Z_*\text{CuH}_xO_y}(T, P_{O_2}, P_{H_2O}) = \left(E_{\text{DFT}}^{Z_*\text{CuH}_xO_y} - E_{\text{DFT}}^{Z_*\text{Cu}}\right) - T\left(S^{Z_*\text{CuH}_xO_y} - S^{Z_*\text{Cu}}\right)$$

$$- \frac{x}{2}\mu^{H_2O}(T, P_{H_2O}) - \frac{y - x/2}{2}\mu^{O_2}(T, P_{O_2}) \tag{8}$$

and the figure reports the composition that minimizes free energy at the specified conditions. These diagrams are a powerful tool for predicting the evolution of sites as a function of conditions. For instance, in Fig. 10, both Cu ions are seen to evolve from a fully hydrated, Cu^{II} state at ambient conditions (state 1) to a fully dehydrated, zeolite-bound state at elevated temperature (state 2). Thermal "autoreduction" to a Cu^{I} state is predicted to occur for the ZCuOH sites but not the Z_2Cu sites. All these observations are in good agreement with experiment, in particular X-ray spectroscopic characterization of oxidation states and coordination numbers
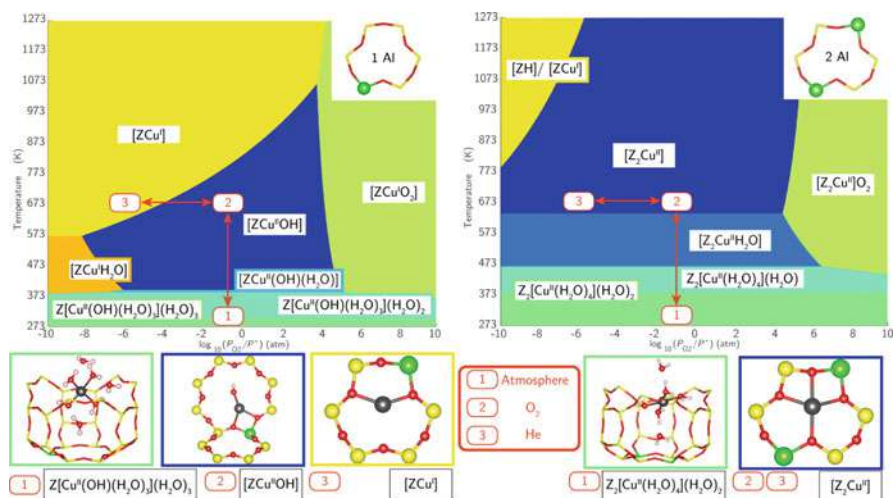
Fig. 10 Ex situ Cu speciation phase diagrams based on HSE06-Tsvdw calculations on 1Al (left) and 2Al (right) Cu exchange sites. Regions indicate site composition that minimizes free energy at 2% H_2O and given T and P_{O_2}. Labeled on the phase diagram and illustrated below are minimum free energy species at (1) ambient (298 K, 20% O_2), (2) oxidizing (673 K, 20% O_2), and (3) inert (673 K, 10^{-6} atm O_2 in He) (Paolucci et al. 2016b)

(Paolucci et al. 2016b). A similar strategy applied to these sites under conditions representative of standard NO_x SCR reveals the role of NH_3 both as a Cu ion ligand and a reductant in this redox chemistry.

4   Outlook and Challenges

The structural and chemical diversity of zeolites presents a rich set of challenges for first-principles computations. We have focused here on the chemistry of Brønsted and Lewis (Cu ion) sites in CHA as two examples for which computational models are well developed and have led to significant insights. The dynamic nature of both types of sites motivates careful consideration of the state of these sites at finite temperature and in the presence of reactive gases. The three-dimensional structure of CHA can be captured in supercells small enough to be amenable to AIMD, but zeolites of lower symmetry and larger supercell volumes remain a challenge for fully first-principles models, as do models able to reliably predict free energy changes upon adsorption and reaction at highly dynamic sites. Zeolite chemistry is closely associated with framework-substituted heteroatoms. While we have focused here on Al, other types of iso- and aliovalent dopants are possible. Further, as we have seen, the distribution of these heteroatoms can itself have an influence on zeolite chemistry by creating local sites and ensembles of differing reactivity, and similar ensemble effects likely are related to zeolite hydrothermal stability as

well. Computational models must account for potential ensemble and environmental effects in any robust comparisons with experiment. For instance, recent experiments demonstrate that ZCuOH but not Z_2Cu sites participate in the selective oxidation of methane to methanol over Cu-CHA (Pappas et al. 2017), a conclusion quite different from that for NO_x SCR (Paolucci et al. 2016b). In fact, recent results highlight the coupled role of Al spatial distributions and adsorbate-induced active site diffusion and pairing in controlling the rate of the O_2-consuming step in standard NO_x SCR (Paolucci et al. 2017). Thus, computational descriptions of zeolite structure and dynamics present a coupled challenge and opportunity for both understanding and predicting heterogeneous catalytic performance.

References

Auerbach S, Carrado K, Dutta PK (2003) Handbook of zeolite science and technology. Dekker, New York

Brüssel M, Di Dio PJ, Munñiz K, Kirchner B (2011) Comparison of free energy surfaces calculations from ab initio molecular dynamic simulations at the example of two transition metal catalyzed reactions. Int J Mol Sci 12(2):1389–1409

Campbell CT, Sellers JRV (2012) The entropies of adsorbed molecules. J Am Chem Soc 134(43):18109–18115

Colella C (1996) Ion exchange equilibria in zeolite minerals. Mineral Deposita 31(6):554–562

Corma A (1989) Application of zeolites in fluid catalytic cracking and related processes. In: Jacobs PA, van Santen RA (eds) Zeolites: facts, figures, future part A – Proceedings of the 8th international zeolite conference. Studies in surface science and catalysis, vol 49. Elsevier, Amsterdam pp 49–67. https://doi.org/10.1016/S0167-2991(08)61708-5, http://www.sciencedirect.com/science/article/pii/S0167299108617085

den Otter WK (2013) Revisiting the exact relation between potential of mean force and free-energy profile. J Chem Theory Comput 9(9):3861–3865

Erichsen MW, Wispelaere KD, Hemelsoet K, Moors SL, Deconinck T, Waroquier M, Svelle S, Speybroeck VV, Olsbye U (2015) How zeolitic acid strength and composition alter the reactivity of alkenes and aromatics towards methanol. J Catal 328(Supplement C):186–196

Fletcher RE, Ling S, Slater B (2017) Violations of Lowenstein's rule in zeolites. Chem Sci 8:7483–7491

Gonzales NO, Chakraborty AK, Bell AT (1998) A density functional theory study of the effects of metal cations on the Brønsted acidity of H-ZSM-5. Catal Lett 50(3):135–139

Gunter T, Carvalho HWP, Doronkin DE, Sheppard T, Glatzel P, Atkins AJ, Rudolph J, Jacob CR, Casapu M, Grunwaldt JD (2015) Structural snapshots of the SCR reaction mechanism on Cu-SSZ-13. Chem Commun 51:9227–9230

Hafner J (2008) Ab-initio simulations of materials using VASP: density-functional theory and beyond. J Comput Chem 29(13):2044–2078

Hass KC, Schneider WF, Curioni A, Andreoni W (1998) The chemistry of water on alumina surfaces: reaction dynamics from first principles. Science 282(5387):265–268

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113(22):9901–9904

Hernandez-Tamargo CE, Roldan A, de Leeuw NH (2016) A density functional theory study of the structure of pure-silica and aluminium-substituted MFI nanosheets. J Solid State Chem 237:192–203

Iannuzzi M, Laio A, Parrinello M (2003) Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. Phys Rev Lett 90:238302

Jones AJ, Iglesia E (2015) The strength of Brønsted acid sites in microporous aluminosilicates. ACS Catal 5(10):5741–5755

Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Berne BJ, Ciccotti G, Coker DF (eds) Classical and Quantum Dynamics in Condensed Phase Simulations. World Scientific, Singapore, pp 385–404

Kramer GJ, De Man AJM, Van Santen RA (1991) Zeolites versus aluminosilicate clusters: the validity of a local description. J Am Chem Soc 113(17):6435–6441

Kwak JH, Varga T, Peden CH, Gao F, Hanson JC, Szanyi J (2014) Following the movement of Cu ions in a SSZ-13 zeolite during dehydration, reduction and adsorption: a combined in situ TP-XRD, XANES/DRIFTS study. J Catal 314:83–93

Laio A, Rodriguez-Fortea A, Gervasio FL, Ceccarelli M, Parrinello M (2005) Assessing the accuracy of metadynamics. J Phys Chem B 109(14):6714–6721

Li H, Paolucci C, Schneider WF (2018) Zeolite adsorption free energies from ab initio potentials of mean force. J Chem Theory Comput 14(2):929–938. https://doi.org/10.1021/acs.jctc.7b00716

Li S, Zheng Y, Gao F, Szanyi J, Schneider WF (2017) Experimental and computational interrogation of fast SCR mechanism and active sites on H-form SSZ-13. ACS Catal 7(8):5087–5096

Lopez A, Tuilier M, Guth J, Delmotte L, Popa J (1993) Titanium in MFI-Type zeolites: a characterization by XANES, EXAFS, IR, and ^{74,49}Ti and ^{17}O MAS NMR spectroscopy and H_2O adsorption. J Solid State Chem 102(2):480–491

Matsubayashi N, Shimada H, Imamura M, Sato T, Okabe K, Yoshimura Y, Nishijima A (1996) Determination of Fe-substituted sites in the MFI structure by Fe K-edge EXAFS. Catal Today 29(1):273–277

Melson GA, Crawford JE, Crites JW, Mbadcam KJ, Stencel JM, Rao VUS (1983) Evaluation of some new zeolite-supported metal catalysts for synthesis gas conversion, Chap 24. In: Stucky GD, Dwyer FG (eds) Intrazeolite chemistry. American Chemical Society, Washington, pp 397–408

Moors SLC, De Wispelaere K, Van der Mynsbrugge J, Waroquier M, Van Speybroeck V (2013) Molecular dynamics kinetic study on the zeolite-catalyzed benzene methylation in ZSM-5. ACS Catal 3(11):2556–2567

Nicholas CP (2017) Applications of light olefin oligomerization to the production of fuels and chemicals. Appl Catal A Gen 543(March):82–97. https://doi.org/10.1016/j.apcata.2017.06.011

Paolucci C, Verma AA, Bates SA, Kispersky VF, Miller JT, Gounder R, Delgass WN, Ribeiro FH, Schneider WF (2014) Isolation of the copper redox steps in the standard selective catalytic reduction on cu-ssz-13. Angew Chem Int Ed 53(44):11828–11833

Paolucci C, Di Iorio J, Ribeiro F, Gounder R, Schneider WF (2016a) Catalysis science of NO_x selective catalytic reduction with ammonia over Cu-SSZ-13 and Cu-SAPO-34. In: Advances in Catalysis, vol 59. Academic Press, pp 1–107. https://doi.org/10.1016/bs.acat.2016.10.002

Paolucci C, Parekh AA, Khurana I, Di Iorio JR, Li H, Albarracin Caballero JD, Shih AJ, Anggara T, Delgass WN, Miller JT, Ribeiro FH, Gounder R, Schneider WF (2016b) Catalysis in a cage: condition-dependent speciation and dynamics of exchanged cu cations in SSZ-13 zeolites. J Am Chem Soc 138(18):6028–6048

Paolucci C, Khurana I, Parekh AA, Li S, Shih AJ, Li H, Di Iorio JR, Albarracin-Caballero JD, Yezerets A, Miller JT, Delgass WN, Ribeiro FH, Schneider WF, Gounder R (2017) Dynamic multinuclear sites formed by mobilized copper ions in NO_x selective catalytic reduction. Science 357:898–903

Pappas DK, Borfecchia E, Dyballa M, Pankin IA, Lomachenko KA, Martini A, Signorile M, Teketel S, Arstad B, Berlier G, Lamberti C, Bordiga S, Olsbye U, Lillerud KP, Svelle S, Beato P (2017) Methane to methanol: structure-activity relationships for Cu-CHA. J Am Chem Soc 139(42):14961–14975. https://doi.org/10.1021/jacs.7b06472

Psofogiannakis GM, McCleerey JF, Jaramillo E, van Duin ACT (2015) ReaxFF reactive molecular dynamics simulation of the hydration of Cu-SSZ-13 zeolite and the formation of Cu dimers. J Phys Chem C 119(12):6678–6686

Rozanska X, van Santen RA, Demuth T, Hutschka F, Hafner J (2003) A periodic DFT study of isobutene chemisorption in proton-exchanged zeolites: dependence of reactivity on the zeolite framework structure. J Phys Chem B 107(6):1309–1315

Rybicki M, Sauer J (2015) Acidity of two-dimensional zeolites. Phys Chem Chem Phys 17:27873–27882

Sauer J, Hobza P, Zahradnik R (1980) Quantum chemical investigation of interaction sites in zeolites and silica. J Phys Chem 84(24):3318–3326

Schneider W, Hass K, Ramprasad R, Adams J (1998) Density functional theory study of transformations of nitrogen oxides catalyzed by cu-exchanged zeolites. J Phys Chem B 102(19):3692–3705

Sherry HS (1966) The ion-exchange properties of zeolites. I. Univalent ion exchange in synthetic faujasite. J Phys Chem 70(4):1158–1168

Sherwood P, de Vries AH, Collins SJ, Greatbanks SP, Burton NA, Vincent MA, Hillier IH (1997) Computer simulation of zeolite structure and reactivity using embedded cluster methods. Faraday Discuss 106:79–92

Sierka M, Sauer J (2001) Proton mobility in chabazite, faujasite, and ZSM-5 zeolite catalysts. Comparison based on ab initio calculations. J Phys Chem B 105(8):1603–1613

Sierka M, Sauer J (2005) Hybrid quantum mechanics/molecular mechanics methods and their application. Springer, Netherlands, pp 241–258

Thang HV, Rubeš M, Bludský O, Nachtigall P (2014) Computational investigation of the Lewis acidity in three-dimensional and corresponding two-dimensional zeolites: UTL vs IPC-1P. J Phys Chem A 118(35):7526–7534

Trout BL, Chakraborty AK, Bell AT (1996) Local spin density functional theory study of copper ion-exchanged zsm-5. J Phys Chem 100(10):4173–4179

Tuma C, Sauer J (2006) Treating dispersion effects in extended systems by hybrid MP2:DFT calculations-protonation of isobutene in zeolite ferrierite. Phys Chem Chem Phys 8:3955–3965

Verma AA, Bates SA, Anggara T, Paolucci C, Parekh AA, Kamasamudram K, Yezerets A, Miller JT, Delgass WN, Schneider WF, Ribeiro FH (2014) No oxidation: a probe reaction on Cu-SSZ-13. J Catal 312(Supplement C):179–190

Wong KY, York DM (2012) Exact relation between potential of mean force and free-energy profile. J Chem Theory Comput 8(11):3998–4003

Microkinetic Modeling of Surface Catalysis

57

Gerhard R. Wittreich, Konstantinos Alexopoulos,
and Dionisios G. Vlachos

Contents

G. R. Wittreich · K. Alexopoulos · D. G. Vlachos (✉)
Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE,
USA
e-mail: wittregr@udel.edu; konalex@udel.edu; vlachos@udel.edu

Abstract

Microkinetic modeling (MKM) breaks down a reaction mechanism into all known elementary steps making no a priori assumptions about dominant reaction paths, rate determining steps, and most abundant reactive intermediates. Instead this information emerges from the solution of the model. Aside from mechanistic understanding, MKM can be utilized to optimize reaction conditions and/or reactor configuration and provide guidelines for catalyst design. This chapter focuses on describing the basics of mean-field MKM. It also details how first-principles calculations or fast-screening methods can be used in conjunction with transition state theory and statistical mechanics to derive kinetic and thermodynamic parameters that abide to thermodynamic consistency constraints. Finally, the chapter covers analysis techniques that provide key insights into the reaction mechanism. We focus primarily on the ammonia decomposition chemistry as an illustrative example.

1 Introduction

In 1864, Waage and Guldberg first proposed mass action kinetics where reaction rates are proportional to the concentration of the reactants (Lund 1965). In 1884, Van't Hoff published his now famous work "Etudes de dynamique chimique" (van't Hoff 1884) where he described how to determine reaction orders and to link equilibrium to thermodynamics. Deriving simplified rate equations and estimating reaction rate and equilibrium constants is still a main activity of chemical reaction engineering today. For a simple bimolecular reaction, one has

$$
\begin{aligned}
&A + B \rightarrow C \\
&r_{\text{eff}} = k_{\text{eff}}[A]^x[B]^y \\
&k_{\text{eff}} = A_{\text{eff}} e^{\frac{-E_{a,\text{eff}}}{RT}}
\end{aligned}
\tag{1}
$$

Here r_{eff} is the observed reaction rate, k_{eff} is the associated rate constant, [A] and [B] the concentrations of the reactants with x and y the observed reactions orders for each reactant, A_{eff} is the pre-exponential factor, and $E_{a,\text{eff}}$ is the reaction barrier.

In the early 1900s, researchers proposed more complex rate expressions. In 1913, Michaelis and Menten recognized that enzymatic reaction rates depend on a reversible reaction between an enzyme and a substrate producing the activated form of the enzyme (Cornish-Bowden 2015). The Langmuir-Hinshelwood-Hougen-Watson (LHHW) mechanism proposed in the 1920s considers the following elementary steps, namely, adsorption of a gas-phase molecule onto a heterogeneous catalyst surface, subsequent reaction between adsorbed species, and finally desorption of the resulting product(s) from the surface. Instead of reactions between adsorbed species, the Eley-Rideal (ER) mechanism formulated in the 1930s proposes an elementary step where a gas-phase species reacts directly with an adsorbed surface

species. Finally, the Mars-van Krevelen (MvK) mechanism describes a surface redox reaction where oxygen is provided by the metal oxide catalyst to an adsorbate reducing the metal oxide which is subsequently oxidized by adsorbed oxygen to complete the cycle. These reaction schemes form the backbone of modern surface reaction mechanisms. Because several elementary steps are postulated in these mechanisms, this modeling approach requires parameters for each step, including activation energies, pre-exponential factors, sticking coefficients for adsorption, and catalyst site densities. In early years, closed-form LHHW or ER rate expressions were derived by postulating the quasi-steady-state approximation (QSSA), partially equilibrated reactions, rate-determining step(s) (RDS), and most abundant surface intermediates (MASI). Regression to experimental data is still commonly used to estimate the parameters of these overall rate expressions. These reduced rate equations avoid the estimation of a large number of unknown parameters, but they are usually less predictive than full microkinetic models because their assumptions are often not valid over a broad range of reaction conditions.

The birth of microkinetic modeling (MKM) is often attributed to Dumesic and Rudd due to their 1993 book of "The Microkinetics of Heterogeneous Catalysis" (Dumesic and Rudd 1993). MKM breaks down a reaction mechanism into all known elementary steps without the aforementioned a priori assumptions. Instead, the rate constants of all elementary steps are estimated and the governing equations for surface and fluid (gas- or liquid-phase) species are numerically solved. Determining accurately the many kinetic parameters is a limiting factor of this modeling approach. In the early days of MKM, this requirement limited the application of MKM to simple reactions, like CO and H_2 oxidation, where enough experimental data was available to estimate individual parameters. In the late 1990s, the bond-order conservation (BOC) method (Shustorovich 1986, 1990) was employed for the first time in parametrizing MKMs (Park et al. 1999; Aghalayam et al. 2000). BOC estimates activation energies from heats of adsorption and even allows for the estimation of molecular heats of adsorption from the atomic ones, such as C, O, N, H, and S. These atomic heats of adsorption are in essence descriptors of thermochemistry and activation energies. In addition, due to the possible dependence of heats of adsorption on coverage (the fraction of surface sites occupied by a species), BOC provides a systematic way of incorporating coverage effects (non-idealities) on the activation energies and obtaining insights into reactivity (Park et al. 1999; Aghalayam et al. 2000).

BOC is a simple but empirical method whose accuracy remains (in most cases) unknown until tested against experimental data. The accuracy of this method becomes even more questionable for complex species (dehydrogenated intermediates, multifunctional molecules) anchoring on a surface in many, often multi-dentate configurations. These shortcomings obviously limit BOC's applicability. With the parallel evolution of both computing power and computational techniques, such as density functional theory (DFT), first-principles-based MKM is flourishing. Parameterization of moderate size reaction mechanisms using DFT is now possible; yet, parameterization of large reaction mechanisms remains a challenge (Sutton and Vlachos 2015). MKM results enable clear assessment of reaction paths, MASI,

RDS, and provide guidelines for catalyst design that promote desired reaction paths with the ultimate goal to improve catalyst selectivity and activity.

This chapter focuses on describing the basics of MKM. It also details how kinetic and thermodynamic parameters are derived from first principles using transition state theory and statistical mechanics while also addressing thermodynamic consistency. Finally, the chapter covers analysis techniques that provide key insights into the reaction mechanism. We focus primarily on the ammonia decomposition chemistry as an illustrative example.

2   Reaction Mechanism

A reaction mechanism, composed of elementary steps, is key to MKM. As an example, the overall ammonia synthesis reaction from nitrogen and hydrogen is written as

$$N_2 + 3H_2 \rightleftarrows 2NH_3 \tag{2}$$

An elementary step is a reaction that describes a molecular event and cannot be broken down into more steps. In the case of ammonia synthesis, one starts with the adsorption of each of the reactants on the active site of the catalyst surface:

$$N_2 + * \rightleftarrows N_2^* \qquad\qquad H_2 + * \rightleftarrows H_2^* \tag{3}$$

Here * represents an unoccupied active site while N_2^* and H_2^* represent the adsorbed hydrogen and nitrogen molecules on the surface. Subsequently, each species dissociates from its molecular form to its atomic form

$$N_2^* + * \rightleftarrows 2N^* \qquad\qquad H_2^* + * \rightleftarrows 2H^* \tag{4}$$

Hydrogen atoms are then added sequentially to nitrogen

$$NH_{x-1}^* + H^* \rightleftarrows NH_x^* + * \tag{5}$$

Ammonia finally desorbs from the surface

$$NH_3^* \rightleftarrows NH_3 + * \tag{6}$$

Alternative reaction paths may also be considered. For example, does molecular hydrogen associate directly with N* to form adsorbed NH_2? Does H_2 physisorb or dissociate directly? Previous studies have shown that hydrogen dissociates rapidly on active metal surfaces, and thus, the hydrogen adsorption is taken as dissociative

$$H_2 + 2^* \rightleftarrows 2H^* \tag{7}$$

As another example, the CO_2 formation in the water-gas-shift reaction was assumed in early days to proceed via water dissociation to oxygen followed by oxidation of CO: $H_2O \rightarrow OH + H$, $OH \rightarrow O + H$, and $CO + O \rightarrow CO_2$. It was subsequently realized that the carboxyl pathway, following partial water dissociation, i.e., $H_2O \rightarrow OH + H$, $CO + OH \rightarrow COOH \rightarrow CO_2 + H$, is in fact the key (preferred) path that was missing in early mechanisms (Mhadeshwar and Vlachos 2004). These examples underscore the diversity of possible reaction schemes and showcase the experience required in assembling a valid reaction mechanism.

An implicit assumption of the above mechanisms is the presence of a single type of active site. As is often the case for heterogeneous catalysts, multiple sites of different activity usually exist on a catalyst. For example, ruthenium (Ru), a well-known ammonia synthesis catalyst, has at least two well-described sites; terrace sites on planes of low Miller indices and B5 step sites located at the edges of nanocrystals. Representing both site types requires parallel sets of elementary steps.

Terrace Sites

$$N_2 + *(t) \underset{k^t_{1_b}}{\overset{k^t_{1_f}}{\rightleftarrows}} N_2(t)$$

$$N_2(t) + *(t) \underset{k^t_{2_b}}{\overset{k^t_{2_f}}{\rightleftarrows}} 2N(t)$$

$$H_2 + 2^*(t) \underset{k^t_{3_b}}{\overset{k^t_{3_f}}{\rightleftarrows}} 2H(t)$$

$$N(t) + H(t) \underset{k^t_{4_b}}{\overset{k^t_{4_f}}{\rightleftarrows}} NH(t) + *(t)$$

$$NH(t) + H(t) \underset{k^t_{5_b}}{\overset{k^t_{5_f}}{\rightleftarrows}} NH_2(t) + *(t)$$

$$NH_2(t) + H(t) \underset{k^t_{6_b}}{\overset{k^t_{6_f}}{\rightleftarrows}} NH_3(t) + *(t)$$

$$NH_3(t) \underset{k^t_{7_b}}{\overset{k^t_{7_f}}{\rightleftarrows}} NH_3 + *(t)$$

Step Sites

$$N_2 + *(s) \underset{k^s_{1_b}}{\overset{k^s_{1_f}}{\rightleftarrows}} N_2(s)$$

$$N_2(s) + *(s) \underset{k^s_{2_b}}{\overset{k^s_{2_f}}{\rightleftarrows}} 2N(s)$$

$$H_2 + 2^*(s) \underset{k^s_{3_b}}{\overset{k^s_{3_f}}{\rightleftarrows}} 2H(s)$$

$$N(s) + H(s) \underset{k^s_{4_b}}{\overset{k^s_{4_f}}{\rightleftarrows}} NH(s) + *(s)$$

$$NH(s) + H(s) \underset{k^s_{5_b}}{\overset{k^s_{5_f}}{\rightleftarrows}} NH_2(s) + *(s)$$

$$NH_2(s) + H(s) \underset{k^s_{6_b}}{\overset{k^s_{6_f}}{\rightleftarrows}} NH_3(s) + *(s)$$

$$NH_3(s) \underset{k^t_{7_b}}{\overset{k^t_{7_f}}{\rightleftarrows}} NH_3 + *(s)$$

Diffusion (Step ↔ Terrace)

$$N_2(t) + *(s) \underset{k^{s \to t}_{1_b}}{\overset{k^{t \to s}_{1_f}}{\rightleftarrows}} N_2(s) + *(t)$$

$$N(t) + *(s) \underset{k^{s \to t}_{2_b}}{\overset{k^{t \to s}_{2_f}}{\rightleftarrows}} N(s) + *(t)$$

$$H(t) + *(s) \underset{k^{s \to t}_{3_b}}{\overset{k^{t \to s}_{3_f}}{\rightleftarrows}} H(s) + *(t)$$

$$NH(t) + *(s) \underset{k^{s \to t}_{4_b}}{\overset{k^{t \to s}_{4_f}}{\rightleftarrows}} NH(s) + *(t)$$

$$NH_2(t) + *(s) \underset{k^{s \to t}_{5_b}}{\overset{k^{t \to s}_{5_f}}{\rightleftarrows}} NH_2(s) + *(t)$$

$$NH_3(t) + *(s) \underset{k^{s \to t}_{6_b}}{\overset{k^{t \to s}_{6_f}}{\rightleftarrows}} NH_3(s) + *(t)$$

$$(8)$$

The simple ammonia synthesis reaction is represented by 14 reversible adsorption, surface, and desorption reactions, and six reversible diffusion steps between sites (these represent the overall process; the distance among sites can also be important). In total, there are 20 distinct gas and surface species. Bimolecular reactions between species in adjacent but different sites, e.g., a terrace and a step site, entail interfacial reactions that require different parameters compared to reactions between species on the same type of site. An example of such study revealing this complexity is the water-gas shift reaction (Stamatakis et al. 2011). As a result, while consideration of multiple sites is in principle

straightforward, the number of reactions and species in these multisite or multifunctional catalysts increases greatly, making the parameterization task more demanding.

3   Mean-Field Approximation

A common assumption in MKM is the so-called mean-field approximation. In mean-field MKM, all the surface species and active sites are assumed to be uniformly distributed at the nanometer scale; this does not imply a fully spatially uniform system. Indeed, nonuniformity over larger, reactor scales can easily be handled in the mean-field framework using a spatially distributed reactor model. Reaction-diffusion partial differential equations-based (mean-field) models on catalyst surfaces have also been employed to describe long-range patterns, e.g., traveling waves (Raimondeau and Vlachos 2003). The application of Fickian diffusion is often another approximation in these diffusion-reaction models.

In mean-field modeling, rate constants depend on the "mean," i.e., the average, coverage, rather than the microscopic distribution of adsorbates and catalyst sites. This does not preclude the ability to define different catalyst sites or diffusion from one type of site to another; rather this is done based on the average density of each site type. In contrast, kinetic Monte-Carlo (KMC) methods describe surface species and catalyst sites with atomic resolution and explicitly account for nanoscale inhomogeneities (Stamatakis and Vlachos 2012). The large computational demand and need for expert knowledge of statistical mechanics for data analysis make currently widespread use of KMC simulation rather limited. In this chapter, we present the mean field approach because it is instructive and the only viable method for large mechanisms. In a later section, there is a discussion about adding lateral interactions (coverage effects). For the interested reader, other approximations for the distribution of adsorbates are discussed in Hill (1962). What is important is that all these methods for spatially averaging microscopic information allows us to compute the reaction rate at different levels of accuracy. In mean-field, we use a uniform distribution of adsorbates; in KMC, we solve for the exact distribution of adsorbates (within numerical accuracy of the stochastic method) subject to the overall kinetics imposed as constraints (this is not an equilibrium Monte Carlo calculation typically taught in statistical mechanics books; rather it is often far from equilibrium conditions).

4   Mass-Action Kinetics

Once the reaction mechanism is described in elementary steps, mass-action kinetics is typically used to describe each step. The rate (r_i) of a single elementary step i can be written as:

$$r_i = k_i \prod_j \theta_j^{\alpha_{ij}} \tag{9}$$

where k_i is the rate constant of elementary step i, while θ_j are the coverages of surface species j with α_{ji} their stoichiometric coefficients in elementary step i.

5   Transition-State Theory (TST)

TST assumes that for any activated elementary step, there exists equilibrium between reactants and the activated complex. In 1910, French chemist Marcelin proposed that the rate constant is proportional to the exponential of the change in standard Gibbs free energy from reactant to transition state ($\Delta G^{\circ}{}_i^{\ddagger}$) with the pre-exponential factor as the proportionality constant (Laidler and King 1983)

$$k_i \propto e^{-\left(\frac{\Delta G^{\circ \ddagger}_i}{k_B T}\right)} \qquad\qquad A_i = \frac{k_B T}{h} \tag{10}$$

The pre-exponential part of this expression ($k_B T/h$), where k_B is Boltzmann's constant, h is Planck constant and T is temperature, approximates the contribution from the frequency associated with the negative curvature of the potential energy surface in the direction of the reaction coordinate (Salciccioli et al. 2011). By breaking down the Gibbs free energy change to its enthalpic and entropic contributions (see Fig. 1), the above equation can be rewritten as

$$k_i = \left(\frac{k_B T}{h}\right) e^{-\left(\frac{\Delta G^{\circ \ddagger}_i}{k_B T}\right)} = \left(\frac{k_B T}{h}\right) e^{\left(\frac{-\Delta H^{\circ \ddagger}_i}{k_B T}\right)} e^{\left(\frac{\Delta S^{\circ \ddagger}_i}{k_B}\right)} \tag{11}$$

Fig. 1 Schematic of reaction coordinate. (Taken from Salciccioli et al. (2011))
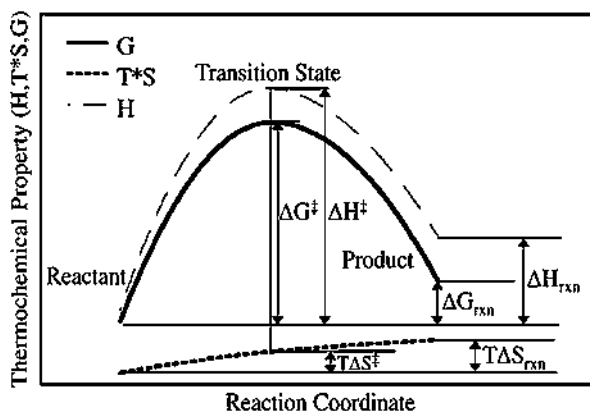
As will be explained later, the resulting equation connects first-principles DFT and statistical mechanics to the rate constant. TST is applied to all surface processes that are activated. Examples of such processes include Langmuir-Hinshelwood surface reaction, desorption, and surface diffusion.

6   Collision Theory and Adsorption Rate

Unlike activated processes described using TST (see above), adsorption is often nonactivated. Collision theory defines the molecular flux (N_j) of adsorbing gases on a catalyst surface

$$N_{j,collisions} = \frac{P_j}{\sqrt{2\pi m_j k_B T}} \ [=] \left(\frac{1}{m^2 \cdot s}\right) \tag{12}$$

Here P_j is the partial pressure of species j, m_j is the molecular weight of species j, k_B is the Boltzmann constant, and T is the temperature. Since not all collisions result in a molecule adsorbing on the surface, the adsorption rate (r_j) is obtained by the product of the incident molecular flux and a dimensionless probability factor known as the sticking coefficient ($0 \leq \gamma \leq 1$)

$$r_j = N_{j,collisions} \cdot \gamma_j = \frac{P_j \gamma_j}{\sqrt{2\pi m_j k_B T}} \tag{13}$$

The adsorption rate constant (k_j) can then be expressed as

$$k_j = \frac{\gamma_j}{\sqrt{2\pi m_j k_B T}} \ [=] \left(\frac{1}{Pa \cdot m^2 \cdot s}\right) \tag{14}$$

An assumption implicit in the sticking coefficient is that it is much less than one (Coltrin et al. 1996; Motz and Wise 1960). That suggests that the gases near the surface behave randomly and are not impacted by surface reactions. As the sticking coefficient nears one, the implicit Maxwell velocity distribution becomes distorted and the rate of adsorption increases. To correct for this, the simple sticking coefficient is replaced with a Motz-Wise corrected sticking coefficient

$$\gamma_{j,corrected} = \frac{\gamma_j}{1 - \gamma_j/2} \tag{15}$$

The corrected versus uncorrected sticking coefficient is shown in Fig. 2. For activated adsorption, the rate constant is multiplied by a Boltzmann factor to account for the reduced probability of incoming molecules to overcome the adsorption barrier and stick to the surface.

Fig. 2 Comparison of an uncorrected sticking coefficient with the Motz-Wise corrected sticking coefficient. (Motz and Wise 1960; Coltrin et al. 1996)
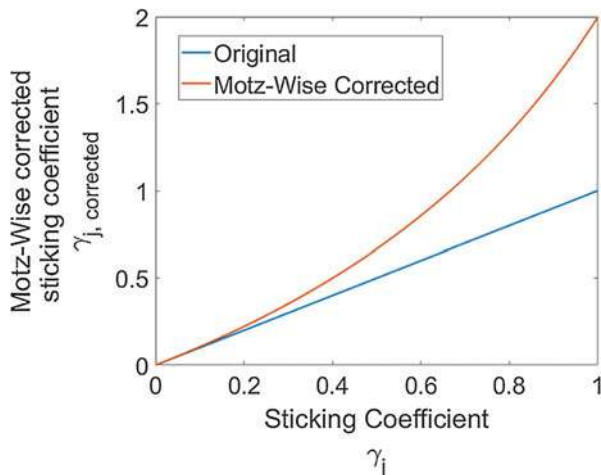
7 Thermodynamic Consistency

A microkinetic model for heterogeneous catalysis involves both gas and surface species. As thermodynamic data are developed for each phase often separately, it is critical to ensure thermodynamic consistency between surface and gas-phase. This constraint is illustrated with the thermodynamic cycle of Fig. 3. Since DFT is not very accurate for gas-phase species, use of DFT for estimation of gas-phase thermochemistry can lead to inaccurate free energies and large errors in overall equilibrium constants. Instead, high-level ab initio or experimental methods (see NIST database) are used to estimate – with chemical accuracy – gas-phase thermochemistry. On the other hand, surface thermochemistry is typically calculated at the DFT theory level since the investigated surface models are of considerable size and higher-level theoretical methods are often not applicable due to the very unfavorable scaling of computational costs. The use of a different methodology for each phase causes a mismatch between gas and surface thermochemistries. There are two ways to remedy this mismatch.

7.1 Adjusting Surface Reaction Properties

The surface reaction state properties can be related to those of the corresponding gas-phase reactions by a thermodynamic cycle (Fig. 3)

$$\zeta_{i,surf}(T,\theta) = \zeta_{i,gas}(T) + \Delta\zeta_{i,ads}(T,\theta) \tag{16}$$

Here ζ is a surface or gas species thermodynamic (state) property (i.e., enthalpy, entropy, or free energy) of reaction i and $\Delta\zeta_{i,ads}$ is the difference in the correspond-

ing adsorption properties of products from those of reactants for reaction i. Such thermodynamic cycles can be written for each reaction.
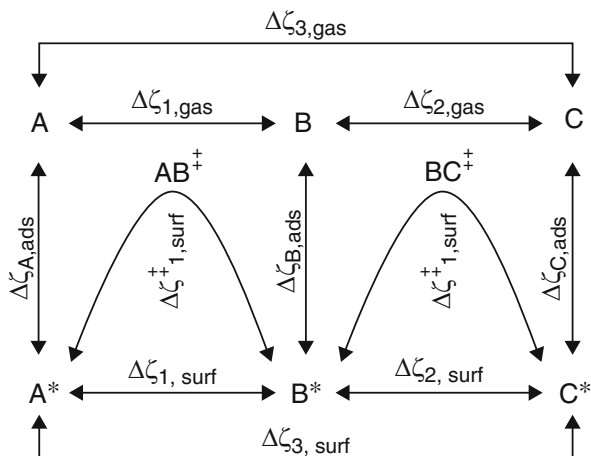
Practically, temperature dependent (NASA) polynomials, describing the thermochemistry of gas-phase species, are commonly used as given below. Surface thermochemistry is described by similar polynomials with the same coefficients a_1-a_5. The coefficients $a_{6,surf}$ and $a_{7,surf}$ of surface species are related to those in the gas a_6 and a_7 as follows (it can easily be shown that this choice yields thermodynamic consistency).

$$\frac{C_p^\circ}{R} = a_1 + a_2 T + a_3 T^2 + a_4 T^3 + a_5 T^4$$

$$\frac{H^\circ}{RT} = a_1 + a_2 \frac{T}{2} + a_3 \frac{T^2}{3} + a_4 \frac{T^3}{4} + a_5 \frac{T^4}{5} + \frac{a_6}{T} \qquad \begin{aligned} a_{6,surf} &= a_{6,gas} + \frac{\Delta H^\circ_{ads}(T,\theta)}{R} \\ a_{7,surf} &= a_{7,gas} + \frac{\Delta S^\circ_{ads}(T,\theta)}{R} \end{aligned}$$

$$\frac{S^\circ}{R} = a_1 \ln T + a_2 T + a_3 \frac{T^2}{2} + a_4 \frac{T^3}{3} + a_5 \frac{T^4}{4} + a_7$$

$$(17)$$

Here ΔH°_{ads} and ΔS°_{ads} are the standard heat and entropy of adsorption, respectively. When adsorption data are well-known, e.g., from experiments, one estimates $a_{6,surf}$ and $a_{7,surf}$. Then the surface reaction thermochemistry is estimated. A more detailed discussion is found in Mhadeshwar et al. (2003).

7.2   Adjusting Adsorption Properties

It is generally accepted that DFT is more accurate in estimating surface reaction properties due to cancelation of errors. Given this fact, one can estimate the surface thermochemistry via DFT and adjust the adsorption properties to close the thermodynamic cycles (Fig. 3). The thermodynamic properties of the surface species are determined by regressing the heat capacity DFT data computed via

statistical mechanics for the surface species. To achieve thermodynamic consistency, a reference set of gas species is selected as a basis such that (1) all the atoms in the surface species are represented in the reference set, (2) there are at least as many gas species in the reference set as the different atoms in the reference set, and (3) the reference gas species have well-known formation energies. These reference gases form the basis set upon which the surface species thermochemistry data is referenced to achieve thermodynamic consistency. The difference in formation energy of a gas species in the basis set computed by DFT from its experimental or high-level ab initio value represents the error in the formation energy of the adsorbed species. This difference is attributed to each atom of the reference species via least squares fit. Adding the total correction for each species to the $a_{6,surf}$ and $a_{7,surf}$ coefficients results in a surface species thermodynamic polynomial with the same reference state as the gas species (Nicolaides et al. 1996; Wijaya et al. 2003).

The error introduced by the least squares fitting in this approach can be quantified by a fourfold cross-validation. Using this methodology, a root mean squared error of ~1.65 kcal/mol is found for a reference set of 13 gas-phase molecules (C_2H_4, C_2H_6, CH_4, H_2, H_2O, $(CH_2)_4O$, $(CH_3)_2CO$, C_4H_{10}, C_5H_{12}, C_5H_{10}, C_3H_8, $C_2H_5OCH_3$, CH_3OCH_3). This value is a good indicator of the order-of-magnitude of the error introduced.

8   Surface Species Lateral Interactions

One of the shortcomings of the mean-field model is that it ignores many-body interactions. To correct this deficiency, the enthalpy of a surface species is modified based on the coverage of each adsorbate. DFT is used to calculate the adsorption energy versus coverage. The change in adsorption energy is regressed and used in the MKM. An example is shown in Fig. 4.

Fig. 4 N-N (self) pair lateral interaction adsorption energy [kcal/mol] versus N coverage varying linearly (blue) or piecewise with a zero-effect below 0.25 coverage of N (orange). (Nørskov 1993; Guo and Vlachos 2013, 2015a)
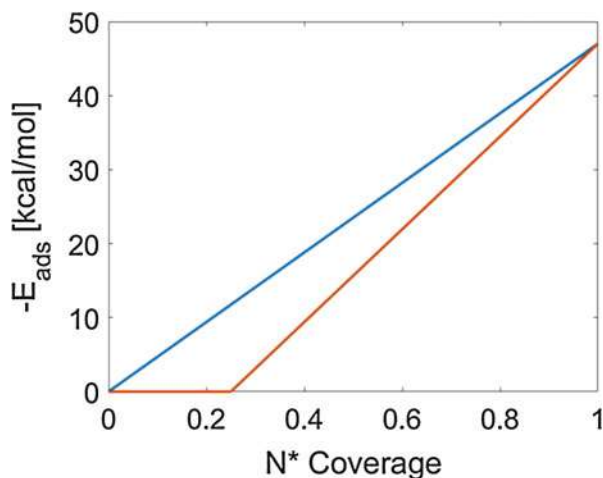
9   Kinetic and Thermodynamic Parameter Estimation

9.1   Partition Functions

The translational, rotational, vibrational, electronic, and nuclear partition functions contribute to the overall as

$$Q_{tot} = q_{trans} \cdot q_{rot} \cdot q_{vib} \cdot q_{elect} \cdot q_{nuclear} \qquad (18)$$

The model often uses the first three partition functions since the electronic and nuclear contributions are much smaller than the uncertainty in the first three. The translational partition function is defined as

$$q_{trans} = \frac{V}{\Lambda^3}, \text{ where} \left. \begin{array}{l} V \text{ is the volume in which the molecule is contained} \\ \Lambda \text{ is the de Broglie } wavelength \text{ and defined as } \Lambda = \sqrt{\frac{h^2}{2\pi m k_B T}} \end{array} \right.$$
$$h = \text{Planck}'s \text{ constant}; \quad k_B = \text{Boltzmann}'s \text{ constant};$$
$$m = \text{mass of } a \text{ molecule}; \quad T = \text{temperature}$$
$$(19)$$

The relationship can be further simplified for a gas adsorbing on the surface of a heterogeneous catalyst into

$$q_{trans} = q_{trans,2D} \cdot \ell_z \cdot \frac{\sqrt{2\pi m k_B T}}{h}; \quad (\ell_z \text{ distance of the gas from the surface})$$
$$q_{trans,2D} = A_{st} \cdot \frac{2\pi m k_B T}{h^2}; \quad \left(\begin{array}{l} A_{st} = \text{the effective area of the site } (s) \\ \text{where the reaction takes place.} \end{array} \right)$$
$$(20)$$

The rotational molecular partition function for a linear molecule is

$$q_{rot} = \frac{8\pi^2 I k_B T}{\sigma h} \qquad (21)$$

For a nonlinear molecule with three moments of inertia, one has

$$q_{rot} = \frac{\left(\pi \cdot I_x \cdot I_y \cdot I_z \right)^{1/2}}{\sigma} \left(\frac{8\pi^2 k_B T}{h^2} \right)^{\frac{3}{2}} \qquad (22)$$

where σ is the number of indistinguishable orientations (i.e., symmetry number) of the molecule and I is the moment of inertia. Lastly, the vibrational contribution is the product of the contribution of each vibrational mode.

$$q_{vib,ZPE} = \prod_k \frac{\exp\left(-\frac{h \cdot \nu_k}{2 \cdot k_B \cdot T}\right)}{1-\exp\left(-\frac{h \cdot \nu_k}{k_B \cdot T}\right)}; \quad \text{where} \left.\begin{array}{l}\\ \end{array}\right\} \quad \begin{array}{l} \nu_k = \text{Each of the vibration frequencies of} \\ \qquad \text{that molecular configuration } \left(s^{-1}\right) \end{array}$$

$$E_{ZPE} = \sum_k \frac{h \cdot \nu_k}{2}$$

Set zero-point energy to 0 in the vibrational partition function and account for it in the *activation* energy.

$$q_{vib} = \prod_k \frac{1}{1-\exp\left(-\frac{h \cdot \nu_k}{k_B \cdot T}\right)}$$

$$(23)$$

9.2   Thermodynamic-Temperature Relationships

Thermodynamic quantities such as heat capacity, enthalpy, and entropy expressed as a function of temperature (Eq. 17) are necessary to estimate reverse rate constants. While it is possible to calculate these quantities directly using the statistical mechanics equations while solving the MKM, it is more efficient and common to use simple polynomial relationships. The statistical mechanics treatment is found in more detail in books, e.g., Sandler (2010) and Hill (1962). The relevant quantities are calculated as follows.

Heat Capacity

Linear molecules

$$\frac{C_p}{R} = \underbrace{\sum_i \left(\frac{\Theta_{v,i}}{T}\right)^2 \frac{e^{-\Theta_{v,i}/T}}{\left(1-e^{-\Theta_{v,i}/T}\right)^2}}_{\substack{\text{Vibrational} \\ \text{contribution}}} + \underbrace{\frac{3}{2}}_{\substack{\text{Rotational} \\ \text{contribution}}} + \underbrace{1}_{\substack{\text{Translational} \\ \text{contribution}}} + \underbrace{1}_{\substack{C_v \text{ to } C_p \\ C_v = C_p + R \\ \frac{C_v}{R} = \frac{C_p}{R} + 1}}$$

$$(24)$$

Non-linear molecules

$$\frac{C_p}{R} = \underbrace{\sum_i \left(\frac{\Theta_{v,i}}{T}\right)^2 \frac{e^{-\Theta_{v,i}/T}}{\left(1-e^{-\Theta_{v,i}/T}\right)^2}}_{\substack{\text{Vibrational} \\ \text{contribution}}} + \underbrace{\frac{3}{2}}_{\substack{\text{Rotational} \\ \text{contribution}}} + \underbrace{\frac{3}{2}}_{\substack{\text{Translational} \\ \text{contribution}}}$$

$$(25)$$

$$+ \underbrace{1}_{\substack{C_v \text{ to } C_p \\ C_v = C_p + R \\ \frac{C_v}{R} = \frac{C_p}{R} + 1}}$$

Enthalpy

Linear molecules

$$\frac{H}{RT} = \underbrace{\sum_i \frac{\Theta_{v,i}}{2T} + \frac{\Theta_{v,i}}{T} \frac{e^{-\Theta_{v,i}/T}}{1 - e^{-\Theta_{v,i}/T}}}_{\substack{\text{Vibrational} \\ \text{contribution}}} + \underbrace{\frac{3}{2}}_{\substack{\text{Rotational} \\ \text{contribution}}} + \underbrace{1}_{\substack{\text{Translational} \\ \text{contribution}}}$$

$$+ \underbrace{RT}_{\substack{\text{Convert U to H} \\ H = U + PV}} + \underbrace{E_{\text{zero point}} + E_{\text{dft total}}}_{\text{Energies from DFT}}$$

(26)

Non-linear molecules

$$\frac{H}{RT} = \underbrace{\sum_i \frac{\Theta_{v,i}}{2T} + \frac{\Theta_{v,i}}{T} \frac{e^{-\Theta_{v,i}/T}}{1 - e^{-\Theta_{v,i}/T}}}_{\substack{\text{Vibrational} \\ \text{contribution}}} + \underbrace{\frac{3}{2}}_{\substack{\text{Rotational} \\ \text{contribution}}} + \underbrace{\frac{3}{2}}_{\substack{\text{Translational} \\ \text{contribution}}}$$

$$+ \underbrace{RT}_{\substack{\text{Convert U to H} \\ H = U + PV}} + \underbrace{E_{\text{zero point}} + E_{\text{dft total}}}_{\text{Energies from DFT}}$$

(27)

Entropy

Linear molecules

$$\frac{S}{R} = \underbrace{\sum_i \frac{\Theta_{v,i}}{T} \frac{e^{-\Theta_{v,i}/T}}{1 - e^{-\Theta_{v,i}/T}} - \ln\left(1 - e^{-\Theta_{v,i}/T}\right)}_{\substack{\text{Vibrational} \\ \text{contribution}}} + \underbrace{1 + \ln\frac{T}{\sigma \Theta_r}}_{\substack{\text{Rotational} \\ \text{contribution}}}$$

$$+ \underbrace{\frac{5}{2} + \ln\left[\left(\frac{2\pi \sum_i m_i k_B T}{h^2}\right)^{\frac{3}{2}} \frac{1}{P}\right]}_{\substack{\text{Translational} \\ \text{contribution}}}$$

(28)

Non-linear molecules

$$\frac{S}{R} = \underbrace{\sum_i \frac{\Theta_{v,i}}{T} \frac{e^{-\Theta_{v,i}/T}}{1 - e^{-\Theta_{v,i}/T}} - \ln\left(1 - e^{-\Theta_{v,i}/T}\right)}_{\substack{\text{Vibrational} \\ \text{contribution}}} + \underbrace{\frac{3}{2} + \ln\left(\frac{\sqrt{\pi}}{\sigma}\sqrt{\frac{T^3}{\Theta_x\Theta_y\Theta_z}}\right)}_{\substack{\text{Rotational} \\ \text{contribution}}}$$

$$\underbrace{+\frac{5}{2} + \ln\left[\left(\frac{2\pi \sum_i m_i k_B T}{h^2}\right)^{\frac{3}{2}} \frac{1}{P}\right]}_{\substack{\text{Translational} \\ \text{contribution}}}$$

where the vibrational and inertial temperatures are defined as

$\Theta_{v,i} = \frac{v_i h}{k_B}$; $v_i = i^{\text{th}}$ vibrational frequency $\left[s^{-1}\right]$

$\Theta_k = \frac{8\pi^2 I_k k_B}{h^2}$; $I_k =$ moment of inertia in the k-plane

$$(29)$$

Gas species include all three molecular partition functions while surface species often include only the vibrational component. Also, since PV is omitted for surface species (Reuter and Scheffler 2001; Mhadeshwar et al. 2004), the terms that convert C_V to C_P and U to H are not included in the calculation for surface species

$$H = U + PV; \quad \text{If } PV \ll 1 \quad \therefore \quad U \approx H;$$
$$U = \int C_V dT \quad and \quad H = \int C_P dT$$
$$\int C_V dT = \int C_P dT \quad \therefore \quad C_V = C_P$$

$$(30)$$

The heat capacity NASA fourth-order polynomial (Eq. 17) is commonly used and fitted to heat capacity-temperature data over a desired temperature range. Depending on the size of the temperature range and the goodness of the fit, it may be appropriate to divide the polynomial into two temperature ranges while ensuring that there is no discontinuity between the two.

9.3 Activation Energy and Pre-exponential Factor

Activation energies are calculated from the heats of formation of the transition state [TS] and the reactants as

$$E_a = \Delta E^{\ddagger} = FE\,[\text{TS}] - FE\,[\text{Reactants}]$$

$$FE\,[\text{TS}] = E_{DFTE}\,[\text{TS}] + E_{ZPE}\,[\text{TS}] - E_{DFTE}\,\big[\text{Catalyst Surface}\big]$$
$$\quad - \sum_i E\,[\text{Gas}_i]$$

$$FE\,[\text{Reactants}] = \sum_i \big(E_{DFTE}\,[\text{Reactant}_i] + E_{ZPE}\,[\text{Reactant}_i]$$
$$\quad - E_{DFTE}\,\big[\text{Catalyst Surface}\big] - E\,[\text{Gas}_i]\big) \tag{31}$$

$$FE = \text{Formation Energy};\ E_{DFTE} = \text{DFT\ Energy};$$

$$E_{ZPE} = \sum_k \frac{h \cdot v_k}{2} = \text{Zero Point Energy}$$

$$E\,[\text{Gas}] = \text{Reference gas DFT computed energy including zero}$$
point energy.

The rate constant for Langmuir-Hinshelwood surface reactions between two adsorbed species ($X^* + Y^* \to Z^* + ^*$) can be written as

$$k_{surf}^{fwd} = \underbrace{\frac{q_{vib}^{\ddagger}}{q_{vib,X} \cdot q_{vib,Y}} \cdot \frac{k_B \cdot T}{h}}_{\text{Pre-exponential factor}} \exp \underbrace{\left(-\frac{\Delta E^{\ddagger}}{k_B \cdot T}\right)}_{\Delta E^{\ddagger} = \ \text{Activation Energy}} \ \big[\sec^{-1}\big] \tag{32}$$

$$k_{surf}^{rev} = \frac{q_{vib}^{\ddagger}}{q_{vib,Z}} \cdot \frac{k_B \cdot T}{h} \exp\left(-\frac{\Delta E^{\ddagger} - \Delta E_{rxn}}{k_B \cdot T}\right) \ \big[\sec^{-1}\big]. \tag{33}$$

It is worth noting that surface species often have no translational or rotational degrees of freedom and have instead vibrational degrees of freedom (mobile species exist when adsorption is weak).

The expression for the rate constant of ER reactions ($X_{gas} + Y^* \to Z^*$) is

$$k_{\text{Eley-Rideal}}^{fwd} = \frac{q_{vib}^{\ddagger}}{Q_{gas} \cdot q_{vib,Y^*}} \cdot \frac{p_{gas} A_{st}}{\sqrt{2\pi m_X k_B T}} \exp\left(-\frac{\Delta E^{\ddagger}}{k_B T}\right) \ \big[\sec^{-1}\big] \tag{34}$$

$$Q_{gas} = q_{trans2D,X_{Gas}} \cdot q_{rot,X_{Gas}} \cdot q_{vib,X_{Gas}}$$

$$k_{\text{Eley-Rideal}}^{rev} = \frac{q_{vib}^{\ddagger}}{q_{vib,Z^*}} \cdot \frac{k_B T}{h} \exp\left(-\frac{\Delta E^{\ddagger} - \Delta E_{rxn}}{k_B T}\right) \ \big[\sec^{-1}\big] \tag{35}$$

These equations depend on the 2D-translational partition function and the effective area of the reaction site(s) (A_{st}) which is the area of the catalyst site on which adsorption takes place.

Finally, the rate constant for both activated and non-activated adsorption of gases to/from the catalyst surface described by the LHHW mechanism ($X_{gas} + * \rightarrow X*$) is calculated as

$$k^{fwd}_{\text{Activated adsorption}} = \frac{q^{\ddagger}_{vib}}{Q_{gas}} \cdot \frac{A_{st}}{\sqrt{2\pi m_X k_B T}} \exp\left(-\frac{\Delta E^{\ddagger}}{k_B T}\right) \quad \left[\sec^{-1} \cdot \text{pressure}^{-1}\right]$$

$$Q_{gas} = q_{trans2D,X_{Gas}} \cdot q_{rot,X_{Gas}} \cdot q_{vib,XGas} \tag{36}$$

$$k^{rev}_{\text{Activated adsorption}} = \frac{q^{\ddagger}_{vib}}{q_{vib,Z^*}} \cdot \frac{k_B T}{h} \exp\left(-\frac{\Delta E^{\ddagger} - \Delta E_{ads}}{k_B T}\right) \quad \left[\sec^{-1}\right] \tag{37}$$

$$k^{fwd}_{\text{Non-activated adsorption}} = \frac{A_{st}}{\sqrt{2\pi m_X k_B T}} \quad \left[\sec^{-1} \cdot \text{pressure}^{-1}\right]. \tag{38}$$

Alternatively, the reverse rate constant for the aforementioned reactions can also be calculated from the equilibrium constant, which in turn is calculated from the Gibbs free energy

$$K_{eq} = \frac{k^{fwd}}{k^{rev}} \qquad\qquad K_{eq} = e^{\frac{-\Delta G}{RT}} = e^{\frac{-(\Delta H - T\Delta S)}{RT}} = e^{-\left(\frac{\Delta H}{RT} - \frac{\Delta S}{R}\right)} \tag{39}$$

10   Brønsted–Evans–Polanyi (BEP) Relationships

DFT calculations become impractical for large reaction networks. To overcome this challenge, BEP relationships can be used as they enable the prediction of reaction barriers from the heat of reactions for an entire homologous series of reactions (i.e., reactions with similar transition states). A BEP describes the reaction barrier $\Delta E_{a,i}$ of reaction i as a linear function of its heat of reaction

$$\Delta E_{a,i} = \omega \Delta H_{rxn,i} + \Delta E_0 \tag{40}$$

Here ω is the proximity factor (0–1) indicating the reaction coordinate position of the TS relative to the reactant and products (0 implies reactant-like whereas 1 product-like). The intercept ΔE_0 is called the intrinsic reaction barrier. Typical BEP relationships for NH_3 chemistry are shown in Fig. 5 (Falsig et al. 2014; Vojvodic et al. 2014). BEPs for common homologous series are available (Sutton and Vlachos 2012; Vojvodic et al. 2014; Guo and Vlachos 2015b). For new reactions, DFT calculations for a handful reactions of a homologous series can be carried out to build the BEP.

Fig. 5 BEP relationships for NH₃ decomposition. (Falsig et al. 2014; Vojvodic et al. 2014; Guo and Vlachos 2015a, b)
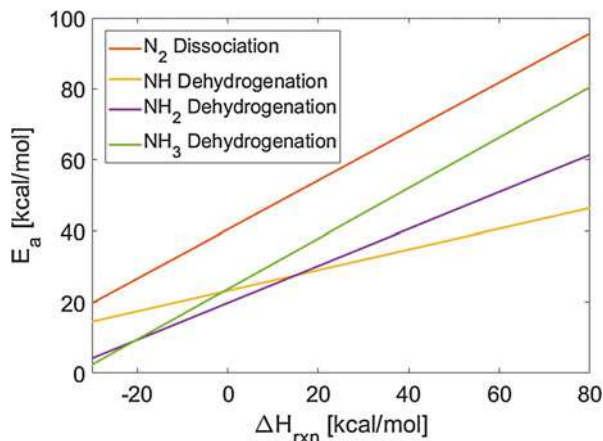
10.1   Linear Scaling Relationships (LSRs)

Linear scaling relationships (LSRs) relate the thermodynamic properties of surface species on one metal catalyst surface to those on another (reference) surface. Small adsorbates, such as AH_x, scale with the adsorption energy Q of the surface species A as

$$Q^{AH_x} = \alpha\,(x)\,Q^A + \beta \quad \alpha\,(x) = \frac{x_{\max} - x}{x_{\max}} \tag{41}$$

α is related to the valency of the species where $x_{\max}$ (C = 4, N = 3, O = 2, etc.) is the maximum number of ligands the central heteroatom can take and x is the actual number of ligands of the bound atom (Abild-Pedersen et al. 2007; Fernndez et al. 2008; Jones et al. 2008). Extended LSRs apply this relationship for small adsorbates to large, more complex, potentially multidentate (n = # dentates, j) species.

$$Q_{i,new} = Q_{i,ref} + \sum_{j=1}^{n} \gamma_{i,j} \left(Q_{new}^{ref\ species,j} - Q_{ref}^{ref\ species,j} \right) \tag{42}$$

Here $Q_{i,new}$ refers to the adsorption energy of the ith large species on a surface from a reference energy on another metal $Q_{i,ref}$, and the parenthesis is the corresponding difference for the heteroatoms via which adsorption occurs. The summation extends over all binding atoms (i) and dentates (j). This approach enables one to transfer thermochemistry from one metal to another without relying on expensive DFT calculations. An example where the surface reference species is N* (step) and the reference metal surface is Pt is shown in Fig. 6. These LSRs together with the aforementioned BEPs are used to parameterize the MKM discussed below.
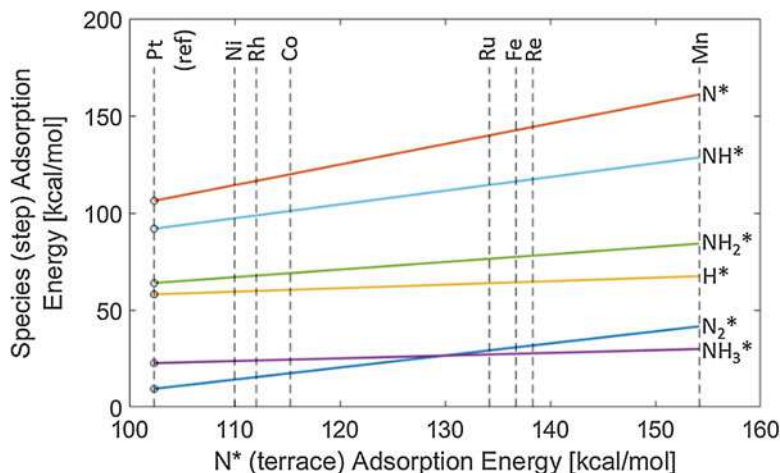
Fig. 6 LSR relationships for ammonia surface species on step sites with Pt as a reference metal and N* adsorption on terrace sites as the reference surface species. (Abild-Pedersen et al. 2007; Falsig et al. 2014)

11   Reactor Equations

The dynamic and steady-state operation of ideal reactors (e.g., batch, continuous stirred-tank reactor [CSTR], or plug flow reactor [PFR]) combines the surface reactions described earlier in this chapter with a series of equations (e.g., ODE's or algebraic) that describe the concentrations of each gas and surface species and unoccupied active catalyst sites. The general extensive property conservation equation is

$$\rho \frac{\partial \overline{\omega}}{\partial t} + \rho \mathbf{u} \cdot \nabla \overline{\omega} = -\nabla \cdot \widehat{\mathbf{J}} + \widehat{\sigma} \tag{43}$$

where $\overline{\omega}$ is an extensive property, ρ the density, $\mathbf{u}$ the velocity vector, $\widehat{\mathbf{J}}$ the diffusive flux vector, and $\widehat{\sigma}$ the source/sink terms. Here the first term on the left is accumulation followed by inflow/outflow by convection and the first term on the right is inflow/outflow by diffusion followed by a generation term (i.e., reactions). This equation is further simplified for a CSTR where the diffusion term is not required and due to perfect mixing, the convective inflow/outflow gradient is just the mass flow in/out. Using the ammonia synthesis/decomposition mechanism described in Eq. 8, and focusing on catalyst step sites with a specific catalyst active-site density (Γ [=] mol_active_sites/area) and catalyst loading (A/V [=] area/volume) results in the following rate expressions (a compact vector notation can be used but we provide an introductory presentation herein for illustration purposes).

$$
\begin{aligned}
R1_f &= k_{1_f} [N_2][*] & R1_b &= k_{1_b} \left[N_2^*\right] \\
R2_f &= k_{2_f} \left[N_2^*\right][*] & R2_b &= k_{2_b} \left[N^*\right]^2 \\
R3_f &= k_{3_f} [H_2]\left[*\right]^2 & R3_b &= k_{3_b} \left[H^*\right]^2 \\
R4_f &= k_{4_f} \left[NH_3^*\right][*] & R4_b &= k_{4_b} \left[NH_2^*\right]\left[H^*\right] \\
R5_f &= k_{5_f} \left[NH_2^*\right][*] & R5_b &= k_{5_b} \left[NH^*\right]\left[H^*\right] \\
R6_f &= k_{6_f} \left[NH^*\right][*] & R6_b &= k_{6_b} \left[N^*\right]\left[H^*\right] \\
R7_f &= k_{7_f} [NH_3][*] & R7_b &= k_{7_b} \left[NH_3^*\right]
\end{aligned}
\tag{44}
$$

A set of ordinary differential equations (ODEs) describe the time-dependent concentration of each species in the reaction mechanism:

$$
\begin{aligned}
\frac{d\left[N_2^*\right]}{dt} &= R1_f + R2_b - R1_b - R2_f \\[6pt]
\frac{d\left[N^*\right]}{dt} &= 2R2_f + R6_f - 2R2_b - R6_b \\[6pt]
\frac{d\left[H^*\right]}{dt} &= 2R3_f + R4_f + R5_f + R6_f \\
&\quad - 2R3_b - R4_b - R5_b - R6_b \\[6pt]
\frac{d\left[NH_3^*\right]}{dt} &= R4_b + R7_f - R4_f - R7_b \\[6pt]
\frac{d\left[NH_2^*\right]}{dt} &= R4_f + R5_b - R4_b - R5_f
\end{aligned}
$$

$$
\begin{aligned}
\frac{d\left[NH^*\right]}{dt} &= R5_f + R6_b - R5_b - R6_f \\[6pt]
\frac{d[N_2]}{dt} &= R1_b - R1_f + \frac{Q_{in}}{V_{reactor}}[N_2]_{feed} - \frac{Q_{out}}{V_{reactor}}[N_2] \\[6pt]
\frac{d[H_2]}{dt} &= R3_b - R3_f + \frac{Q_{in}}{V_{reactor}}[H_2]_{feed} - \frac{Q_{out}}{V_{reactor}}[H_2] \\[6pt]
\frac{d[NH_3]}{dt} &= R7_b - R7_f + \frac{Q_{in}}{V_{reactor}}[NH_3]_{feed} - \frac{Q_{out}}{V_{reactor}}[NH_3] \\[6pt]
\frac{d[*]}{dt} &= R1_b - R1_f + R2_b - R2_f + 2R3_b \\
&\quad - 2R3_f + R4_b - R4_f + R5_b - R5_f \\
&\quad + R6_b - R6_f + R7_b - R7_f
\end{aligned}
\tag{45}
$$

The concentration of surface species above is expressed in [mol/vol] but often as surface coverage by dividing the adsorbate concentrations with the total site concentration ($\Gamma \times A/V$). The gas species ODEs also include terms to account for their volumetric inlet and outlet flow rate (Q_{in} and Q_{out}). While the inlet flow rate of gases is specified, the outlet flow rate must be determined and depends on how the reactor is operated, e.g., isochoric or isobaric. Using the same ammonia equations, the total exit flow rate of the CSTR is written as:

$$
Q_{out} = Q_{in} + \left[\frac{\left(R1_b - R1_f\right)_{N_2} + \left(R3_b - R3_f\right)_{H_2} + \left(R7_b - R7_f\right)_{NH_3}}{[N_2] + [H_2] + [NH_3]} \cdot V_{reactor} \right]
\tag{46}
$$

Here the term in square brackets represents the rate of volume change due to the changing number of moles of the reactant and products and can be excluded for an isochoric reactor or included for an isobaric reactor.

If the flow $Q_{in} = 0$, then the equations represent a batch reactor. A PFR model can similarly be written. In this example, only the material balance is considered (the reactor is assumed to be isothermal).
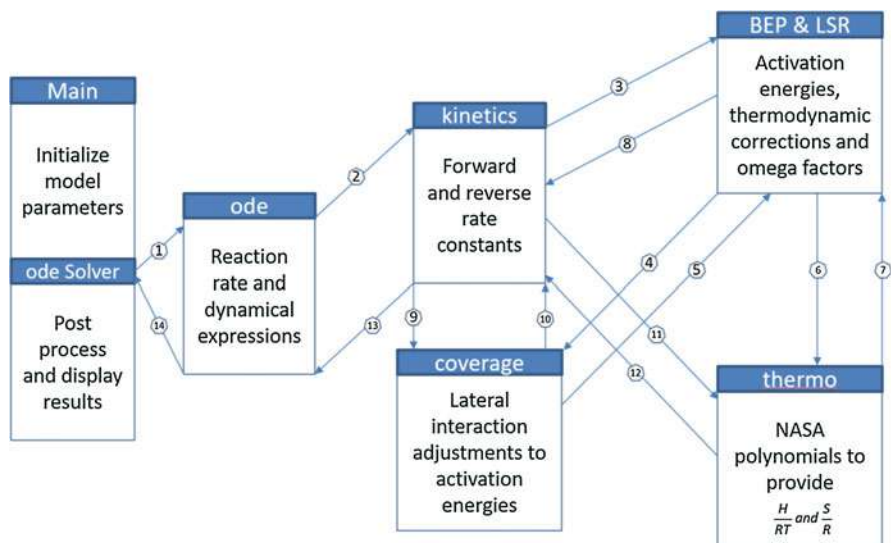
Fig. 7 Typical microkinetic model solution program flow

The system of ODEs is usually stiff, and its solution requires an implicit numerical method. A typical solution flow might look like Fig. 7 where key functions such as ODE solver, thermodynamics, kinetics, lateral interactions and BEP/LSR relationships are modularized and the modules pass relevant information of the model. The modules allow for simple troubleshooting. A module can easily be rewritten and inserted into the model to change the model capability without impacting other modules.

12 Microkinetic Model Data Analysis

The solution of the MKM yields the concentration of gaseous species, coverages of surface species, and reaction rates as well as conversion and selectivity. In addition, analysis of the MKM results can provide significant insights into the mechanism. The following sections describe such analysis tools focusing on the ammonia chemistry as a model example. (Ammonia decomposition simulation examples were performed in a Matlab based MKM using an NH_3 feed and Ru (step sites) catalyst. Model parameters: P = 1 atm, τ = 1 sec, Γ_{total} = 2.6 $\times$ 10^{-9} mol catalyst sites/cm^2, Γ_{step} = 4.4385 $\times$ 10^{-10} mol catalyst sites/cm^2, A/V = 1500 cm^3/cm^2. Kinetic parameters derived from BEP and LSR relationships. Complete model details and Matlab MKM code is found at https://github.com/VlachosGroup/NH3-Matlab-Microkinetic-Model.)
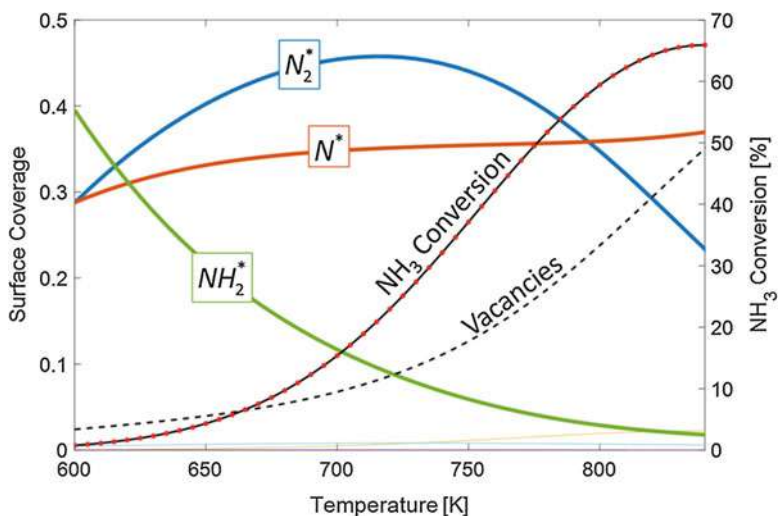
Fig. 8 Surface species coverage and conversion versus temperature for the decomposition of ammonia on ruthenium catalyst in a CSTR

12.1 Species Concentrations

The steady-state ammonia decomposition reaction MKM results versus temperature are shown in Fig. 8. The MASI changes from NH_2^* at very low temperatures to N_2^* and N^* at intermediate temperatures, to N^* and vacancies at sufficiently high temperatures.

12.2 Reaction Path Analysis (RPA)

In more complex reaction mechanisms, there are multiple paths from reactants to products and by-products. Understanding the primary paths and why one path is favored over another allow design of experiments to interrogate the mechanism and potentially predict new catalysts. The ammonia synthesis or decomposition involves only one key path. In the case of ammonia synthesis, this involves N_2 activation followed by subsequent hydrogenation and desorption. The forward, reverse and net reaction rates are computed for each elementary step and path. The net reaction rates (red bars) of all steps are all when balanced by the stoichiometric coefficients of the overall reaction. This simple comparison indicates that the solution obeys mass conservation. The large yellow and blue bars (from top to bottom in Fig. 9) for NH_3 adsorption, NH_2 dehydrogenation, NH dehydrogenation, H_2 desorption, and N_2 desorption indicate these steps are much faster than others, e.g., 2N* association
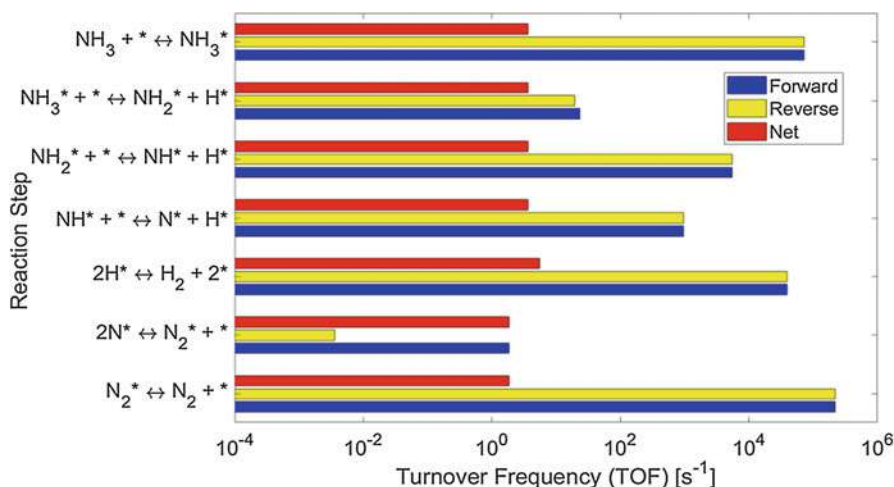
Fig. 9 Comparison of steady-state forward, reverse, and net reaction rates for ammonia decomposition at 750 K

and the first NH$_3$* dehydrogenation and will be most likely equilibrated. We will return to this point next but it suffices to state that the RDS is one or both of the two latter reaction steps.

12.3 Partial Equilibrium Analysis

It is often useful to understand which reaction steps are nearly in equilibrium since those steps tend not to be the RDS. A formal partial equilibrium analysis computes the partial equilibrium index (PEI) from the forward and backward reaction rates (Samant and Vlachos 2005).

$$PEI = \frac{R^{fwd}}{R^{fwd} + R^{rev}} \qquad (47)$$

Heuristically, it has been found that a reaction step with a PEI between 0.45 and 0.55 is in partial equilibrium. Again, using the ammonia decomposition as an example, it is clear that the nitrogen association reaction is non-equilibrated, whereas the first dehydrogenation reaction is nearly out of equilibrium. These steps are the likely RDS for this reaction. Analysis of partial equilibrium combined with MASI can be used to perform systematic mechanism reduction (Mhadeshwar and Vlachos 2004; Deshmukh and Vlachos 2007) (Fig. 10).
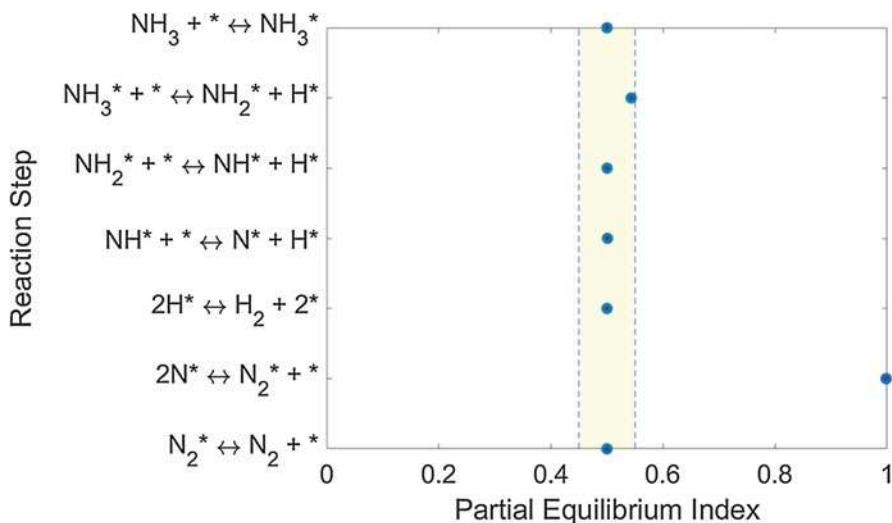
Fig. 10 Steady-state partial equilibrium index for ammonia decomposition at 750 K

12.4 Sensitivity Analysis

A sensitivity analysis reveals the RDS by intentionally varying the rate constant of each reaction by a differential factor and determining its impact on the overall reaction rate (e.g., represented by conversion). This is accomplished by multiplying both the forward and backward pre-exponential factors of each reaction step by the same small factor. The normalized sensitivity coefficient (NSC) is calculated from

$$NSC = \frac{d\,(\ln X)}{d\,(\ln A)} \approx \frac{\Delta\,(\ln X)}{\Delta\,(\ln A)} \tag{48}$$

where $d(\ln A)$ is a differential change in the pre-exponential factor of a specific reaction step and $d(\ln X)$ is the differential response of the system, in this case, the conversion; other quantities of interest can also be used.

Figure 11 shows that most reaction steps are kinetically irrelevant. The nitrogen association reaction is the most sensitive at all temperatures, while the first dehydrogenation reaction is relevant at temperatures above 750 K. At low temperatures, NH* dehydrogenation is significant too. The results are consistent with the partial equilibrium analysis. It is important to note that the RDS is not a single elementary reaction, but often a pair of reactions and the significance of these reactions can change with operating conditions, e.g., temperature, and the catalyst. MKM provides a systematic way of determining the RDS. Given that H abstraction can be kinetically significant, isotopically labeled ammonia can be used at various conditions to confirm model predictions. A kinetic isotope effect is expected at low and high temperatures but not one at 700–800 K. These model predictions can be assessed with experiments.

Fig. 11 Normalized sensitivity coefficients for ammonia decomposition on Ru catalyst over a range of temperatures
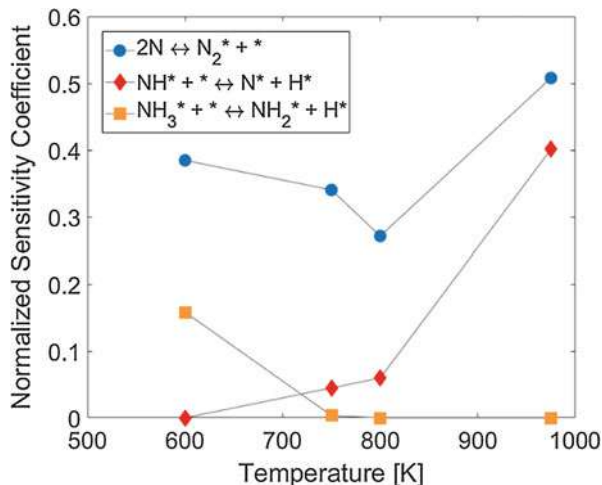
12.5   Potential Energy Diagram

The potential energy diagram is often presented at the zero-coverage limit, and the step with the highest barrier is typically assumed to be RDS. This oversimplification can lead to incorrect conclusions as the rate of the reaction depends on the coverage of species participating in a reaction and these coverages are often the result of local equilibrium of partially equilibrated reactions. MKM can provide coverages under reaction conditions (similar to "operando" spectroscopy) and correct the potential energy diagram.

The potential energy diagram for the ammonia decomposition mechanism on a clean catalyst surface and at steady-state coverages shows the significant impact of species coverages on the energetics for this reaction (orange and blue lines in Fig. 12, respectively). Specifically, the reaction barrier for the nitrogen association is significantly increased to 1.8 eV due to repulsive interactions and becomes the dominant reaction barrier to overcome at typical surface coverages. In contrast, the zero-coverage potential energy diagram shows the last dehydrogenation and 2N* association having comparable barriers. Furthermore, energetics alone cannot explain why the NH_3* and NH* dehydrogenation steps can be kinetically significant given the relatively low barrier of these reactions.

13   Concluding Remarks

Microkinetic models are an essential tool for understanding a reaction mechanism and an important link to reactor modeling and design as well to catalyst prediction. Advances in computing power and DFT functionals have both improved the speed and accuracy of producing first-principles based data, which are essential for the estimation of key kinetic and thermodynamic parameters required for a microkinetic
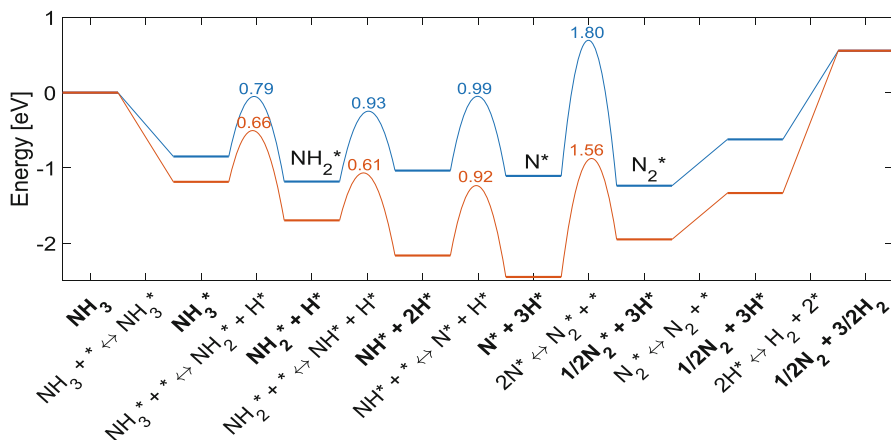
Fig. 12 Potential energy diagram with activation barriers for ammonia decomposition at 750 K at zero-coverage (orange) and at a typical coverage under reaction conditions (blue)

model. Alternatively, BEPs and LSRs can be used to estimate the parameters of large mechanisms. These fast estimation methods can be combined with first-principles calculations to enable building large models at low cost while ensuring DFT accuracy of key thermodynamic and kinetic parameters. This chapter has provided the approach, theory, equations, and examples necessary to use these tools. An example of a microkinetic software is given on Github at https://github.com/VlachosGroup/NH3-Matlab-Microkinetic-Model.

References

Abild-Pedersen F, Greeley J, Studt F et al (2007) Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. Phys Rev Lett 99:4–7. https://doi.org/10.1103/PhysRevLett.99.016105

Aghalayam P, Park YK, Vlachos DG (2000) Construction and optimization of complex surface-reaction mechanisms. AICHE J 46:2017–2029. https://doi.org/10.1002/aic.690461013

Coltrin ME, Kee RJ, Rupley FM (1996) Surface Chemkin-III: a Fortran package for analyzing heterogeneous chemical kinetics at a solid-surface – gas-phase interface. Sandia National Laboratory, Albuquerque, NM

Cornish-Bowden A (2015) One hundred years of Michaelis–Menten kinetics. Perspect Sci 4:3–9. https://doi.org/10.1016/j.pisc.2014.12.002

Deshmukh SR, Vlachos DG (2007) A reduced mechanism for methane and one-step rate expressions for fuel-lean catalytic combustion of small alkanes on noble metals. Combust Flame 149:366–383. https://doi.org/10.1016/j.combustflame.2007.02.006

Dumesic JA, Rudd DF (1993) The microkinetics of heterogeneous catalysis. American Chemical Society, Washington, DC

Falsig H, Shen J, Khan TS et al (2014) On the structure sensitivity of direct NO decomposition over low-index transition metal facets. Top Catal 57:80–88. https://doi.org/10.1007/s11244-013-0164-5

Fernández EM, Moses PG, Toftelund A et al (2008) Scaling relationships for adsorption energies on transition metal oxide, sulfide, and nitride surfaces. Angew Chem Int Ed 47:4683–4686. https://doi.org/10.1002/anie.200705739

Guo W, Vlachos DG (2013) Effect of local metal microstructure on adsorption on bimetallic surfaces: atomic nitrogen on Ni/Pt (111). J Chem Phys 174702:2–4. https://doi.org/10.1063/1.4803128

Guo W, Vlachos DG (2015a) Patched bimetallic surfaces are active catalysts for ammonia decomposition. Nat Commun 6:1–7. https://doi.org/10.1038/ncomms9619

Guo W, Vlachos DG (2015b) Patched bimetallic surfaces are active catalysts for ammonia decomposition: supplementary info. Nat Commun 6:1–7

Hill TL (1962) An introduction to statistical thermodynamics. Dover Publications, New York

Jones G, Bligaard T, Abild-Pedersen F, Nørskov JK (2008) Using scaling relations to understand trends in the catalytic activity of transition metals. J Phys Condens Matter 20:064239. https://doi.org/10.1088/0953-8984/20/6/064239

Laidler K, King M (1983) The development of transition-state theory. J Phys Chem 87:2657–2664

Lund EW (1965) Guldberg and Waage and the law of mass action. J Chem Educ 42:548. https://doi.org/10.1021/ed042p548

Mhadeshwar AB, Vlachos DG (2004) Microkinetic modeling for water-promoted CO oxidation, water-gas shift, and preferential oxidation of CO on Pt. J Phys Chem B 108:15246–15258. https://doi.org/10.1021/jp048698g

Mhadeshwar AB, Wang H, Vlachos DG (2003) Thermodynamic consistency in microkinetic development of surface reaction mechanisms. J Phys Chem B 107:12721–12733. https://doi.org/10.1021/jp034954y

Mhadeshwar AB, Kitchin JR, Barteau MA, Vlachos DG (2004) The role of adsorbate – adsorbate interactions in the rate controlling step and the most abundant reaction intermediate of NH_3 decomposition on Ru. Catal Lett 96:13–22. https://doi.org/10.1023/B:CATL.0000029523.22277.e1

Motz H, Wise H (1960) Diffusion and heterogeneous reaction. III. Atom recombination at a catalytic boundary. J Chem Phys 32:1893–1894. https://doi.org/10.1063/1.1731060

Nicolaides A, Rauk A, Glukhovtsev MN, Radom L (1996) Heats of formation from G2, G2(MP2), and G2(MP2,SVP) total energies. J Phys Chem 100:17460–17464. https://doi.org/10.1021/jp9613753

Nørskov JK (1993) Adsorbate-adsorbate interactions on metal surfaces, Chapter 1. In: King DA, Woodruff DP (eds) Coadsorption, promoters and poisons. Elsevier, Amsterdam, pp 1–27

Park YK, Aghalayam P, Vlachos DG (1999) A generalized approach for predicting coverage-dependent reaction parameters of complex surface reactions: application to H_2 oxidation over platinum. J Phys Chem A 103:8101–8107. https://doi.org/10.1021/jp9916485

Raimondeau S, Vlachos DG (2003) Front propagation at low temperatures and multiscale modeling for the catalytic combustion of H_2 on Pt. Chem Eng Sci 58:657–663. https://doi.org/10.1016/S0009-2509(02)00592-4

Reuter K, Scheffler M (2001) Composition, structure, and stability of RuO_2(110) as a function of oxygen pressure. Phys Rev B 65:35406. https://doi.org/10.1103/PhysRevB.65.035406

Salciccioli M, Stamatakis M, Caratzoulas S, Vlachos DG (2011) A review of multiscale modeling of metal-catalyzed reactions: mechanism development for complexity and emergent behavior. Chem Eng Sci 66:4319–4355. https://doi.org/10.1016/j.ces.2011.05.050

Samant A, Vlachos DG (2005) Overcoming stiffness in stochastic simulation stemming from partial equilibrium: a multiscale Monte Carlo algorithm. J Chem Phys 123:144114. https://doi.org/10.1063/1.2046628

Sandler SI (2010) An introduction to applied statistical thermodynamics, 1st edn. John Wiley & Sons, Hoboken, NJ, USA

Shustorovich E (1986) Chemisorption phenomena: analytic modeling based on perturbation theory and bond-order conservation. Surf Sci Rep 6:1–63. https://doi.org/10.1016/0167-5729(86)90003-8

Shustorovich E (1990) The bond-order conservation approach to chemisorption and hetero-geneous catalysis: applications and implications. Adv Catal. https://doi.org/10.1016/S0360-0564(08)60364-8

Stamatakis M, Vlachos DG (2012) Unraveling the complexity of catalytic reactions via kinetic Monte Carlo simulation: current status and frontiers. ACS Catal 2:2648–2663. https://doi.org/10.1021/cs3005709

Stamatakis M, Chen Y, Vlachos DG (2011) First-principles-based kinetic Monte Carlo simulation of the structure sensitivity of the water-gas shift reaction on platinum surfaces. J Phys Chem C 115:24750–24762. https://doi.org/10.1021/jp2071869

Sutton JE, Vlachos DG (2012) A theoretical and computational analysis of linear free energy relations for the estimation of activation energies. ACS Catal 2:1624–1634. https://doi.org/10.1021/cs3003269

Sutton JE, Vlachos DG (2015) Building large microkinetic models with first-principles' accuracy at reduced computational cost. Chem Eng Sci 121:190–199. https://doi.org/10.1016/j.ces.2014.09.011

van't Hoff JH (1884) Études de dynamique chimique. Frederik Muller, Amsterdam. Google Books

Vojvodic A, James A, Studt F et al (2014) Exploring the limits: a low-pressure, low-temperature Haber – Bosch process: supplement. Chem Phys Lett 598:108–112. https://doi.org/10.1016/j.cplett.2014.03.003

Wijaya CD, Sumathi R, Green WH (2003) Thermodynamic properties and kinetic parameters for cyclic ether formation from hydroperoxyalkyl radicals. J Phys Chem A 107:4908–4920. https://doi.org/10.1021/jp027471n

Computational Fluid Dynamics of Catalytic Reactors

58

Matthias Hettel, Martin Wörner, and Olaf Deutschmann

Contents

M. Hettel
Institute for Chemical Technology and Polymer Chemistry, Karlsruhe Institute of Technology
(KIT), Karlsruhe, Germany
e-mail: matthias.hettel@kit.edu

M. Wörner
Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology (KIT),
Karlsruhe, Germany
e-mail: martin.woerner@kit.edu

O. Deutschmann (✉)
Institute for Chemical Technology and Polymer Chemistry and Institute of Catalysis Research
and Technology, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: deutschmann@kit.edu

Abstract

Today, the challenge in chemical and material synthesis is not only the development of new catalysts and supports to synthesize a desired product, but also the understanding of the interaction of the catalyst with the surrounding flow field. Computational Fluid Dynamics (CFD) is the analysis of fluid flow, heat and mass transfer, and chemical reactions by means of computer-based numerical simulations. CFD has matured into a powerful tool with a wide range of applications in industry and academia. From a reaction engineering perspective, main advantages are reduction of time and costs for reactor design and optimization, and the ability to study systems where experiments can hardly be performed, e.g., hazardous conditions or beyond normal operation limits. However, the simulation results will always remain a reflection of the uncertainty in the underlying models and physicochemical parameters so that in general a careful experimental validation is required.

This chapter introduces the application of CFD simulations in heterogeneous catalysis. Catalytic reactors can be classified by the geometrical design of the catalyst material (e.g. monoliths, particles, pellets, washcoats). Approaches for modeling and numerical simulation of the various catalyst types are presented. Focus is put on the principal concepts for coupling the physical and chemical processes on different levels of details, and on illustrative applications. Models for surface reaction kinetics and turbulence are described and an overview on available numerical methods and computational tools is provided.

1 Introduction

Today, the challenge in chemical and material synthesis is not only the development of new catalysts and supports to synthesize a desired product but also the understanding of the interaction of the catalyst with the surrounding reactive flow field. Often, only the exploitation of these interactions can lead to the desired product selectivity and yield. Hence, a better understanding of gas-solid, liquid-solid, and gas-liquid-solid flows in chemical reactors is understood as a critical need in chemical technology calling for the development of reliable simulation tools that integrate detailed models of reaction chemistry and computational fluid dynamics (CFD) modeling of macroscale flow structures.

The ultimate objective of CFD simulations of catalytic reactors is (1) to understand the interactions of physics (mass and heat transport) and chemistry in the reactor, (2) to support reactor design and engineering, and, (3) eventually, to find optimized operating conditions for the maximization of the yield of the desired product and minimization of undesired side products or pollutants.

This chapter introduces the application of CFD simulations to obtain a better understanding of the interactions between mass and heat transport and chemical reactions in reactors for chemical and materials synthesis, in particular in the presence of catalytic surfaces. Catalytic reactors are generally characterized by the complex interaction of various physical and chemical processes. Bundles or tubular fixed bed reactors can serve as an example. They are widespread in chemical industry and in particular used for reactions with large heat release or supply. Figure 1 illustrates typical time scales and length scales within such a catalytic tube bundle reactor.

Computational fluid dynamics (CFD) is able to predict complex flow fields, even combined with heat transport, due to the recently developed numerical algorithms and the availability of fast computer hardware. The consideration of detailed models for chemical reactions, in particular for heterogeneous reactions, however, is still very challenging due to the large number of species mass conservation equations,
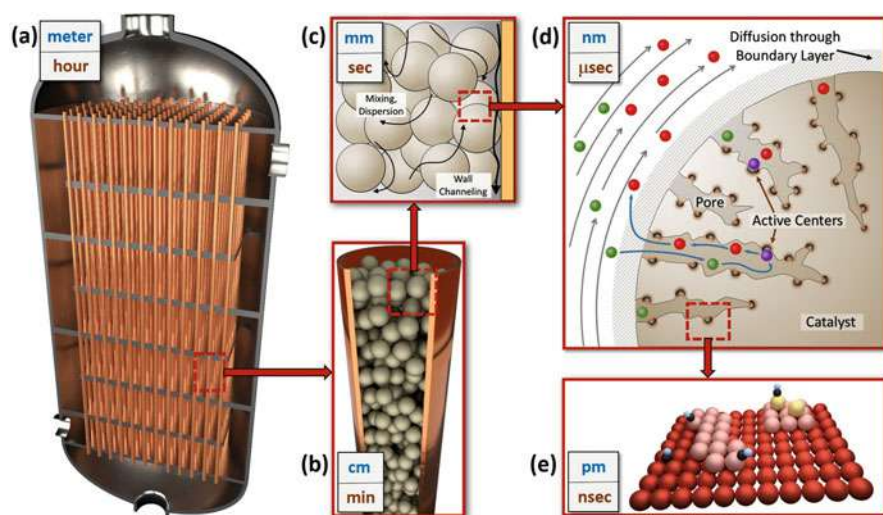
Fig. 1 Typical length scales (blue text) and length time scales (brown text) in a catalytic bundled tube reactor. (**a**) Principle of the reactor. The mixture flows from top to bottom through a bundle of tubes where the conversion takes place. The tubes are bathed by a fluid (gaseous or liquid) which is guided around the tubes in a crossflow arrangement. This fluid serves as a cooling or heating agent depending on the type of reaction (exothermic or endothermic). The tubes (**b**) are filled with catalytic particles. In each tube, the transport of momentum, energy, and chemical species occurs not only in flow direction but also into radial direction. Channeling in the near-wall region (**c**) affects the radial heat transfer; radiation plays a significant role at higher temperatures; local kinetics influence local transport phenomena and vice versa; reaction rates are limited due to film diffusion. The catalyst material is often dispersed in porous structures like washcoats or pellets (**d**). Mass transport in the fluid phase and chemical reactions are superimposed by diffusion of the species to the active catalytic centers in the pores. The temperature distribution depends on the interaction of heat convection and conduction in the fluid, heat release due to chemical reactions, heat transport in the solid material, and thermal radiation. The chemical reaction itself takes place at active sites on the inner surface of the porous catalyst structure (**e**)

their highly nonlinear coupling, and the wide range of time scales introduced by the complex reaction networks.

In the following, concepts for modeling and numerical simulation of catalytic reactors are presented, which describe the coupling of the physical and chemical processes in detail. The elementary kinetics and dynamics as well as ways for modeling the intrinsic chemical reaction rates (micro kinetics) by the mean field approximation (MF) and by lumped kinetics are discussed. It is assumed that models exist that can compute the local heterogeneous, but also homogeneous reaction rate, as a function of the local conditions such as temperature and species concentration in the gas phase and of the local and temporal state of the surfaces. These chemical source terms are here coupled with the fluid flow and used to numerically simulate the chemical reactor.

The focus is on the principal ideas and the potential applications of CFD in heterogeneous catalysis; textbooks and specific literature are frequently referenced for more details. Specific examples taken from literature in combination with own work will be used for illustration of the state-of-the-art CFD simulation of chemical reactors with heterogeneously catalyzed reactions.

2 Governing Equations of Multicomponent Single-Phase Flows

As long as a fluid can be treated as a continuum, the most accurate description of the flow field of multicomponent mixtures is given by the transient three-dimensional (3D) Navier-Stokes equations coupled with the mass continuity, energy, and species transport equations, which will be summarized in this section. More detailed introductions into fluid dynamics and transport phenomena can be found in a number of textbooks (Bird et al. 2001, Kee et al. 2003, Warnatz et al. 1996, Hayes and Kolaczkowski 1997). Other alternative concepts such as the Lattice-Boltzmann method are not discussed here, because the state of development does not allow a broad application.

Governing equations, which are based on conservation principles, can be derived by consideration of the fluid flow for a differential control volume. The principle of mass conservation leads to the mass continuity equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho \cdot u_i)}{\partial x_i} = 0, \tag{1}$$

with ρ being the mass density, t the time, x_i ($i = 1,2,3$) are the Cartesian coordinates, and u_i the velocity components. The principle of momentum conservation for Newtonian fluids leads to three scalar equations (Navier-Stokes Equations) for the momentum components ($\rho \cdot u_i$):

$$\frac{\partial (\rho \cdot u_i)}{\partial t} + \frac{\partial \left(\rho \cdot u_i \cdot u_j\right)}{\partial x_j} + \frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} = \rho \cdot g_i, \tag{2}$$

where p is the static pressure, τ_{ij} is the stress tensor, and g_i are the components of the gravitational acceleration. The above equation is written for Cartesian coordinates. The stress tensor is given as

$$
\tau_{ij} = -\mu \cdot \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \left(\frac{2}{3}\mu - \kappa \right) \delta_{ij} \cdot \frac{\partial u_k}{\partial x_k}.
\tag{3}
$$

Here, κ and μ are the bulk viscosity and mixture viscosity, respectively, and δ_{ij} is the Kronecker delta, which is unity for $i = j$, else zero. The bulk viscosity vanishes for low-density mono-atomic gases and is also commonly neglected for dense gases and liquids. The coupled mass continuity and momentum equations have to be solved for the description of the flow field.

In multicomponent mixtures, not only the flow field is of interest but also mixing of the chemical species and reactions among them, which can be described by an additional set of partial differential equations. Here, the mass m_i of each of the N_g gas-phase species obeys a conservation law that leads to

$$
\frac{\partial (\rho \cdot Y_i)}{\partial t} + \frac{\partial (\rho \cdot u_j \cdot Y_i)}{\partial x_j} + \frac{\partial (j_{i,j})}{\partial x_j} = R_i^{\text{hom}},
\tag{4}
$$

where Y_i is the mass fraction of species i in the mixture ($Y_i = m_i/m$, with m as total mass) and R_i^{hom} is the net rate of production due to homogeneous chemical reactions. The components $j_{i,j}$ of the diffusion mass flux caused by concentration and temperature gradients are often modeled by the mixture-average formulation:

$$
j_{i,j} = -\rho \cdot \frac{Y_i}{X_i} \cdot D_i^{\text{M}} \cdot \frac{\partial X_i}{\partial x_j} - \frac{D_i^{\text{T}}}{T} \cdot \frac{\partial T}{\partial x_j}.
\tag{5}
$$

D_i^{M} is the effective diffusion coefficient of species i in the mixture; D_i^{T} is the thermal diffusion coefficient, which is significant only for light species; and T is the temperature. The molar fraction X_i is related to the mass fraction Y_i using the species molar masses M_i by

$$
X_i = \frac{1}{\sum_{j=1}^{N_g} Y_j/M_j} \cdot \frac{Y_i}{M_i}.
\tag{6}
$$

Heat transport and heat release due to chemical reactions lead to spatial and temporal temperature distributions in catalytic reactors. The corresponding governing equation for energy conservation is commonly expressed in terms of the specific enthalpy h:

$$
\frac{\partial (\rho \cdot h)}{\partial t} + \frac{\partial (\rho \cdot u_j \cdot h)}{\partial x_j} + \frac{\partial j_{q,j}}{\partial x_j} = \frac{\partial p}{\partial t} + v_j \cdot \frac{\partial p}{\partial x_j} - \tau_{jk} \cdot \frac{\partial v_j}{\partial x_k} + S_{\text{h}},
\tag{7}
$$

with S_h being the heat source, for instance, due to chemical reactions. In multicomponent mixtures, diffusive heat transport is significant due to heat conduction and mass diffusion, hence

$$j_{q,j} = -\lambda \cdot \frac{\partial T}{\partial x_j} + \sum_{i=1}^{N_g} h_i \cdot j_{i,j}. \tag{8}$$

λ is the thermal conductivity of the mixture. The temperature is then related to the enthalpy by the definition of the mixture specific enthalpy:

$$h = \sum_{i=1}^{N_g} Y_i \cdot h_i(T), \tag{9}$$

with h_i being the specific enthalpy of species i, which is a monotonic increasing function of temperature. The temperature is then commonly derived from Eq. (9) for known h and Y_i.

Heat transport in solids such as reactor walls and catalyst materials can also be modeled by an enthalpy equation

$$\frac{\partial (\rho \cdot h)}{\partial t} - \frac{\partial}{\partial x_j} \left(\lambda_s \cdot \frac{\partial T}{\partial x_j} \right) = S_h, \tag{10}$$

where h is the specific enthalpy and λ_s the thermal conductivity of the solid material. S_h accounts for heat sources, for instance, due to electric or radiative heating of the solid.

This system of governing equations is closed by the equation of state to relate the thermodynamic variables density ρ, pressure p, and temperature T. The simplest model of this relation for gaseous flows is the ideal gas equation:

$$p = \frac{\rho \cdot R \cdot T}{\sum\limits_{i=1}^{N_g} X_i \cdot M_i}, \tag{11}$$

with the universal gas constant $R = 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$.

The transport coefficients of the fluid μ, D_i^M, D_i^T, and λ appearing in Eqs. (3, 5, and 8) depend on temperature and mixture composition. They are derived from the transport coefficients of the individual species and the mixture composition by applying empirical approximations.

The specific enthalpy h_i is a function of temperature and can be expressed in terms of the heat capacity:

$$h_i = h_i\,(T_{\text{ref}}) + \int\limits_{T_{\text{ref}}}^{T} c_{p,i}(T)\;\mathrm{d}T, \tag{12}$$

where $h_i\,(T_{\text{ref}})$ is the enthalpy at reference conditions ($T_{\text{ref}} = 298.15$ K, $p_0 = 1$ bar) and $c_{p,i}$ is the specific heat capacity at constant pressure.

3   Coupling of the Flow Field with Heterogeneous Chemical Reactions

Depending on the spatial resolution of the different catalyst structures, e.g. flat surface, gauzes, and pellets, embedded in porous media, the species mass fluxes due to catalytic reactions at these structures are differently coupled with the flow field.

3.1   Modeling of Fluid above Nonporous Catalytic Surfaces

In the simplest case considered, the catalytic surface consists of a single nonporous homogeneous material, e.g. wires or plates made of platinum. The chemical processes at the surface are then coupled with the surrounding flow field by boundary conditions for the species continuity equations, Eq. (4), at the gas-solid interface:

$$\vec{n} \cdot \left(\vec{j}_i + \rho \cdot \vec{u}_{\text{Stef}} \cdot Y_i \right) = R_i^{\text{het}}. \tag{13}$$

Here, $\vec{n}$ is the outward-pointing unit vector normal to the surface, $\vec{j}_i$ is the diffusion mass flux of species i as discussed in Eq. (4), and R_i^{het} is the heterogeneous surface reaction rate, which is given per unit geometric surface area (kg m^{-2} s^{-1}). It should be noted that in this case the *catalytic surface* (red outline in Fig. 2, left)
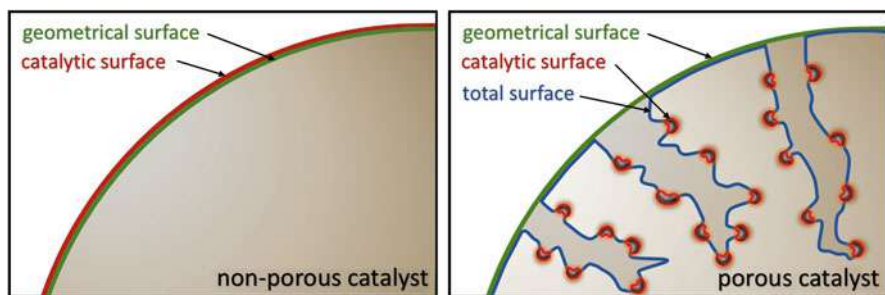
Fig. 2 Left: Definition of surface areas for a nonporous catalyst. Right: Definition of surface areas for porous catalysts (or porous layers such as washcoats)

corresponds to the *geometrical surface* (green outline in Fig. 2, left) of the fluid-solid interphase of the flow field simulation.

The Stefan velocity $\vec{u}_{\text{Stef}}$ occurs at the surface if there is a net mass flux between the surface and the gas phase:

$$\vec{n} \cdot \vec{u}_{\text{Stef}} = \frac{1}{\rho} \cdot \sum_{i=1}^{N_g} R_i^{\text{het}}. \tag{14}$$

At steady-state conditions, this mass flux vanishes unless mass is deposited on the surface, e.g. chemical vapor deposition, or ablated, e.g. material etching. Equation (13) basically means that for $\vec{u}_{\text{Stef}} = 0$, the amount of gas-phase molecules of species i, which are consumed/produced at the catalyst by adsorption/desorption, have to diffuse to/from the catalytic wall (Eq. 5).

3.2   Fluid Flow above Porous Catalysts

Most catalysts exhibit a certain structure, for instance, they may occur as dispersed particles on a flat or in a porous substrate. Examples are thin catalytically coated walls in honeycomb structures, foams, disks, plates, and well-defined porous media (e.g. particles). A representative is a washcoat (Fig. 3), which is a thin layer of supporting material where small particles of the catalytic active material (e.g. precious metals) are embedded in a supporting material (Fig. 2, right). The numerical grid resolves only the flow region bounded by the geometrical structure of the catalyst. The simplest way to account for the active catalytic surface area consists in scaling the intrinsic reaction of Eq. (14) in the following form:

$$R_i^{\text{het}} = \eta \cdot F_{\text{cat/geo}} \cdot M_i \cdot \dot{s}_i. \tag{15}$$

Here, $\dot{s}_i$ is the molar net production rate of gas-phase species i, given in mol m^{-2} s^{-1}. The area now refers to the actual *catalytically active surface area* (red outline in Fig. 2, right). The parameter $F_{\text{cat/geo}}$ represents the amount of *catalytically*
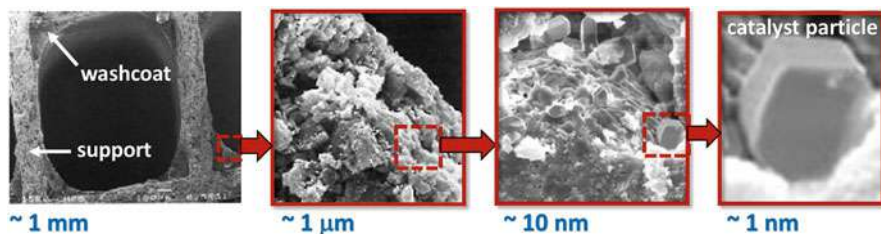
Fig. 3 Washcoat on a channel of a monolith with typical length scales

active surface area in relation to the *geometric surface area* of the fluid-solid interphase (green outline in Fig. 2, right). Depending on porosity of the support and the dispersion (e.g. of particles of precious metals), this factor can range up to the order of hundred. The *catalytically active surface area* is the surface area (layer on which we find adsorbed species) of the catalytic active particles exposed to the ambient gas (fluid) phase. This area can be determined experimentally (e.g. by chemisorption) with sample molecules such as CO and hydrogen. The *catalytically active surface area* should not be confused with the BET (Brunauer-Emmett-Teller) surface area representing the *total inner surface area* of a porous structure (blue outline in Fig. 2, right).

The simplest model to include the effect of internal mass transfer resistance for catalyst dispersed in a porous media is the effectiveness factor η based on the Thiele modulus (Hayes and Kolaczkowski 1997). The effectiveness factor of species i, η_i, is defined as

$$\eta_i = \frac{\dot{s}_{i,\text{mean}}}{\dot{s}_i} \tag{16}$$

with $\dot{s}_{i,\text{mean}}$ as mean surface reaction rate in the porous structure. Assuming a homogeneous porous medium, time-independent concentration profiles, and a rate law of first order, the effectiveness factor can be analytically calculated in terms of

$$\eta_i = \frac{\tanh(\Phi_i)}{\Phi_i} \tag{17}$$

with Φ_i as Thiele module defined as

$$\Phi_i = L \cdot \sqrt{\frac{\dot{s}_i \cdot \gamma}{D_{\text{eff},i} \cdot c_{i,0}}}. \tag{18}$$

Here, L is the thickness of the porous medium (washcoat), γ is the ratio of catalytic active surface area to washcoat volume, and $c_{i,0}$ are the species concentrations at the fluid/porous media interface (geometrical surface in Fig. 2, right). The Thiele module is a dimensionless number. The value in the root term of Eq. (18) represents the ratio of intrinsic reaction rate to diffusive mass transport in the porous structure. Since mass conservation has to be obeyed (Eq. 15), the same effectiveness factor has to be applied for all chemical species. Therefore, this simple model can only be applied at conditions at which the reaction rate of one species determines overall reactivity. Furthermore, this model then implies that mass diffusion inside the porous media can be described by the same diffusion coefficient for all species.

In fixed bed reactors with large numbers of catalytic pellets or for catalyst dispersed in porous media, the structure of the catalyst cannot be resolved geometrically. In those cases, the catalytic reaction rate is expressed per volume, which means R_i^{het} is now given in kg m^{-3} s^{-1}; the volume here refers to the volume of

a computational cell in the geometrical domain of a discretized flow region. Then R_i^{het} simply represents an additional source term on the right side of the species continuity equation, Eq. (4), and is computed by

$$R_i^{\text{het}} = \eta \cdot S_{\text{V}} \cdot M_i \cdot \dot{s}_i, \tag{19}$$

where S_{V} is the active catalytic surface area per volumetric unit, given in m^{-1}, determined experimentally or estimated. $F_{\text{cat/geo}}$ as well as S_{V} can be expressed as a function of the position inside a reactor and time to account for inhomogeneously distributed catalysts and loss of activity, respectively.

3.3   Porous Catalyst as a Homogeneous Media

The aforementioned approach fails at conditions under which the reaction rate and diffusion coefficient of more than a single species determine overall reactivity. Hence, the interaction of diffusion and reaction demands more adequate models if mass transport in the porous media is dominated rather by diffusion than by convection. Contrary to the approaches shown above, not only the fluid region has to be discretized but also the porous solid region that is considered as a homogeneous media.

Concentration gradients inside the porous media result in spatial variations of the surface reaction rates $\dot{s}_i$. In thin catalyst layers (washcoats), these are primarily significant in normal direction to the fluid/washcoat boundary. Therefore, one-dimensional reaction-diffusion equations are applied with their spatial coordinate in that direction. Each chemical species leads to one reaction-diffusion equation, which is written in steady state as

$$\frac{\partial}{\partial x} \left(-D_i^{\text{eff}} \cdot \frac{\partial c_i^{\text{W}}}{\partial x} \right) - S_{\text{V}} \cdot \dot{s}_i = 0. \tag{20}$$

Here, c_i^{W} denotes the species concentration inside the washcoat. D_i^{eff} is the effective diffusion coefficient, which can account for the different diffusion processes in macro- and micropores.

Eq. (20) is coupled with the surrounding flow field, Eq. (5), at the interface between open fluid and catalytic layer/pellet. In this model the species concentrations, catalytic reaction rates, and surface coverages do not only depend on the position of the catalytic layer/pellet in the reactor but also vary inside the catalyst layer/particle leading to CPU-time-consuming computations.

Fluxes within porous media which are driven by gradients in concentration and pressure, i.e. diffusion and convection, can be described by the dusty gas model (DGM) (Kerkhof and Geboers 2005). This model, which is also applicable for three-dimensional and larger porous media, leads to more sophisticated computational efforts.

3.4   Resolved Modeling of Porous Catalyst

If a catalytic layer is porous (e.g. washcoat in Fig. 3), the effective catalyst surface area is often significantly larger than the geometrical surface area. The multiplicative factor is $F_{\text{cat/geo}}$ (see above). Albeit with knowledge of measured properties such as porosity, BET surface area, and chemisorption characteristics, the effective catalyst area is usually adjusted empirically. Moreover, $F_{\text{cat/geo}}$ is often taken to be a constant, independent of changes in reactor operating conditions or age of the catalyst.

Karakaya et al. 2016 performed a detailed analysis of reaction-diffusion processes within a porous rhodium-alumina catalyst washcoat based on a reconstruction of the actual catalyst-support microstructure. In this case, the total surface of a washcoat was used as the catalytic surface area (blue outline in Fig. 2, right). The catalytic chemistry was represented simply as a first-order irreversible reaction A→B, with a rate constant of k. The molar net production rate of the gas-phase species B is therefore

$$R_B^{\text{het}} = F_{\text{total}} \cdot k \cdot c_A, \tag{21}$$

where F_{total} is the total surface of the solid (see Fig. 2, right). They showed that characteristics such as effective area and pore effectiveness are strong functions of the reaction chemistry and the pore microstructure. The results depend greatly on the Damköhler number, which defines the ratio of the diffusive time scale to the reaction time scale. At relatively high Damköhler numbers, the reaction is mainly concentrated at the pore openings leaving most of the catalytic area deeper in the pores underutilized. However, as the Damköhler number decreases, the reaction is able to penetrate further into the pores. In this case, the net reaction can depend on local diffusion resistances, such as small necking regions (see Figs. 4 and 5). In the work of Blasi et al. (2016), the results from the microscale reconstructed pore volumes from the work of Karakaya et al. (2016) were used to develop macroscale models that can be applied at larger length scales.

4   Numerical Methods and Computational Tools

There are a variety of methods to solve the coupled system of partial differential and algebraic equations (PDE), which were presented in the previous sections for modeling catalytic reactors. Very often, the transient three-dimensional governing equations are simplified (no time dependence, symmetry conditions, preferential flow direction, infinite diffusion etc.) as much as possible but still taking care of all significant processes in the reactor. Simplifications often are not straightforward and need to be conducted with care. Special algorithms were developed for special types of reactors to achieve a converged solution or to speed up the computation solution.
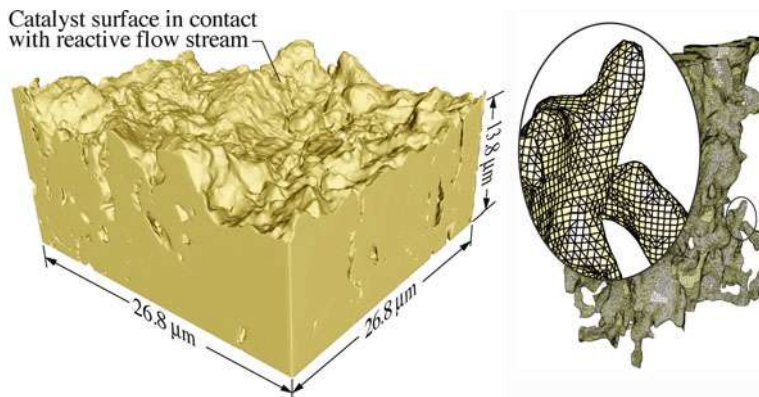
Fig. 4 Left: Three-dimensional reconstruction of a small section of catalyst washcoat. Right: Generated cut-cell mesh of a single pore. (Karakaya et al. 2016, reprinted with permission from Elsevier)
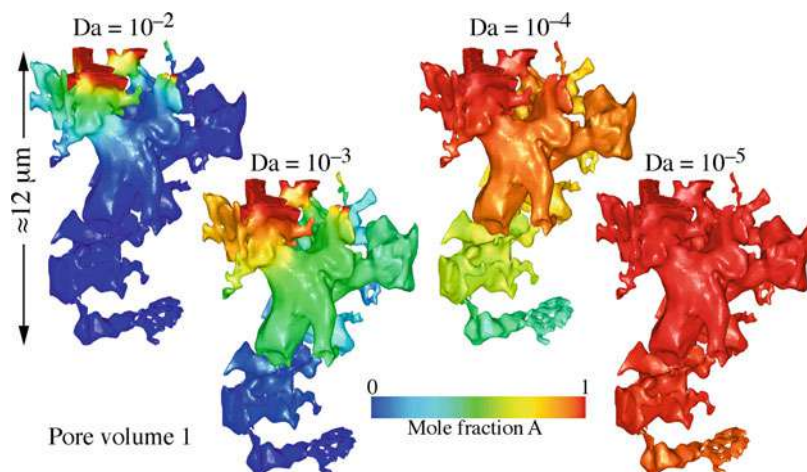
Fig. 5 Predicted reactant mole fractions within a pore at different Damköhler numbers. (Karakaya et al. 2016, reprinted with permission from Elsevier)

4.1   Numerical Methods for the Solution of the Governing Equations

An analytic solution of the PDE system is only possible in very limited special cases; for all practical cases, a numerical solution is needed. A numerical solution means that algebraic equations are derived approximating the solution of the PDE system at discrete points of the geometrical space of the reactor only at discrete points in time. The way of selection of these grid points and the derivation of algebraic

equations, which are finally solved by the computer, is called discretization. The grid points are representatives for small flow regions, which are now represented by the cells of a computational grid. Since the solution of the discretized equations is only an approximation of the solution of the PDE system, an error analysis (e.g. a study of grid independency) is an essential feature of the interpretation of every CFD simulation.

4.2   Turbulence Models

Turbulent flows are characterized by continuous fluctuations of velocity, which can lead to fluctuations in scalars such as density, temperature, and mixture composition. Turbulence can be desired in catalytic reactors to enhance mixing and reduce mass transfer limitations but is also unwanted due to the increased pressure drop and energy dissipation. An adequate understanding of all facets of turbulent flows is still missing.

Today, the engineer or scientist can choose one of the numerous commercial or public CFD codes including a wide variety of turbulence models. However, the complexity of the codes and the bandwidth of potential applications are enormous, and almost nobody is able to overlook all aspects of a modern CFD code. Even scientists often confine themselves to improve numerical algorithms or to develop models for the description of physical or chemical sub-processes and use the other parts of the codes in good faith of their accurateness. The idea that turbulence or turbulent reaction models may be used without knowledge of the underlying theory is misleading.

This section should give the reader an overview of numerical turbulence and reaction modeling. For a deeper insight, some basic additional literature should be mentioned: turbulence (Wilcox 1998; Pope 2000; Lemos 2006), numerical methods (Patankar 1990; Oran and Boris 1987; Fox 2003; Versteeg and Malalasekera 2007), and combustion modeling (Libby and Williams 1994; Peters 2000; Poinsot and Veynante 2001).

4.2.1   Length and Time Scales of Turbulent Reacting Flows

The length scales of turbulent flows range from the size of Kolmogorov eddies η up to the size L_t (integral length scale) of the large energy-containing eddies. In technical applications, L_t is typically one order smaller and η typically three orders smaller than the characteristic size of the flow system. Corresponding to the length scales – via a characteristic flow velocity – there exist time scales, accordingly. The time scales of the turbulent flow range from the lifetime of the fine-grained Kolmogorov eddies up to the characteristic lifetime τ_t of the large energy-containing eddies. On the chemical conversion side, the time scales of reactions range from very small time scales (formation of radicals) up to large time scales (e.g., formation of NO_X), and thus, many time scales exist, which are able to interact with those of turbulent mixture.

4.2.2   Calculation of the Flow Field
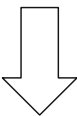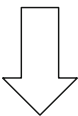
For the calculation of multidimensional turbulent flows, different methods with various levels of detail exist. The most common methods used in CFD are listed in Table 1. The more universal a method is, the larger is the computational effort. Besides that, there are numerous specialized methods which are only applicable to distinct flow problems. The acceptable level of complexity is restricted by the available computer capacity.

Direct numerical simulations (DNS) solve the time-dependent Navier-Stokes equations without using any turbulence model. For technical applications, there are at least three orders of magnitude between the size L of the flow system and the mesh size Δ needed to resolve the smallest scales. In 3D, this would result in at least 10^9 grid points. As the length scale of the small vortices decreases with the Reynolds number of the flow, the number of grid cells needed to resolve also the small structures increases. Therefore, DNS is still out of reach as a method to predict even isothermal turbulent flows for practical engineering applications (high Reynolds number) for many years to come. Therefore, this method is restricted to either flows with small Reynolds number (e.g., in the order of some thousands or laminar flows) or to higher Reynolds number but only small calculation domains (in the order of centimeters).

Consequently, turbulent flow configurations are generally modeled with other approaches, i.e., Reynolds-averaged Navier-Stokes simulations (RANS) or large eddy simulations (LES). Within these two basic approaches, the balance equations are filtered in time or space, respectively. All flow variables are subdivided into a resolved part and an unresolved part. Both approaches introduce additional terms, which have to be modeled to achieve closure, in the Navier-Stokes equations.

The balance equations for RANS techniques are obtained by time averaging the instantaneous balance equations. If the turbulent flow is steady state in the sense that there is no overall time scale of the transient process (statistically steady-state flow, e.g. no periodic variations), the calculation yields only the time-averaged quantities. The most popular model is the k-ε model, which solves two additional transport equations (partial differential equations) for the quantities k (turbulent kinetic energy) and ε (turbulent dissipation). These two variables are used to determine the turbulent viscosity $\mu_t \sim k^2/\varepsilon$ representing the effect of the turbulent fluctuations on the exchange of momentum in the flow. This quantity is dependent on the local conditions of turbulence and is no property of the medium. The RANS

Table 1 Methods for the calculation of fluid dynamics

Universalism	Method	Resolution, Effort
	1. Solving the Reynolds-averaged Navier-Stokes Equations (RANS-approach) 2. Solving the Space-filtered Navier-Stokes Equations (LES: Large-Eddy-Simulation) 3. Solving of the Navier-Stokes Equations (DNS: Direct-Numerical-Simulation)	

technique is still the standard approach used for the calculation of catalytic reactors with turbulent flow.

LES is always formulated in a transient way, which makes it a time-consuming calculation. In deriving the basic LES equations, the instantaneous balance equations are spatially filtered with a filter size, which usually is the size of the grid. The objective of LES is explicitly to compute the largest structures of the flow field (typically structures larger than the mesh size), whereas the effects that are dependent on the motion of the smallest eddies are modeled using subgrid closure rules (employing length-scale separation). The evolving local subgrid eddy viscosity μ_t has to be modeled.

Compared to RANS simulations, LES simulations are expensive in time and storage space. As modeling is only applied on the smaller unresolved scales, the results are usually less sensitive to modeling assumptions such as RANS models. However, usual simplifications, which can be applied by RANS as symmetry conditions or two-dimensional flows, cannot be retained. The solution has to be time-dependent, even for statistically steady-state flows.

Figure 6 shows schematically the temporal variation of the velocity u at a fixed spatial position in a statistically steady-state flow for different modeling approaches. The DNS claims to predict the same spatial and temporal variations inside the flow as a measurement would yield. The concept of LES implies that the determination of the flow will be the better the finer the grid is. Asymptotically, this leads to a DNS solution. In the frame of a RANS model, the velocity is constant and represents the time averaged value.

The fluid flow in small-scaled flow regions (in the order of cm and below) is highly affected by particle surfaces and reactor tube walls, respectively. Consequently, the turbulence intensity is also changed by the presence of a near wall. Getting closer to a wall, the tangential velocity fluctuations are diminished by viscous damping, and the normal fluctuations are damped by kinematic blocking. Therefore, it is important to use appropriate models for the flow laminarization at the solid surfaces. The traditional $k-\varepsilon$ model is not suited for near-wall velocity

Fig. 6 Temporal variation of velocity u at one position in a statistically steady-state flow for different modeling approaches
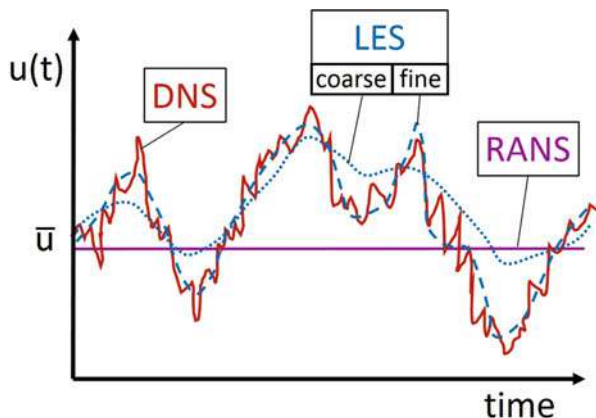
calculation, since the damping in this region is not accounted for. This and some other RANS models are based on wall functions, so that the sublayer and buffer layer do not have to be resolved. This is often enough if the flow region is large scaled with respect to the dimensions of the near wall regions. As a consequence, turbulence models that are based on the resolution of the near-wall zones have to be used for small-scaled (strongly confined by walls) flow regions. These are termed low-Reynolds number models such as the Spalart-Allmaras and $k - \omega$ model, where ω is a typical frequency of turbulence. Sometimes the flow region consists of distinct laminar and turbulent zones. An example is an automotive converter, where the flow inside the channels of the monolith is laminar whereas the flow upstream and downstream of the monolith is turbulent. RANS models cannot predict the laminarization in front of and the transition to turbulence behind the monolith.

Contrary, the LES approach is able to resolve the near-wall turbulence as well as laminarization and transition processes. Dixon et al. (2011) investigated different RANS turbulence models on a single sphere in terms of drag coefficient and heat transfer Nusselt number. As a reference case, large eddy simulations (LES) were carried out showing good agreement with correlations from literature. However, the small time step and fine mesh required are unpractical for entire bed simulations. The authors suggest using the shear-stress transport (SST) $k - \omega$ turbulence model, which is capable to predict heat transfer accurately. Cornejo et al. (2018) calculated the gas flow inside an entire automotive converter applying the porous body approach and using a RANS model. The decay of the turbulence inside of the single channels of the monolith was investigated using a LES model.

4.2.3 Turbulence and Chemistry

Turbulence implies a large range of both length scales and time scales. This is valid for both flow and chemical reactions. A basic concept for the development of turbulent reaction models is the physical principle stating that the probability of the interaction of scales decreases with the amount of scale separation. There are many circumstances where only limited ranges of chemical and turbulent time scales are involved and where the overlap of these ranges is small. According to Damköhler, the interaction of time scales can be expressed by the ratio of the typical times scales of chemical conversion τ_c and turbulence τ_t called the Damköhler number: $Da = \tau_c / \tau_t$.

Classical turbulent reaction models describing homogeneous reactions (e.g. combustion) explicitly or implicitly assume scale separation (fast chemistry). If the turbulent mixing is relatively fast or the chemistry is relatively slow (Damköhler number < 1), there is no time-scale separation. In this case, PDF approaches are claimed. Here, probability density functions (PDFs) (Warnatz et al. 1996), either derived by transport equations or empirically constructed, are used to consider the turbulent fluctuations when calculating the chemical source terms. In combination with RANS models for turbulence, these combustion models yield the time-averaged chemical reaction rates (Warnatz et al. 1996; Libby and Williams 1994). The realization of homogeneous reaction in the LES approach is handled in a similar

way for the subgrid scales. As the reaction is determined by molecular mixing, reaction takes place only on this dimension of scales.

There is a lack in literature with respect to the modeling of systems where turbulence and heterogeneous chemistry concur. To the best knowledge of the authors, no paper can be found where the impact of turbulence on the chemical conversion on a surface is discussed with respect to the interaction of time scales.

4.3   CFD Software

The three major approaches of discretization are the finite difference method (FDM), the finite volume method (FVM), and the finite element method (FEM). Today, nearly all commercial CFD codes are based on the FVM; only few use the FEM. Currently available multipurpose commercial CFD codes can simulate very complex flow configurations including turbulence and multicomponent transport. However, CFD codes still have difficulties to implement complex models for the chemical processes. One problem is the insufficient number of reactions and species the codes can handle. An area of recent development is the implementation of detailed models for heterogeneous reactions.

Several software packages have been developed for modeling complex reaction kinetics in CFD such as CHEMKIN, CANTERA, and DETCHEM, which also offer CFD codes for special reactor configurations such as channel flows and monolithic reactors. These kinetic packages and a variety of user written subroutines for modeling complex reaction kinetics have meanwhile been coupled to several commercial CFD codes. Aside from the commercially widespread multipurpose CFD software packages such as FLUENT, STAR-CD, FIRE, CFD-ACE+, and CFX, a variety of multipurpose and specialized CFD codes have been developed in academia and at research facilities. A widespread CFD code is the tool OpenFOAM (OpenFOAM 2017). By being open-source, OpenFOAM offers users complete freedom to customize and extend its existing functionality.

5   Reactor Simulations

From the modeling point of view, catalytic reactors can be classified by the geometrical design of the catalyst material. This is the main characteristic, because the numerical calculation needs a spatial discretization of the fluid region and of the solid material (surface or whole body). Therefore, it is essential how the catalyst is shaped and if it is immobile or not.

On the one hand, the catalyst can consist of one single continuous solid material that is penetrated by regularly distributed straight channels (as in monoliths or honeycomb structures) or by randomly distributed and shaped voids (as in foams). Here, the catalyst material is mostly deposited on the surface of the support material as a washcoat (see Fig. 3). This is a thin layer of supporting material where small particles of the catalytic active material (e.g. precious metals) are embedded.

On the other hand, the catalyst can consist of single particles with undefined form (e.g. crushed material) or pellets showing a defined geometrical shape (e.g., spheres or cylinders). The single elements itself can consist of a coated supporting material or can be built up of the catalyst material itself. In a packed bed, the particles/pellets are immobile. They touch each other in distinct contact points (for spheres) or contact lines/areas (for pellets) only. In a fluidized bed, the particles are moved by the flow and touch only frequently due to collisions.

This section explores the modeling of catalytic reactors. It starts with an approach where the solid structure is considered as a porous media. This model needs the smallest effort in computer capacities (memory and time) and can be applied for all of the configurations where the catalyst is immobile. The discretization of both the fluid and the solid structure is more complex and is called *resolved modeling* in the following. For each of the basic macroscopic morphologies of the catalyst (monolith, foam, fixed bed), some of the most used modeling methods will be presented consecutively. The simulation of mobile catalysts (e.g., fluidized beds) and reactors with multiple fluid phases demands special methods and is discussed afterward. Various special types of catalysts (e.g., wire gauzes) will not be discussed here.

For all the methods presented for the modeling of a reactor, it holds that the microscopic structure of the active part of the catalyst (washcoat or pelleted material) is not resolved by the numerical grid. This is because the morphology of the catalyst layer itself implies very small length scales. The thickness of washcoats is in the order of 0.1 mm (Fig. 3); the typical size of pores in this material is in the order of nm and below. However, the modeling of a small part of a reactor (e.g., one channel of a monolith) including the resolved washcoat can lead to very valuable insights in the interaction of processes (see Sect. 3.4).

5.1   Modeling of Immobile Catalyst as a Porous Media

A permeable structure can be seen as a multiphase system with one of the phases being a solid matrix. It is considered as a continuous media consisting partially of fluid and partially of solid (monolith, foam, fixed bed). Within this approach, the whole region of the porous body itself is discretized with a numerical grid. That means the real geometry of the flow region or the solid structure is not reproduced by the grid. The volumetric ratio between fluid and solid structure defines the porosity. Using volume averaging leads to a loss of information at the microlevel but provides insight into the macroscopic effects (e.g., flow maldistribution), which are needed for the full understanding of the system operation.

The flow of fluids through consolidated porous media and through beds of granular solids is similar, both having the general function of pressure drop versus flow rate. Transition from laminar to turbulent flow is gradual. Depending on the Reynolds number based on length scales of the macroscopic pores (e.g. diameter of channels in a monoliths/foams or particle diameter of a fixed bed), four distinct flow regimes can be observed (Dybbs and Edwards 1984): Darcy or creeping flow,

steady laminar inertial flow, unsteady laminar flow, and highly unsteady and chaotic flow regime. If the Reynolds number is very small, the empirical Darcy's law is applied traditionally for the determination of the pressure loss of the flow. For larger Reynolds number, the models of Ergun, Brinkman, and Forchheimer followed by numerous extended models were developed (Lemos 2006). The pressure loss may be strongly anisotropic, depending on the structure of the solid (e.g. for monoliths).

The properties of the porous medium (e.g. heat conductivity) which is a combination of the fluid and the solid part have to be deduced with additional models. Heat transport in porous media involves conductive heat transfer both in solid struts and in the fluid phase as well as convective heat transport due to the flow. The overall conductive heat transport depends on the material and structural properties of the solid and on the properties of the working fluid. Under the assumption of a local thermal equilibrium between fluid and solid phases, mixture models are used traditionally for heat conduction in porous media. By this, the temperatures of solids and fluids are assumed to be the same locally, and the heat conduction equations for the solid and fluid phase are lumped together. If the solid and fluid are in thermally non-equilibrium state, the heat conductions in the fluid and solid phases have to be considered with two separate balance equations. For heat exchange between fluid and solid structure, Nusselt number correlations are needed. If the temperature is high, the radiative heat transport has to be included.

The prediction of the mass transfer of species between flow and surface in the case of catalytic reactions demands additional models (Sherwood number correlations or effectiveness factor approaches). Within the porous media approach, the catalytic surface is parametrized as a specific area (square meter surface per volume porous body).

The modeling of a monolith as a continuous porous medium can be found in Hayes et al. (2012) and Porter et al. (2016). By pre-computing the required data at the micro- and mesoscales and storing them in a look-up table, small-scale effects (such as washcoat diffusion) can be captured in a macroscale model that executes with practicable execution times (Nien et al. 2013). A multiscale model for a fixed bed catalytic reactor is given in Park (2018). The macroscale part is for the transport across the reactor, and the microscale part is for the intra-particle transport and reaction. The effects of pellet size, intra-pellet diffusivity and conductivity on the reactor yield, and catalyst effectiveness factor are investigated.

5.2 Resolved Modeling of Monoliths (Honeycomb-Structured Catalysts)

In environmental catalysis (e.g., automotive catalytic converters, high-temperature catalysis, catalytic combustion and reforming), honeycomb-structured catalysts are often applied. Reactors that are based on monolithic structures include many equally sized small channels. To resolve these structures with CFD methods in detail, grids with high spatial resolution are needed, which leads to large numbers of grid cells and very large simulation times. To reduce the computer resources needed for

modeling of structured catalytic reactors, several strategies with various levels of complexity and accuracy have been developed.

One possibility is to model only one representative channel of the whole monolith in 1D/2D/3D with or without inclusion of diffusion effects in the catalytic coating (washcoat). This approach is successful if only one channel is a good representative for all other channels of the monolith. However, a typical monolith consists of thousands of channels running in parallel, and the operational conditions for each of them might vary significantly. If the conditions inside the monolith are inhomogeneous, it is far better to calculate a certain number of channels and assume that each of those channels is a representative for a group of them (Tischer and Deutschmann 2005; Sharma and Birgersson 2016; Hettel et al. 2015). Another approach is to exclude the channels of the catalyst from the CFD domain and to solve them using 1D/2D grids applying a second solver (Hettel et al. 2015; Sui et al. 2016).

Studies where all channels of a real converter are resolved by the numerical grid include restrictions in dimensionality, complexity of models, or in the number of channels considered. Either the number of channels resolved is large (some thousands) and there is no chemistry involved or the number of channels is small (some hundreds) and global chemistry or reduction in dimensionality is used.

Publications presenting the calculation of multichannel systems resolving the 3D geometry and using detailed chemistry are scarce in literature. Some papers focus on numerical methods and computer resources (Kumar and Mazumder 2010); others are using new computer architectures (Choudary and Mazumder 2014). Generally, there is a lack of publications where the results of numerical calculations are compared to spatially resolved experimental data from inside the channels. In the work of Hettel et al. (2018), the 3D modeling of a quarter of a CPOX (catalytic partial oxidation) reformer (diameter 20 mm) including two monoliths with ca. 300 channels each is discussed (Figs. 7 and 8). Detailed chemistry was applied, and radiation effects between the two solids were considered. The results are compared to detailed measured data.

5.3   Resolved Modeling of Foams

The application of ceramic foams as catalyst carriers represents a promising alternative to fixed beds or structured packings. Ceramic foams are characterized by their low specific pressure drop, high mechanical stability at relatively low specific weight, enhanced radial transport, as well as a high geometric surface area. Their open-cell structure consists of stiff, interconnected struts building a continuous network. In literature, these structures are typically named as foams or sponges. They can be found in a wide range of technical apparatuses, such as molten metal filters, burner enhancers, soot filters for diesel engine exhausts, and biomedical devices.

Several catalytic applications have been investigated during the last decades (combustion, exhaust gas treatment, syngas production). Maestri et al. (2005)
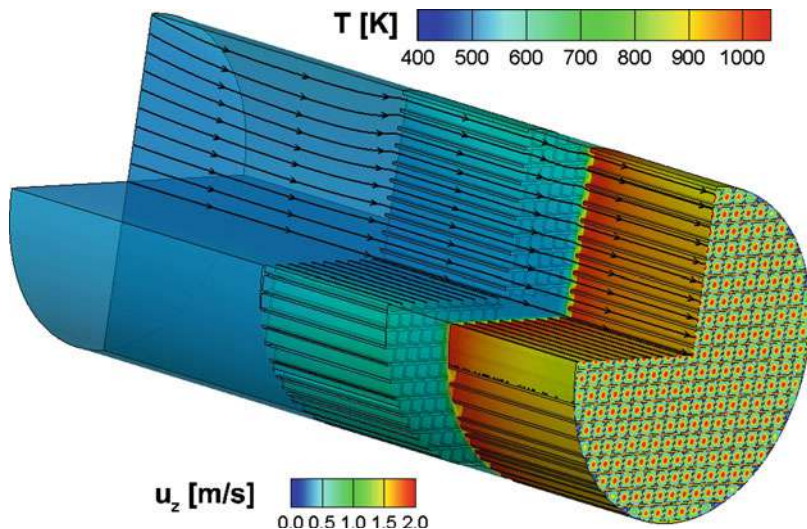
Fig. 7 Temperature field of the fluid part and the two solids (the left monolith is uncoated; the right one is coated). Additionally, some streaklines are plotted. Only the slice at the end of the calculation domain shows the distribution of the velocity into z-direction. (Hettel et al. 2018, reprinted with permission from Elsevier)

compared by simulation the performance of three different catalyst supports for the CPOX of methane, i.e., honeycomb, foam, and fixed bed reactor. The authors conclude that the foam catalyst support shows the best performance regarding transport properties, syngas selectivity, and pressure drop.

The sponge-like structure consists of solid edges and open faces through which fluid can pass. The irregularly shaped strut network can be expressed by its morphological parameters, i.e., cell and window diameter, strut thickness, and porosity. However, industrial manufacturers often use pores per inch (PPI) as a value of characterization, ranging typically between 10 and 100, and porosity values greater than 75%. One of the most import properties of ceramic foams is the specific surface area, which is significant for the transport of momentum, energy, and mass.

An alternative to transport correlation-based models (see sect. 5.1) are structure-resolved 3D CFD simulations. With such detailed simulations, quantitative descriptions can be made of transport phenomena within complex structures without a priori assumptions. Numerical 3D simulations of foam structures can be divided into real foam models and model structures of different complexity. The first type is carried out by converting the actual geometry by means of 3D computer tomography (CT) or 3D magnetic resonance imaging (MRI) into CAD data appropriate for 3D simulations. These models are limited to the scanned structure that is determined by its morphological characteristics. The procedures are time and cost intensive and include several consecutive steps of preparation.
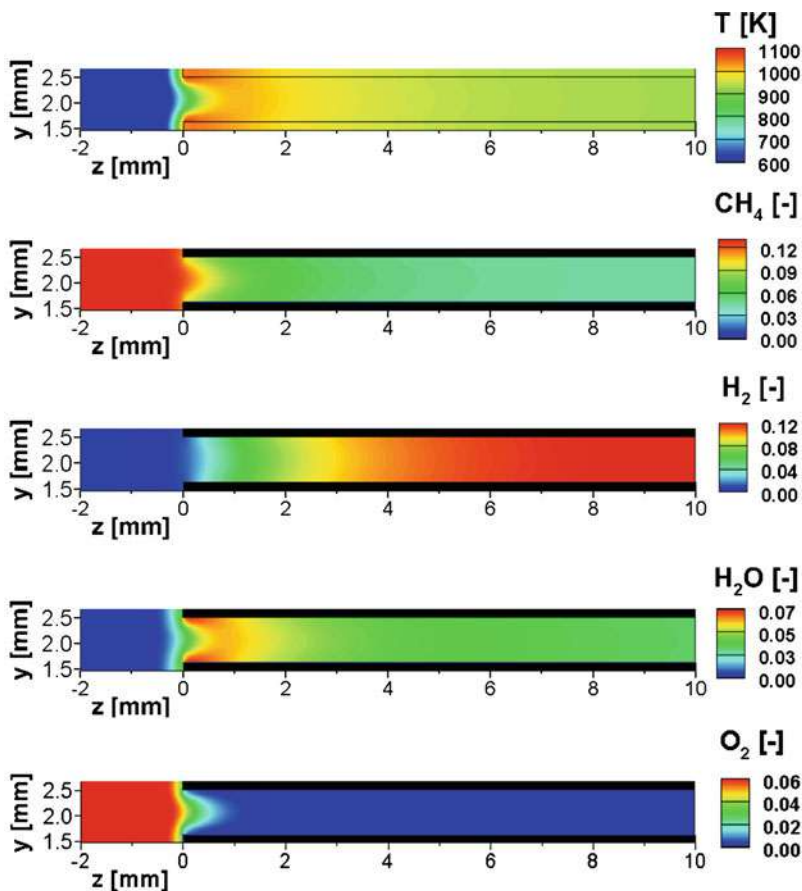
Fig. 8 Intersection through the midplane of only the central channel of the coated monolith. Distributions of temperature and species. (Hettel et al. 2018, reprinted with permission from Elsevier)

Model structures can be separated into deterministic models (e.g. unit cells) and stochastic models. An artificial foam modeler is able to generate structures that can be varied in terms of the most significant morphological characteristics, i.e., porosity, specific surface area, and strut dimensions. The modeler does not have to use scanned data but generates the structure automatically, as well as it generates a format that is ready to use for a CFD simulation. However, the usage of idealized unit cells does not account for random variations typical for natural foam structures. This leads to preferable flow directions in the cell structure. The second type of model structures applies stochastic approaches. For example, a perfectly ordered cell structure can be randomized by relocating nodes with a stochastic algorithm. However, the prediction of pressure drop or residence time is still unsatisfactory. Habisreuther et al. (2009) demonstrated that the parameters porosity and specific
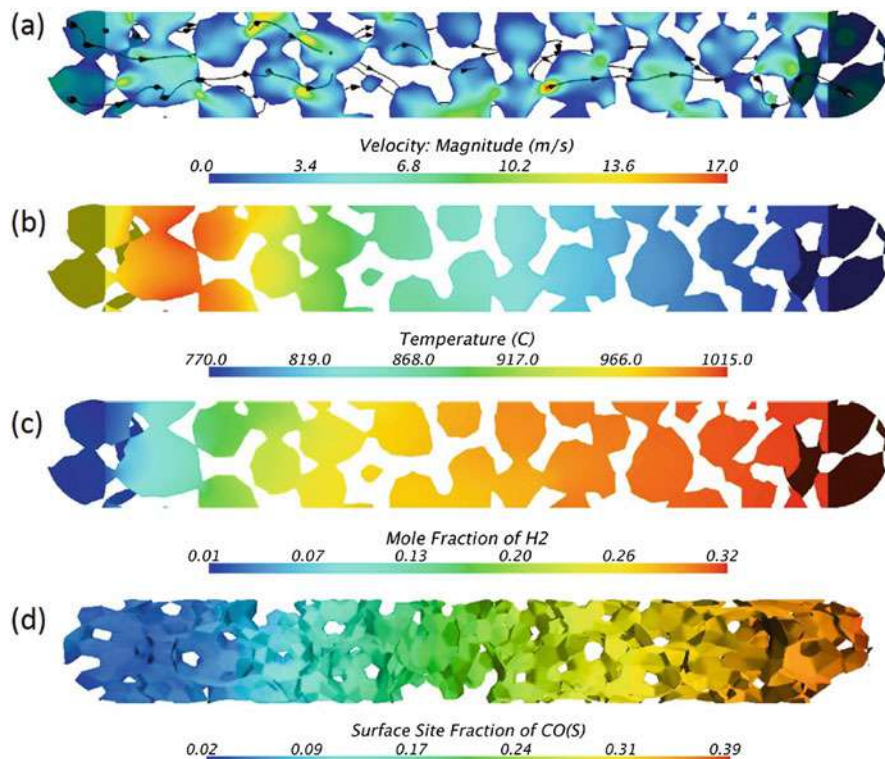
Fig. 9 Modeling of CPOX on a foam coated with rhodium. (**a**) Velocity magnitude with streamlines, (**b**) temperature, and (**c**) hydrogen mole fractions on a plane cut through the CPOX foam. (**d**) Adsorbed C* on the foam surface. (courtesy of G. Wehinger, related papers are Wehinger 2016 and Wehinger et al. 2016)

surface alone are insufficient to describe the transport phenomena in such complex geometries.

Works showing the modeling of foams including surface reactions are very scarce. Figure 9 shows examples of the calculation of CPOX in a tubular reactor filled with a foam (see also Wehinger 2016; Wehinger et al. 2016).

5.4 Particle Resolved Modeling of Fixed Bed Reactors

A fixed bed reactor is based on a container filled with randomly positioned particles. The understanding of fluid dynamics and their impact on conversion and selectivity in fixed bed reactors is still very challenging. For large ratios of reactor width to pellet diameter, simple porous media models are usually applicable. This simple approach becomes questionable as this ratio decreases. At small ratios, the individual local arrangement of the particles and the corresponding flow field are significant

for mass and heat transfer and, hence, the overall product yields. Therefore, the method of direct numerical simulation (DNS) is the most promising approach. Over the last years, researchers have used particle-resolved CFD simulations to model this kind of packed-bed arrangement more rigorously (Dudukovic 2009). The concept is to take into account the actual shape of each individual particle inside the reactor. Consequently, the fluid flow field is determined by the particles it has to pass by. The advantage is the absence of transport correlations, which are necessary in pseudo-homogeneous or heterogeneous models to describe the transport of momentum, heat, and mass. Even though the governing equations are relatively simple for laminar flows, this approach can usually be applied for small and periodic regions of the reactor only, which is caused by the huge number of computational cells needed to resolve all particles.

Several authors used detailed fluid dynamics in combination with pseudo-homogeneous kinetics due to the small number of reaction equations and species (Kuroki et al. 2009). However, these kinetics are often limited to a certain range of process parameters and could therefore not be applied to other flow regimes or reactor types. Especially, the species distribution inside fixed bed reactors is often insufficiently reproduced with such kinetics in contrast to the exit concentrations (Korup et al. 2011).

For a first-principles approach, modeling spatially resolved fluid dynamics must be combined with reliable kinetics, i.e., microkinetics. Additionally, the diffusion and reaction of the species within the catalyst pellets, which may be made up of microporous materials, have to be considered. More recently, a larger bed of catalytic spheres was investigated both experimentally and by particle-resolved CFD simulations (Wehinger 2016; Wehinger et al. 2016b). Figures 10 and 11 depict some typical results.

5.5   Modeling of Mobile Catalyst with one Fluid Phase

At the present, gas-fluidized beds (FBs) find a widespread application in the petroleum, chemical, metallurgical, and energy industries. The large increase in computational resources during recent decades made theoretical approaches for the investigation more applicable, since the governing equations describing FBs can now be solved within accessible time scales. Hence, theoretical modeling of FBs with CFD has become a trend in the field. Three different types of simulation approaches can be distinguished: direct numerical simulations (DNS), Eulerian-Eulerian (EE), and Eulerian-Lagrangian (EL) models. Contrary to DNS, the flow field is not resolved at the particle level in the last approaches mentioned, and consequently closures need to be applied, e.g., for fluid-particle interaction forces.

In DNS, the incompressible Navier-Stokes equations are solved for the gas phase, and the flow in the interstices of the particles bed is directly computed. DNS is the most popular tool for developing closure relations required in the averaged models mentioned below. Unfortunately, these simulations are restricted to a few thousand particles. In Eulerian-Eulerian (EE) models, in contrast, the particulate and the gas
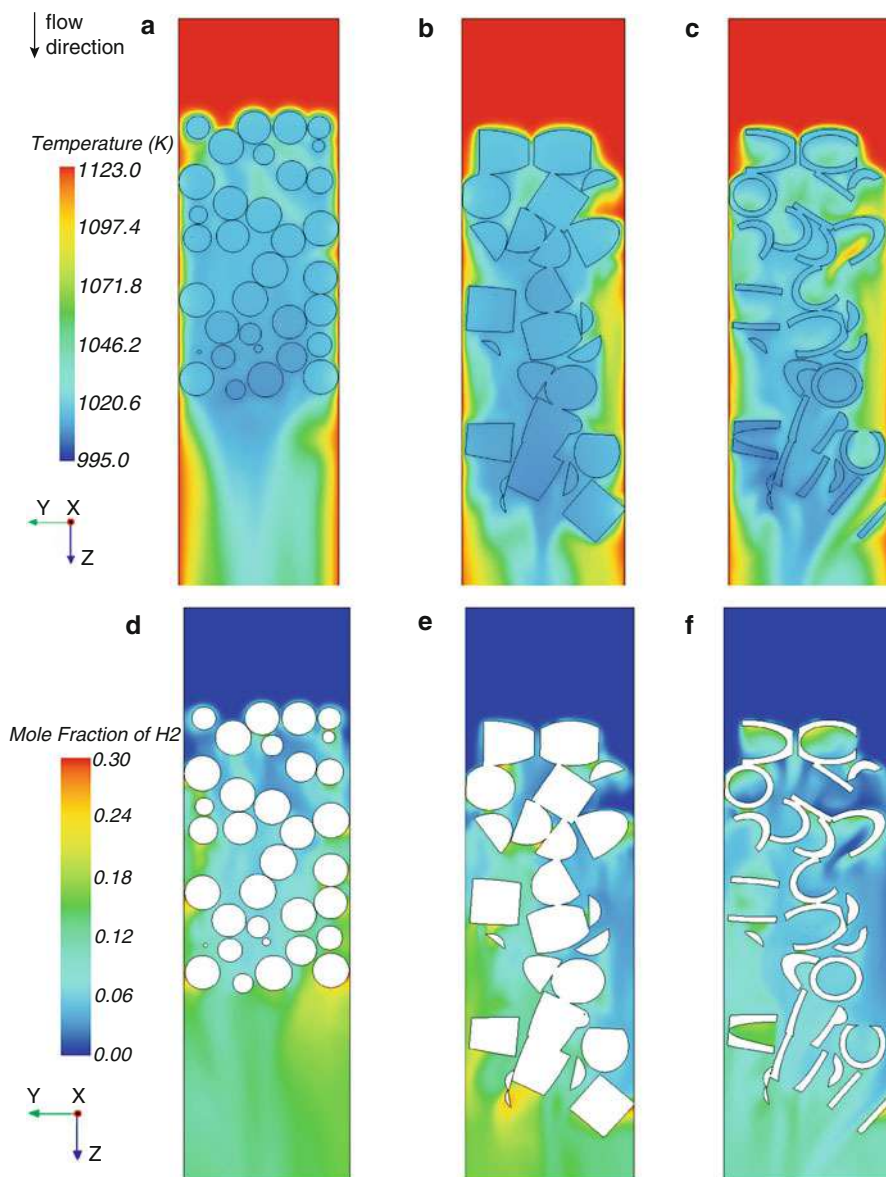
Fig. 10 Modeling of dry reforming of methane in a resolved fixed bed using three different particle shapes. Plane cut through the fixed bed. Temperature distribution: (**a**) spheres, (**b**) cylinders, (**c**) 1-hole cylinders. Mole fraction of hydrogen: (**d**) spheres, (**e**) cylinders, (**f**) 1-hole cylinders. (courtesy of G. Wehinger, related paper is Wehinger 2016)
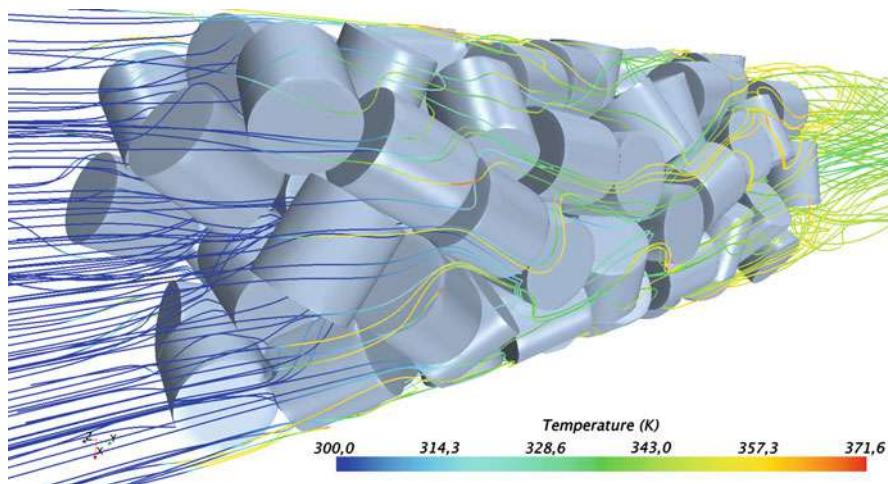
Fig. 11 Streamlines colored with temperature inside the bed of cylinders. (courtesy of G. Wehinger, related paper is Wehinger et al. 2016b)

phase are considered as interpenetrating continua, necessitating additional closure laws. EE models are preferred for the simulation of large-scale devices. Figure 12 shows an example for the EE modeling of an industrial scale fluidized bed for olefin polymerization. The height of the calculation domain is about 5 m.

The use of an Eulerian-Lagrangian (EL) model (e.g., the discrete particle model, DPM) holds a much greater promise when it comes to the full-physics simulations of heterogeneous reactors. EL models describe the particle phase with discrete entities, e.g., using the discrete element method (DEM) where every particle is tracked individually. EL approaches typically compute the particle-phase interactions directly and hence do not require complex rheological models and boundary conditions for the granular phase. This opens the door for the application to systems involving non-spherical and irregular particles. While a large number of work focused on inert systems, CFD-DEM calculations accounting for chemical reactions are relatively scarce. The latter are mainly found in the field of biomass and coal conversion (for a recent review, see Zhong et al. (2016)). Typically, the employed model assumes a homogeneous concentration and temperature distribution within the particle. This is unrealistic in situations involving fast chemical reactions, small pore diameters, or in case depositions form on the surface of the particles. CFD-DEM calculations accounting for intra-particle transport phenomena are even more rarely done; recently universal open-source tools were developed (Radl et al. 2015). Figure 13 shows an example of the calculation of a fluidized bed including surface reactions with an EL approach. In the left plot, the temperatures of the particles are shown and in the right plot (Hettel et al. 2016) the burst of particles at the upper surface of the bed. Additionally, a part of the CFD grid on the surrounding wall is shown.
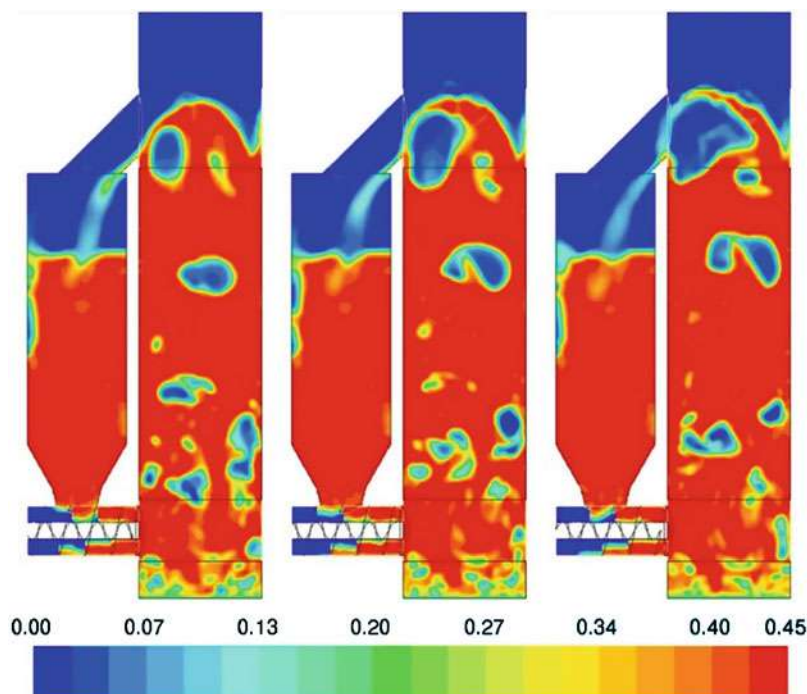
Fig. 12 Snapshots of the solids volume fraction. The time interval between two consecutive pictures is $\Delta t = 0.1$s. (Schneiderbauer et al. 2015, reprinted with permission from Elsevier)

5.6   Modeling of Catalytic Reactors with Multiple Fluid Phases

The catalytic reactors considered so far contain a solid phase (the catalyst and its supports) and one fluid phase (i.e. a gas or a liquid). However, there are also various classes of reactions performed in catalytic reactors containing multiple fluid phases (usually a gas and a liquid or less often two immiscible liquids) (Önsan and Avci 2016). Examples for heterogeneously catalyzed gas-liquid reactions are hydrogenations, oxidations, hydroformylations, and Fischer-Tropsch synthesis (FTS). The corresponding reactors can be classified in types with immobile catalyst (e.g. fixed beds, trickle beds, solid sponges, monoliths, and other structured reactors) and types with significant motion of the catalyst (i.e. slurry reactors such as bubble columns or stirred tanks).

Relevant phenomena in three-phase catalytic reactors arise on multiple scales reflecting multiple physics. The dimensions range from catalyst pores to single particles/bubbles/drops and over particle/bubble clusters to reactor internals toward the entire vessel. Gas-to-liquid processes such as FTS are frequently carried out in slurry reactors (Wang et al. 2007). There, the catalyst particles (diameter 1 to 100 μm) are suspended in the liquid phase while the injected gas rises in the form of
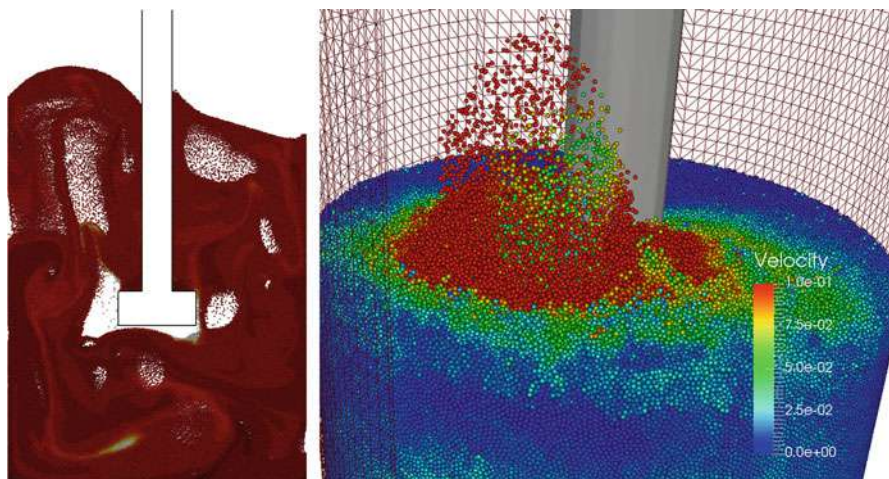
Fig. 13 Modeling of a two-zone fluidized bed reactor for the conversion of butane to butadiene. Left: Snapshot of the particle temperature in a 2D domain (white, T = 710 K; black, T = 700 K). Diameter/number of particles: 70 µm/250,000 (courtesy of Prof. S. Radl, University Graz, Austria). Right: Snapshot of the particle velocity in a 3D domain. Diameter/number of particles: 200 µm/2,500,000. (Hettel et al. 2016)

bubbles (diameter 1 to 10 mm) providing the mixing and catalyst suspension within the vessel (height up to 25 m). Coupling of hydrodynamics, heat and mass transfer, and reaction kinetics takes place at molecular and particle levels where conductive and convective transfer and diffusion within the internal pores of the catalyst are accompanied by the adsorption, surface reaction, and desorption of reactant and product on the surface.

Full interface-resolving multiphase CFD simulations of at least some of the relevant phenomena listed above are only possible for small sections of large-scale reactors and for microreactors. In the general case, multidimensional CFD approaches relying on a time averaging (Ishii and Hibiki 2006) or volume averaging procedure are indispensable instead (Jakobsen 2008). Mathematically, the averaging of the single-phase conservation equations for mass, momentum, and energy for all phases yields the Eulerian multi-fluid model, where the phases are represented as interpenetrating continua. For slurry reactors, the liquid and catalyst particles may optionally be represented by one slurry phase with the gas as additional phase.

As the system of averaged Eulerian multi-fluid equations is not closed, constitutive relations must be provided to model turbulence as well as momentum/heat/mass transfer between the phases. Interfacial transfer models (e.g. for the drag force and interfacial heat/mass transfer) are well established for dilute disperse flows (Sommerfeld et al. 2008), however, often relying on idealized situations (e.g. isolated particles moving through stagnant fluid). The development of interfacial transfer and turbulence models for non-dilute and non-disperse flow regimes which are valid in a wide range of operating conditions is actually a larger scientific

challenge than the numerical solution of the set of averaged multi-fluid equations themselves (Hjertager 2007).

In its standard form, the Euler-Euler model (two-fluid model) requires input of a (mean) particle/bubble/drop diameter. For narrow size distributions, the Sauter mean diameter may serve well for this purpose. For wider size distributions, which are standard rather than exceptional in technical reactors, various types of population balance approaches are available taking into account, e.g., bubble breakup/coalescence (Yeoh et al. 2014; Marchisio and Fox 2013).

A conceptually different approach for disperse flows is the Euler-Lagrange (EL) model, where the continuous phase is still represented in the Eulerian manner while the disperse elements (i.e., particles/drops/bubbles) are described in a Lagrangian way (Sommerfeld 2017; Crowe et al. 2011). There, an equation of motion is solved for a large number of particles or computational parcels yielding particle trajectories. The EL approach is often used for disperse two-phase flows (e.g., bubbly flow or sprays). The interaction between the phases is modeled depending on the disperse volume fraction α_P (see Fig. 14). For very dilute conditions ($\alpha_P < 5 \cdot 10^{-4}$), one-way coupling is sufficient, where the continuous phase affects the disperse phase but not vice versa. As α_P increases, two-way and four-way coupling are required, where in the latter case also the effects of particle-particle interactions are accounted for. For volume fractions $\alpha_P > 0.1$, the particle transport is no longer flow/collisions-driven but contact-driven. Such situations are frequently encountered in fluidized beds; they are best described by the discrete element method, in potential combination with CFD (see above).

Three-phase reactors with immobile catalyst are conceptually simpler to model by CFD since the solid domain is invariant in time. The hydrodynamics thus reduces to the simulation of the two-fluid flow coupled with the solid domain by boundary conditions reflecting the conjugate heat transfer and heterogeneous reactions. By a porous body approach (see above), such reactors may be computed on a large scale by the Eulerian multi-fluid model as well. However, reactor details such as a single monolith channel, a section of a trickle bed, a representative part of structured reactors, a portion of a solid foam, etc. are well amenable nowadays for interface-resolving simulation of the two-phase flow.

Fig. 14 Regimes of dispersed two-phase flows in terms of transport phenomena; left, dilute flow dominated by fluid transport of particles; middle, dense flow regime dominated by interparticle collision; right, dense flow regime that is particle contact dominated. (Sommerfeld 2017, reprinted with permission from Springer Nature)
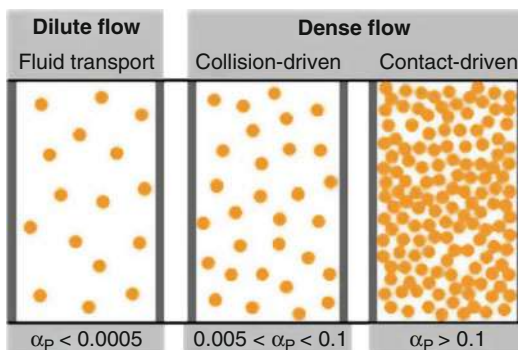
In recent years, microfluidic systems operated in continuous multiphase flows have emerged as useful platforms for synthesizing micro- and nanoparticles with controlled shape and size distribution. They have yielded particles of organic polymers, oxides, semiconductors, and metals as well as hybrid structures combining multiple materials and functionalities (Marre and Jensen 2010). A detailed review on numerical methods for interface-resolving multiphase CFD simulations in microfluidics and microprocess engineering (including the widely used volume-of-fluid and level set methods) along with an overview on related applications is given in Wörner (2012). While the number of computational studies combining two-phase hydrodynamics with mass transfer is increasing (especially for segmented gas-liquid flow), numerical studies incorporating chemical kinetics in microreactors in which the catalyst is dispersed on the solid wall are still rare. Here, the study of Woo et al. (2017) on the hydrogenation of nitrobenzene to aniline in gas-liquid Taylor flow within a single channel may serve as an example. The authors first computed the quasi-steady bubble shape and velocity field by a 2D simulation of a single flow unit cell (consisting of one Taylor bubble and liquid slug). Using these (frozen) hydrodynamics, they computed the transient mass transfer of hydrogen from the gas into the liquid phase toward the catalytic wall, where a detailed kinetic mechanism adapted from a density functional theory study is employed, consisting of four bulk species and ten surface species. Figure 15 illustrates the temporal evolution of educts and product in the bulk phases in combination with wall profiles of all intermediates.
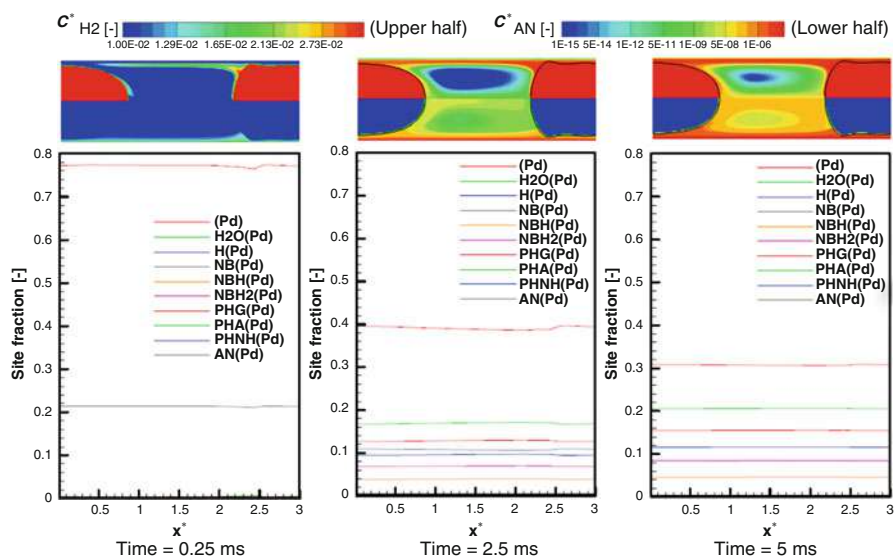
Fig. 15 Top row: Normalized instantaneous concentration distributions of hydrogen (upper half of channel) and aniline (lower half of channel) in gas-liquid Taylor flow. Bottom row: Instantaneous wall profiles of intermediate species. (Woo et al. 2017)

Even though the adequate modeling of physical complexity is challenging for multiphase CFD simulations, computations are a promising tool to achieve a better understanding of three-phase catalytic reactors.

6   Conclusions and Outlook

Computational fluid dynamics (CFD) is the analysis of systems involving fluid flow, heat transfer, and associated phenomena such as chemical reactions by means of computer-based numerical simulations. The technique is powerful and spans a wide range of industrial and nonindustrial applications. The advantages of CFD simulations over experiment-based approaches for reactor system design and optimization are the substantial reduction of development times and costs, the ability to study systems where experiments are difficult or impossible, the ability to study systems under hazardous conditions or beyond their normal performance limits, and the practically unlimited level of detail of results.

From a reaction engineering perspective, CFD simulations have matured into a powerful tool for understanding mass and heat transport in catalytic reactors. Initially, CFD calculations focused on a better understanding of mixing, mass transfer to enhance reaction rates, diffusion in porous media, and heat transfer. Over the last decade, the flow field and heat transport models have also been coupled with models for heterogeneous chemical reactions. So far, most of these models are based on the mean field approximation, in which the local state of the surface is described by its coverage with adsorbed species averaged on a microscopic scale. There is a gap between the length and time scales on the microscopic level and the length and times scales which can be resolved by a computational grid. This gap is bridged with models (e.g., mean field approximation). In the future, the gap will be shortened due to improved models, the combination of existing and advanced numerical approaches (e.g., Monte-Carlo simulations), and extended computer resources (higher discretization level). The currently increasing research activities on surface reactions at practical conditions will certainly boost the application of CFD codes that combine fluid flow and chemistry. New insights into the complexity of heterogeneous catalysis, however, will also reveal the demand for more sophisticated chemistry models. Their implementation into CFD simulations will then require robust numerical algorithms.

The simulation results will always remain a reflection of the models and physical parameters applied. Various commercial and open-source codes are available including a sometimes confusing number of methods and models. The careful choice of the submodels (turbulence, diffusion, species, reactions involved, etc.) and the physical parameters (boundary conditions, properties, etc.) is a precondition for reliable simulation results. Only the use of appropriate models and parameters, which describe all significant processes in the reactor, can lead to reliable results. Numerical algorithms never give the exact solution of the model equations but only an approximated solution. The solution is always dependent on the calculation grid (calculation domain, discretization level, grid type). Hence, error estimation is

needed. Moreover, the practical experience shows that the results from two different codes for the same complex physical problem are not equal.

CFD calculations require careful validation using experimental data or analytical solutions. Not till the model is evaluated, operational conditions or other parameters should be varied or extended to regions where no data is available. Owning the resources of a CFD program and computer hardware only is worthless. Qualified people having long experience in the field of modeling are needed to run the codes and to analyze and interpret the results. The understanding of the physics of fluid flow and the fundamentals of the numerical algorithms is a prerequisite. Having these crucial issues in mind, CFD can really serve as a powerful tool in understanding the behavior in catalytic reactors and in supporting the design and optimization of reactors and processes.

References

Bird RB, Stewart WE, Lightfoot EN (2001) Transport phenomena, 2nd edn. Wiley, New York

Blasi JM, Weddle PJ, Karakaya C, Diercks DR, Kee RJ (2016) Modeling reaction-diffusion processes within catalyst washcoats: II. Macroscale processes informed by microscale simulations. Chem Eng Sci 145:308–316

Choudary C, Mazumder S (2014) Direct numerical simulation of catalytic combustion in a multichannel monolith reactor using personal computers with emerging architectures. Comput Chem Eng 61:175–184

Cornejo I, Nikrityuk P, Hayes RE (2018) Multiscale RANS-based modeling of the turbulence decay inside of an automotive catalytic converter. Chem Eng Sci 175:377–386

Crowe CT, Schwarzkopf JD, Sommerfeld M, Tsuji Y (2011) Multiphase flows with droplets and particles, 2nd edn. CRC Press, Boca Raton

Dixon AG, Taskin ME, Nijemeisland M, Stitt EH (2011) Systematic mesh development for 3D CFD simulation of fixed beds: single sphere study. Comput Chem Eng 35(7):1171–1185

Dudukovic MP (2009) Frontiers in reactor engineering. Science 325:698–701

Dybbs A, Edwards R (1984) A new look at porous media fluid mechanics – darcy to turbulent. In: Bear J, Corapcioglu M (eds) Fundamentals of transport phenomena in porous media. Vol. 82 of NATO ASI series. Springer, Netherlands, pp 199–256

Fox RO (2003) Computational methods for turbulent reacting flows. Cambridge University Press, Cambridge

Habisreuther P, Djordjevic N, Zarzalis N (2009) Statistical distribution of residence time and tortuosity of flow through open-cell foams. Chem Eng Sci 64(23):4943–4954

Hayes RE, Kolaczkowski ST (1997) Introduction to catalytic combustion. Gordon and Breach Science Publ, Amsterdam

Hayes RE, Fadic A, Mmbaga J, Najafi A (2012) CFD modelling of the automotive catalytic converter. Catal Today 188:94–105

Hettel M, Diehm C, Bonart H, Deutschmann O (2015) Numerical simulation of a structured catalytic methane reformer by DUO: the new computational interface for OpenFOAM® and DETCHEM™. Catal Today 258:230–240

Hettel M, Denev JA, Deutschmann O (2016) Two-zone fluidized bed reactors for butadiene Production: a multiphysical approach with solver coupling for supercomputing application. In: Nagel W, Kröner D, Resch M (eds) High performance computing in science and engineering '16. Springer International Publish AG, Cham

Hettel M, Daymo E, Deutschmann O (2018) 3D modeling of a CPOX-reformer including detailed chemistry and radiation effects with DUO. Comput Chem Eng 109:166–178

Hjertager BH (2007) Multi-fluid CFD analysis of chemical reactors. In: Marchisio DL, Fox RO (eds) Multiphase reacting flows: modelling and simulation. Springer, Vienna, pp 125–179

Ishii M, Hibiki T (2006) Thermo-fluid dynamics of two-phase flow. Springer, New York/ London

Jakobsen HA (2008) Chemical reactor modeling: multiphase reactive flows. Springer, Berlin

Karakaya C, Weddle PJ, Blasi JM, Diercks DR, Kee RJ (2016) Modeling reaction-diffusion processes within catalyst washcoats: I. Microscale processes based on three-dimensional reconstructions. Chem Eng Sci 145:299–307

Kee RJ, Coltrin ME, Glarborg P (2003) Chemically reacting flow. Wiley, New Jersey

Kerkhof P, Geboers, MAM (2005) Towards a unified theory of isotropic molecular transport phenomena. Am Inst Chem Eng J 51:79–121

Korup O, Mavlyankariev S, Geske M, Goldsmith CF, Horn R (2011) Measurement and analysis of spatial reactor profiles in high temperature catalysis research. Chem Eng Process Process Intensif 50(10):998–1009

Kumar A, Mazumder S (2010) Toward simulation of full-scale monolithic catalytic converters with complex heterogeneous chemistry. Comput Chem Eng 34:135–145

Kuroki M, Ookawara S, Ogawa K (2009) A high-fidelity CFD model of methane steam reforming in a packed bed reactor. J Chem Eng Jpn 42(supplement):s73–s78

Lemos MJS (2006) Turbulence in porous media, modeling and applications. Elsevier, Amsterdam

Libby PA, Williams FA (eds) (1994) Turbulent reacting flows. Academic Press, London

Maestri M, Beretta A, Groppi G, Tronconi E, Forzatti P (2005) Comparison among structured and packed-bed reactors for the catalytic partial oxidation of CH_4 at short contact times. Catal Today 105:709–717

Marchisio DL, Fox RO (2013) Computational models for polydisperse particulate and multiphase systems. Cambridge series in chemical engineering. Cambridge University Press, Cambridge

Marre S, Jensen KF (2010) Synthesis of micro and nanostructures in microfluidic systems. Chem Soc Rev 39:1183–1202

Nien T, Mmbaga JP, Hayes RE, Votsmeier M (2013) Hierarchical multi-scale model reduction in the simulation of catalytic converters. Chem Eng Sci 93:362–375

Önsan ZI, Avci AK (2016) Multiphase catalytic reactors theory, design, manufacturing, and applications. Wiley, Hoboken

OpenFOAM-The Open Source CFD Toolbox (2017) www.openfoam.org

Oran ES, Boris JP (1987) Numerical simulation of reactive flow. Elsevier, Amsterdam

Park HM (2018) A multiscale modeling of fixed bed catalytic reactors. Int J Heat Mass Transf 116:520–531

Patankar SV (1990) Numerical heat transfer and fluid flow. Series in computational methods in mechanics and thermal science. McGraw-Hill, New York

Peters N (2000) Turbulent combustion. Cambridge University Press, London

Poinsot T, Veynante D (2001) Theoretical and numerical combustion. R. T. Edwards, Inc., Philadelphia

Porter S, Saul J, Aleksandrova S, Medina H, Benjamin S (2016) Hybrid flow modelling approach applied to automotive catalysts. Appl Math Model 40:8435–8445

Pope SB (2000) Turbulent flows. Cambridge University Press, Cambridge

Radl S, Forgber T, Kloss C, Aigner A (2015) ParScale - a compilation of particle scale models. https://github.com/CFDEMproject/ParScale-PUBLIC

Schneiderbauer S, Puttinger S, Pirker S, Aguayo P, Kanellopoulos V (2015) CFD modeling and simulation of industrial scale olefin polymerization fluidized bed reactors. Chem Eng J 264: 99–112

Sharma AK, Birgersson E (2016) Validity and scalability of an asymptotically reduced single-channel model for full-size catalytic monolith converters. Appl Math Comput 281:186–198

Sommerfeld M, van Wachem B, Oliemans R (2008) Best practice guidelines for computational fluid dynamics of dispersed multiphase flows. ERCOFTAC, SIAMUF Swedish Industrial Association for Multiphase Flows

Sommerfeld M (2017) Numerical methods for dispersed multiphase flows. In: Bodnár T, Galdi GP, Nečasová Š (eds) Particles in flows. Springer, Cham, pp 327–396

Sui R, Prasianakis NI, Mantzaras J, Mallya N, Theile J, Lagrange D, Friess M (2016) An experimental and numerical investigation of the combustion and heat transfer characteristics of hydrogen-fueled catalytic microreactors. Chem Eng Sci 141:214–230

Tischer S, Deutschmann O (2005) Recent advances in numerical modeling of catalytic monolith reactors. Catal Today 105:407–413

Versteeg HK, Malalasekera W (2007) An introduction to computational fluid dynamics, 2nd edn. Pearson, London

Wang TF, Wang JF, Jin Y (2007) Slurry reactors for gas-to-liquid processes: a review. Ind Eng Chem Res 46:5824–5847

Warnatz J, Dibble RW, Maas U (1996) Combustion, physical and chemical fundamentals, modeling and simulation, experiments, pollutant formation. Springer, New York

Wehinger GD (2016) Particle-resolved CFD simulations of catalytic flow reactors. PhD Thesis, Technical University Berlin

Wehinger GD, Heitmann H, Kraume M (2016) An artificial structure modeler for 3D CFD simulations of catalytic foams. Chem Eng J 284:543–556

Wehinger GD, Kraume M, Berg V, Korup O, Mette K, Schlögl R, Behrens M, Horn R (2016b) Investigating dry reforming of methane with spatial reactor profiles and particle-resolved CFD simulations. AICHE J 62:4436–4452

Wilcox CC (1998) Turbulence modeling for CFD. DCW Industries, La Canada, California, United States

Wörner M (2012) Numerical modeling of multiphase flows in microfluidics and micro process engineering: a review of methods and applications. Microfluid Nanofluid 12:841–886

Woo M, Wörner M, Maier L, Tischer S, Deutschmann O (2017) A numerical study on gas-liquid Taylor flow for catalytic hydrogenation of nitrobenzene with detailed kinetic mechanism, Annual meeting ProcessNet section Multiphase Flows, March 14–15, 2017, Dresden, Germany, https://doi.org/10.5445/IR/1000068709

Yeoh GH, Cheung CP, Tu J (2014) Multiphase flow analysis using population balance modeling. Butterworth-Heinemann, Oxford

Zhong W, Yu A, Zhou G, Xie J, Zhang H (2016) CFD simulation of dense particulate reaction system: approaches, recent advances and applications. Chem Eng Sci 140:16–43

Structure of Electrode-Electrolyte Interfaces, Modeling of Double Layer and Electrode Potential

59

Axel Groß

Contents

Abstract

Our current understanding of the structure of electrode-electrolyte interfaces and double layers will be presented from a theoretical point of view. Furthermore, recent attempts to consider varying electrode potentials in theoretical studies will be discussed. Although there have been significant advances in recent years as far as the modeling of electrode-electrolyte interfaces is concerned, still further progress is needed to get a more complete understanding of these scientifically interesting and technologically relevant systems.

1 Introduction

There is a growing interest in the structure of electrochemical electrode-electrolyte interfaces because of their importance for a sustainable future energy technology

A. Groß (✉)
Helmholtz Institute Ulm (HIU) Electrochemical Energy Storage, Ulm, Germany

Institute of Theoretical Chemistry, Ulm University, Ulm, Germany
e-mail: axel.gross@uni-ulm.de

(Schlögl 2010). Electrochemical energy storage and conversion processes occur at these interfaces in devices such as batteries, fuel cells, or electrolyzers. Still, there is a long tradition of studying solid-electrolyte interfaces (Schmickler and Santos 2010). They correspond to the interface between an electron conductor (the electrode) and an ionic conductor (the electrolyte) (Schmickler and Santos 2010). At these interfaces, so-called electrical double layers (EDL) form. From a theoretical point of view, the double layers were first described in a continuum approach in which the electrolyte was represented by a dielectric medium, while atomistic aspects such as the orientation of the molecules directly at the electrode were taken into account in an empirical manner. The atomistic modeling of these interfaces is theoretically and numerically rather demanding, in particular for liquid electrolytes, as a proper description requires to perform statistical averages over the many possible configurations of the electrolyte. Furthermore, instead of total energies, free energies need to be evaluated, taking entropy effects into account.

A complete elucidation of the structure of electrode-electrolyte interfaces still requires an atomistic understanding, in particular if one is interested in charge-transfer processes or electrocatalytic reactions at these interfaces. Due to the ever-increasing computer power and the development of more efficient algorithms, an atomistic description of electrode-electrolyte interfaces has become possible. First, such studies were based on parameterized interaction potentials, but nowadays a quantum chemical first-principles modeling is feasible. At the same time, experimental electrochemical tools with atomistic resolution have been developed making an electrochemical surface science approach (Kolb 2002) possible.

Still, it is fair to say that our atomistic understanding of the structure of electrode-electrolyte interfaces is still far from being complete. Experimental techniques based on the scattering and diffraction of electronic beams are typically not operative in solution. As far as theory is concerned, apart from the requirement to take the disordered structure of the liquid electrolyte into account, it is important to note that structures and properties of electrode-electrolyte interfaces are strongly influenced by the electrode potential. This adds considerable complexity to the theoretical treatment (Schnur and Groß 2009) as varying the electrode potential changes the excess charge at the interface. Consequently, even such a seemingly simple issue as the structure of water layers at close-packed metal electrodes is still a matter of debate (Schnur and Groß 2009; Michaelides 2006).

We will now discuss some basic properties of electrode-electrolyte interfaces using the schematic drawing shown in Fig. 1. Any electrolyte contains ions to make it conductive. In principle, any aqueous electrolyte also contains protons and/or hydroxide, depending on the given pH value. Typically concentrations are below the 1 M level. Recall that a concentration of 1 M corresponds to 1 ion per 55 water molecules. Still, at the electrode-electrolyte interface, the concentration of ions can be much higher because of the strong ion-electrode interaction. This is illustrated in Fig. 1 for the anions. Typically, anions interact more strongly than cations with electrodes. Hence anions often adsorb *specifically* to the electrode, which means that their solvation shell becomes broken up upon adsorption. In contrast, the more weakly interacting cations usually adsorb *non-specifically* with their solvation shell still intact.
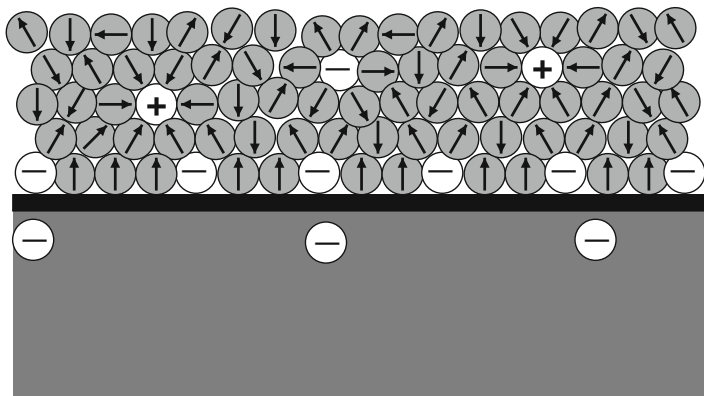
Fig. 1 Schematic drawing of an electrode-electrolyte interface between a metal electrode (lower part) and an aqueous electrolyte (upper part). The gray-shaded balls represent water molecules; the arrows denote their dipole moment. Solvated cations are illustrated by the white balls with a plus sign and anions in the electrolyte by white balls with a minus sign. Anions can either adsorb specifically at the electrode or non-specifically with their solvation shells still intact. In the electrode, the only charge carriers are the electrons, denoted by white balls with a minus sign. Their explicit presence at the electrode-electrolyte interface just illustrates that the electron density at the interface differs from the bulk metal density

The arrows in Fig. 1 denote the dipole moment of the water molecules. In the drawing, it is assumed that all the water molecules directly at the electrode surface are aligned which is the basis of the Stern layer described below. However, ab initio molecular dynamics simulations suggest that the first water layer at metal electrodes is not ordered (Schnur and Groß 2009). Still there might be a preferential orientation of the water molecules in the first layer at the electrode. At oxide electrodes, water molecules are rather bound through direct covalent bonds to the electrode (Cheng and Sprik 2014; Bonthuis and Netz 2013). Here a picture as shown in Fig. 1 might be more appropriate.

In this contribution, the structure of electrode-electrolyte interfaces will be addressed from a theoretical point of view. We will first determine the surface coverage of adsorbed ions within a thermodynamic picture in a grand-canonical approach. Even if within this approach the presence of the electrolyte at the electrode is not explicitly taken into account, still meaningful results can be derived. Then continuum models of electrode-electrolyte interfaces will be discussed. Atomistic studies of aqueous electrolytes at electrodes will be presented in which statistical averaging can be done via molecular dynamics simulations. Finally, recent approaches to incorporate varying electrode potentials into the theoretical description will be introduced. This chapter will show that further progress is needed in order to obtain a more complete picture of electrode-electrolyte interfaces. Still, this is an exciting field of research which is not only scientifically very interesting but also technologically rather relevant.

2   Thermodynamic Considerations

The structure of electrode-electrolyte interfaces is determined by the interaction between the electrode and electrolyte which seems to be a trivial statement. However, in a thermodynamics sense, the interface corresponds to an open system in contact with a reservoir. As mentioned above, electrolytes contain solvated ions in order to make them electrically conducting. In thermodynamic equilibrium, these ions are characterized by their (electro-)chemical potential. The adsorption of an ion species at the electrode can be thermodynamically stable (Reuter and Scheffler 2001) if the Gibbs free energy of adsorption $\Delta\gamma$ of adsorbates bound to a surface area A_s:

$$
\begin{aligned}
\Delta\gamma(T, c) &= \gamma(T, c, N_{\text{ads}}) - \gamma_{\text{clean}}(T, p, 0) \\
&= \frac{1}{A_{\text{S}}}\big(G_{\text{surf,ads}} - G_{\text{surf,0}} - \sum_i n_i \tilde{\mu}_i\big)
\end{aligned}
\tag{1}
$$

is negative. Here, $G_{\text{surf,ads}}$ and $G_{\text{surf,0}}$ are the Gibbs free energies of the adsorbate-covered and the clean surface, respectively, n_i is the number of adsorbed atoms of type i per surface area A_{S}, and $\tilde{\mu}_i$ is the corresponding electrochemical potential:

$$
\tilde{\mu}_i = \mu_i + q_i eU,
\tag{2}
$$

of adsorbate species i where q_i is the charge of the particle and U is the electrode potential. The chemical potential μ_i contains all solvation effects of the species. However, in order to evaluate solvation energies, computationally demanding thermodynamic integration schemes are necessary (Leach 2001).

Concentrating first on solvated protons, these efforts can be avoided by realizing (Nørskov et al. 2004, 2005) that at standard conditions (pH= 0, $p = 1$ bar, $T = 298$ K, $U = 0$), there is an equilibrium between protons in aqueous solution together with an electron at the Fermi level of the electrode $H^+(\text{aq}) + e^-$ and hydrogen in the gas phase, $\frac{1}{2}H_2(\text{g})$. In addition, the variation of the electrochemical potential of the proton and the electron change with proton concentration and electrode potential is well-known (Schmickler and Santos 2010):

$$
\tilde{\mu}(H^+(\text{aq})) + \mu(e^-) = \frac{1}{2}\mu(H_2(\text{g})) - eU_{\text{SHE}} - k_{\text{B}}T\ln(10)\text{pH},
\tag{3}
$$

where U_{SHE} is the electrode potential with respect to the standard hydrogen electrode (SHE). This approach is now referred to as the computational hydrogen electrode (CHE) (Nørskov et al. 2004, 2005); it is equivalent to the concept of ab initio thermodynamics for heterogeneous catalysis at the solid-gas interface (Reuter and Scheffler 2001).

It is not only applicable for hydrogen and protons; it works also for any redox couple such as $\frac{1}{2}A_2 + e^- \rightleftharpoons A^-$ (Hansen et al. 2010) where A_2 can be, for example, a halogen molecule, resulting in an electrochemical potential:

$$\tilde{\mu}(A^-(aq)) - \mu(e^-) = \frac{1}{2}\mu(A_2(g)) + e(U_{SHE} - U^0) + k_B T \ln a_{A^-}, \qquad (4)$$

where U^0 is the reduction potential of the corresponding halide vs. U_{SHE} and a_{A^-} its activity coefficient. Furthermore, this concept can also be extended to more than one solvated species in the electrolyte (Gossenberger et al. 2016).

Up to here the whole derivation is exact. Typically, the electrochemical potential $\tilde{\mu}_i$ is taken as a parameter characterizing the electrochemical conditions of the system. Still, Eq. 1 requires to evaluate surface free energy differences which is computationally very demanding. In heterogeneous catalysis where adsorption occurs at the solid-gas interphase, the dependence of the free energy of the adsorbate-covered surfaces is typically only rather weakly dependent on the environment represented by partial pressures of the gas phase species and temperature (Reuter and Scheffler 2001). Furthermore, changes in entropy and zero-point energies upon adsorption are typically small (Reuter and Scheffler 2001) so that they can be neglected. Hence surface free energies are often replaced by surface total energies (Todorova et al. 2003; Rogal et al. 2007).

At electrochemical interfaces, typically corresponding approximations are made, i.e., the presence of the electrolyte, entropy effects, and varying electrode potentials is neglected, so that the surface free energy is expressed as

$$\Delta\gamma = \frac{1}{A_S}\Big(E_{ads} - \sum_i n_i \Delta\tilde{\mu}_i(T, a_i, U)\Big). \qquad (5)$$

where E_{ads} is the total energy of adsorption in the absence of the electrochemical environment and $\Delta\tilde{\mu}_i(T, c_i, U)$ is the contribution to the electrochemical potential that depends on temperature, concentrations c_i, and electrode potential.

Hence the whole dependence on the particular electrochemical conditions is only taken into account with respect to the solvated species in the electrolyte. This seems to be a very rough approximation. Still, this approach has been rather successful in reproducing and explaining observed trends in electrocatalytic reactions such as oxygen reduction (Nørskov et al. 2004; Keith et al. 2010) and hydrogen evolution (Nørskov et al. 2005) on metal electrodes. This is most probably a consequence of the good screening properties of metallic electrodes which make binding energies hardly dependent on applied electric fields (Nørskov et al. 2004; Rossmeisl et al. 2006; Wang and Liu 2009; Jinnouchi et al. 2014). In addition, water typically interacts rather weakly with metal electrodes (Gohda et al. 2008).

We will now illustrate the power of this approach with respect to the equilibrium coverage of halides on metal electrodes (Gossenberger et al. 2015, 2016; Lin et al. 2016). Experimentally a competitive adsorption of protons and chloride adsorption on Pt(111) was observed (Garcia-Araez et al. 2005a, b), which means that as a function of increasing electrode potential, hydrogen adsorbed on Pt(111) is replaced by chlorine. Similar observations were also made in bromide and hydrogen co-adsorption (Garcia-Araez et al. 2006; Garcia-Araez and Koper 2010).
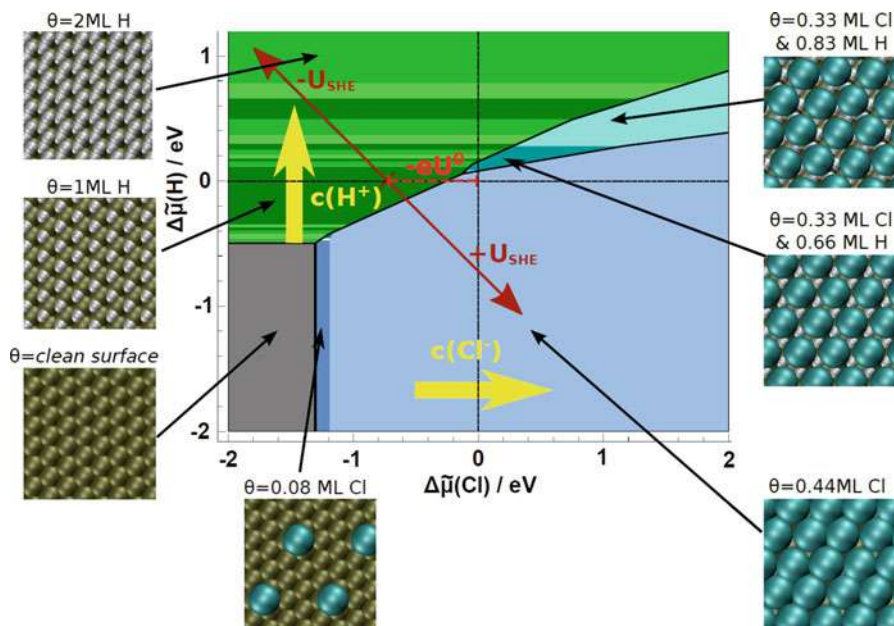
Fig. 2 Equilibrium phases of co-adsorbed chlorine and hydrogen on Pt(111) as a function of the electrochemical potential of hydrogen and chlorine (Gossenberger et al. 2016). The green areas denote hydrogen-containing adsorbate structures, the blue areas chlorine-containing structures, and the cyan areas correspond to mixed structures of both hydrogen and chlorine. The effect of increasing the concentration of the species individually is illustrated by the yellow arrows, whereas the red double arrow shows the effect of varying the electrode potential for fixed concentrations. (Reprinted from Gossenberger et al. (2016), with permission from Elsevier)

Using Eq. 5, the free energies of adsorption of hydrogen and chloride on Pt(111) were determined as a function of their electrochemical potential (Gossenberger et al. 2016). In total, the adsorption energies of 117 adsorbate structures were evaluated. Thus the phase diagram shown in Fig. 2 was derived. Interestingly enough, over a wide range of electrochemical potentials either pure hydrogen or pure chlorine adsorption phases or the clean electrode surface are stable. Naively, one would expect that adsorbed anions (chloride) and cations (protons) experience an attractive interaction. However, there is only a relatively small pocket at positive electrochemical potentials where both hydrogen and chlorine form a stable co-adsorbate structure. Usually one would assume that the adsorption of anions leads to a work function decrease, but in fact chlorine adsorption on Pt(111) causes an anomalous decrease of the work function for chlorine coverage below $\Theta_{Cl} = 0.5$ because of the strong polarization of chlorine upon adsorption (Gossenberger et al. 2014; Roman et al. 2014). This causes an effective dipole-dipole repulsion between hydrogen and chlorine upon co-adsorption, making co-adsorbate structure energetically unfavorable.
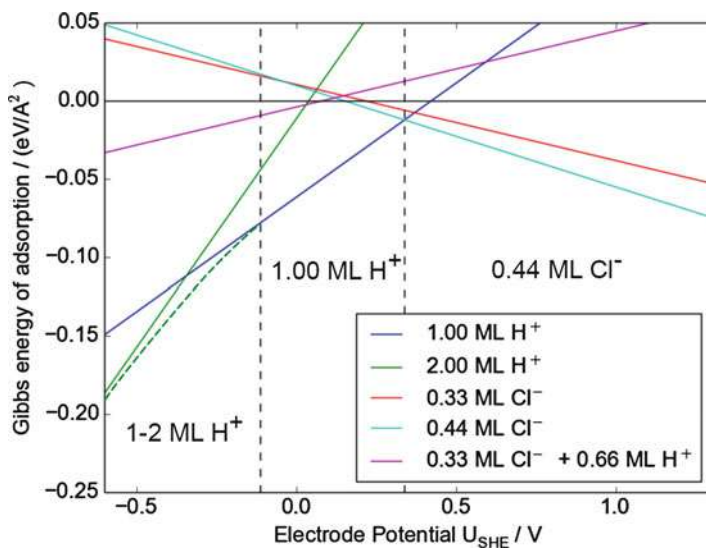
Fig. 3 Calculated electrochemical Gibbs energy of adsorption as a function of electrode potential on a Pt(111) surface under standard conditions (Gossenberger et al. 2016). The structure with the lowest energy as a function of electrode potential corresponds to the thermodynamical equilibrium structure. (Reprinted from Gossenberger et al. (2016), with permission from Elsevier)

Note that a phase diagram as shown in Fig. 2 cannot be directly compared to experiment as in electrochemical experiments the electrochemical potentials are not controlled but rather the electrode potential and the concentration of the solvated species. However, electrochemical potentials are a function of the electrode potential and the concentration (see Eqs. 3 and 4). A pure variation of the electrode potential for given concentrations corresponds to a change parallel to the red double arrow included in Fig. 2. At standard conditions, i.e., for an activity $a_{A^-} = 1$, the free energy of adsorption for a halide can be expressed as a function of the electrode potential according to

$$\Delta\gamma(U_{SHE}) = \frac{N_{ads}}{A_s}\left(E_{ads} - e(U_{SHE} - U^0)\right). \qquad (6)$$

For other concentrations of the species A^- in the electrolyte, the corresponding electrode potential needs to be shifted by $k_B T \ln a_{A^-}$ which at room temperature is equivalent to approximately 60 meV when the activity is changed by one order to magnitude.

The Gibbs energy of adsorption for hydrogen and chlorine on Pt(111) as a function of electrode potential is plotted in Fig. 3. This diagram shows the stable phases along the red double arrow in Fig. 2, but it also provides information about the free energies of adsorption for unstable phases. At low electrode potentials
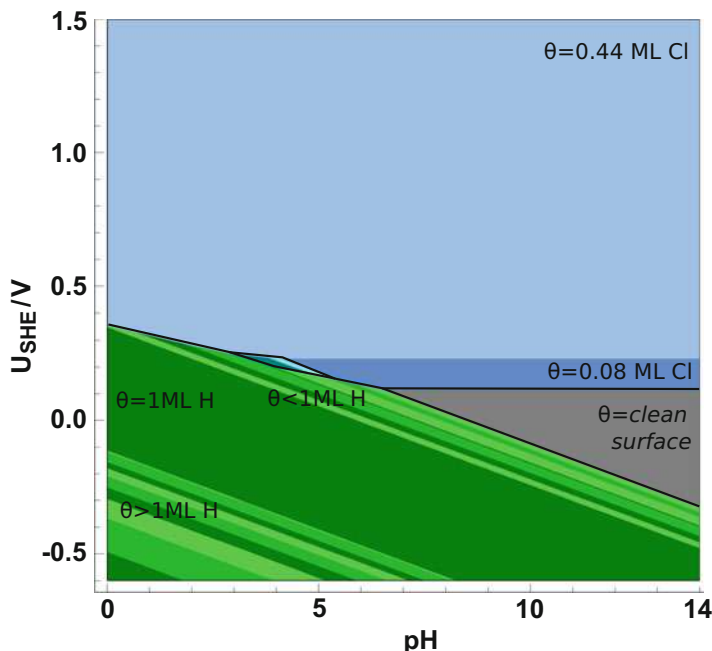
Fig. 4 Calculated Pourbaix diagram showing the stable phases of co-adsorbed chlorine and hydrogen on Pt(111) as a function of pH and electrode potential. The chlorine concentration corresponds to an activity of 0.1. (Reprinted from Gossenberger et al. (2016), with permission from Elsevier)

for which the electrode becomes negatively charged, the adsorption of cations is favored. Consequently, in this region the electrode is hydrogen-covered. As Fig. 3 shows, up to electrode potentials of about 0.45 V hydrogen adsorption is in principle thermodynamically stable with respect to hydrogen evolution. However, already below 0.45 V chlorine adsorption is energetically more favorable. Hence the adsorbed hydrogen atoms will be replaced by chlorine atoms, and the so-called underpotential deposition (upd) of hydrogen is shifted to lower potentials due to the replacement by chlorine. This example shows that phase diagrams based on the concept of the computational hydrogen electrode can successfully reproduce experimental findings with respect to the competitive character of hydrogen and chlorine adsorption on Pt(111) (Garcia-Araez et al. 2005a, b).

Figure 3 shows the stable phases for standard conditions, i.e., for pH = 0. The stable phases as a function of both the electrode potential and pH are plotted in Fig. 4. Such a so-called Pourbaix diagram can directly be compared with experimental finding. Note however that all this information is already implicitly included in Fig. 2, i.e., it corresponds to a transformation of the phase diagram as a function of the electrochemical potentials of the solvated species.

3   Continuum Models of Electrode-Electrolyte Interfaces

In the previous section, the ionic adsorption at electrode-electrolyte interfaces in equilibrium was addressed. However, the presence of the electrolyte was in fact not taken into account in the explicit calculations. We will now consider the properties and structure of the electrode-electrode interface at the first-principles level in more detail. We start by describing the interface in a continuum approach. At this interface, the so-called electrical double layer forms. One layer consists of ions on the electrode surface due to the chemical interaction between the ions and the electrode. This layer is then screened by a charge redistribution in the electrolyte. As this charge is locally not well confined, it is called the diffuse layer. In a simplified picture, the whole charge distribution can be regarded as being formed by two capacitor plates.

Helmholtz is given credit for being the first to realize that the charge distribution in the solution is realized as a single layer of ions adsorbed at the surface (Helmholtz 1853). The potential within this *Helmholtz layer* exhibits a linear behavior as a function of distance from the electrode. Early in the twentieth century, Gouy (1910) and Chapman (1913) refined this model by allowing for a diffuse thermal distribution of the ions.

Applying Boltzmann statistics to the distribution of both anions and cations of absolute charge $|ze|$ with a bulk density of n_0 yields the *Poisson-Boltzmann equation* (Schmickler and Santos 2010) for the potential $\phi(x)$:

$$\frac{d^2\phi}{dx^2} = -\frac{zen_0}{\varepsilon\varepsilon_0}\left(\exp\left(-\frac{ze\phi(x)}{k_BT}\right) - \exp\left(\frac{ze\phi(x)}{k_BT}\right)\right). \tag{7}$$

The exponentials can be linearized for $ze\phi(x)/k_BT \ll 1$, yielding an exponentially decreasing electric potential:

$$\phi(x) = \frac{\sigma}{\varepsilon\varepsilon_0\kappa}\exp(-\kappa z), \tag{8}$$

where σ is the surface charge density of the electrode and

$$\kappa = \left(\frac{2(ze)^2 n_0}{\varepsilon\varepsilon_0 k_BT}\right)^{1/2}, \tag{9}$$

is the inverse *Debye length*. Note that this formulation is rather close to the macroscopic treatment of the interface between an electron conducting n-type and a hole conducting p-type semiconductor, resulting in the formation of a space charge layer. The Gouy-Chapman model, however, does not take into account that ions

Fig. 5 Illustration of the Stern model describing the electric double layer in a continuum picture
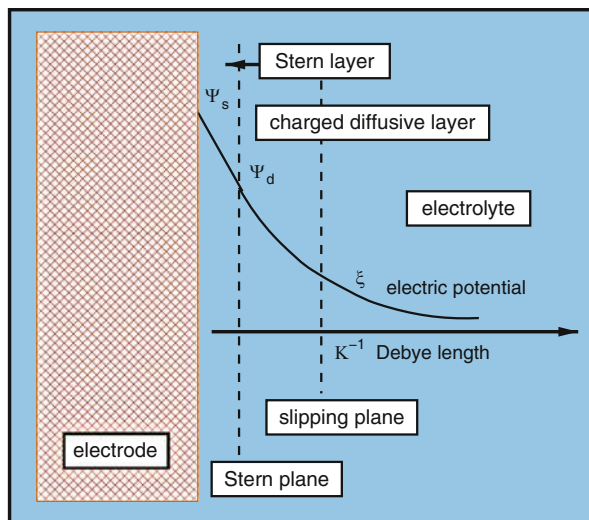
can adsorb specifically at the electrode surface immobilizing them. Therefore it does not allow to describe highly charged double layers. As a consequence, the Gouy-Chapman theory predicts that the capacity always has its minimum at the potential of zero charge. Still, at higher electrolyte concentrations, the capacity exhibits a much more complex dependence on the potential (Schmickler and Santos 2010).

Stern suggested a combination of the Helmholtz and Gouy-Chapman models in order to improve the description of the double layer (Stern 1924), as illustrated in Fig. 5. It consists of an internal Stern layer (i.e., a Helmholtz layer) in which the potential drops linearly and an outer diffuse layer (i.e., a Gouy-Chapman layer) in which the potential falls off exponentially. This yields the following expression for the capacity C:

$$\frac{1}{C} = \frac{1}{C_{GC}} + \frac{1}{C_H}. \tag{10}$$

C_{GC} is the Gouy-Chapman capacity and C_H the Helmholtz capacity which is independent of the electrolyte concentration.

This description yields a phenomenological macroscopic model of the electrode-electrolyte interface. It is very useful in order to understand and analyze general trends of the capacity at electrochemical interfaces. Still, a more quantitative theory of solid-electrolyte interfaces is needed, taking into account the atomistic structure of the interface in order to fully understand structures and processes at these interfaces.

4   Atomistic First-Principles Description of Solid/Water Interfaces

A more realistic picture of the structure of electrode-electrolyte interfaces requires an atomistic modeling associated with performing statistical averages over many different configurations. This calls for numerical efficient methods based on inter-atomic potentials which avoid to solve the quantum mechanical electronic structure problem. In fact, structural properties of aqueous electrolytes can be rather well reproduced by parameterized classical interaction potentials (Siepmann and Sprik 1995; Crozier et al. 2000; Bukowski et al. 2007). Such potentials have also been used quite frequently in molecular dynamics studies addressing the structure of the electrode-electrolyte interfaces at finite temperatures (Spohr 1989, 1997; Raghavan et al. 1991; Xia and Berkowitz 1995). Important insights into the structure of electrode-electrolyte interfaces have been gained from these studies.

However, interatomic potentials usually do not reproduce water *and* metal properties satisfactorily at the same time. On the one hand, force fields do not describe metals well. On the other hand, covalent bonding can typically not been appropriately reproduced by interpolation schemes including many-body effects such as the embedded atom method (EAM) (Daw and Baskes 1984; Daw et al. 1993) that are well suited to reproduce metallic properties. In addition, classical interaction potentials usually fail in describing bond breaking and bond making which would be required to treat electrocatalytic processes at electrode-electrolyte interfaces. Reactive force fields based on bond-order concepts (Goddard et al. 2003; Fantauzzi et al. 2014) allow to treat chemical reactions, but their training effort rises exponentially with the number of elements considered in the parameterization.

Modern machine learning techniques such as interpolation schemes based on artificial networks (Lorenz et al. 2006; Morawietz and Behler 2013; Behler 2014) are able to yield a reliable description of the structural properties of liquid water (Morawietz et al. 2016) and of water-metal interfaces (Natarajan and Behler 2016). However, as no chemical insights enter their construction procedure, their fitting usually requires large training sets (Lorenz et al. 2006).

Due to the ever-increasing computer power and the development of efficient algorithms, first-principles electronic structure methods based on density functional theory (DFT) can now be applied to describe the properties of electrode-electrolyte interfaces (Schnur and Groß 2009; Sakong et al. 2016). DFT calculations combine numerical efficiency with sufficient accuracy (Groß 2002; Hafner et al. 2006; Michaelides 2006). Thus also statistical averages can be performed using ab initio molecular dynamics (AIMD) simulations (Schnur and Groß 2009; Marx and Hutter 2009) in which the forces necessary to integrate the classical equations of motion are determined "on the fly" by DFT calculations. Based on DFT studies, we will first discuss the structure of water layers on close-packed metal electrodes and then later also turn to oxide electrodes.

Although DFT calculations yield often sufficient accuracy for materials properties, it should be noted that the popular PBE functional (Perdew et al. 1996) leads to

an overstructuring of liquid water (Fernández-Serra and Artacho 2004; VandeVondele et al. 2005; Liu et al. 2009), which means that the pair distribution functions are too much peaked compared to the measured distribution function (Soper 2007), i.e., the shell structure of PBE water is too pronounced.

As a simple solution of this problem, PBE water simulations have been run at a higher temperature of 350 K (Fernández-Serra and Artacho 2004; VandeVondele et al. 2005; Liu et al. 2009) which is, however, not a truly satisfactory approach. Recently, there was significant progress concerning the incorporation of the van der Waals interaction into DFT approaches by the development of reliable dispersion-corrected DFT functionals (Grimme 2004; Grimme et al. 2010; Tkatchenko and Scheffler 2009) and the improvement of van der Waals DFT functionals (Dion et al. 2004). It turns out that it is critical to include the van der Waals interaction in DFT calculations addressing water properties (Santra et al. 2007, 2008). The inclusion of dispersion does not only improve the description of liquid water (Zhang et al. 2011; Lin et al. 2012; Forster-Tonigold and Groß 2014; Sakong et al. 2016), but it also yields the correct wetting behavior of water on metal electrodes (Tonigold and Groß 2012; Carrasco et al. 2011, 2013).

For pure water simulations, DFT hybrid functionals including a certain amount of exact Hartree-Fock exchange are rather popular (Do and Besley 2012). However, for metal-water interfaces, they are not appropriate as the inclusion of exact exchange worsens the description of metal properties (Paier et al. 2007). Here the RPBE functional (Hammer et al. 1999) together with the so-called D3 dispersion corrections of Grimme (Grimme et al. 2010) yields a rather satisfactory description of the water-water (Forster-Tonigold and Groß 2014; Imoto et al. 2015; Sakong et al. 2016; Morawietz et al. 2016) and the water-metal interaction (Tonigold and Groß 2012) as it replaces the overestimated directional hydrogen bonds of PBE by nondirectional van der Waals bonds (Forster-Tonigold and Groß 2014).

Another source of error in the description of water is the possible role of nuclear quantum effects because of the low mass of hydrogen atoms. There are conflicting results with respect to the effect of these nuclear quantum effects on the properties of liquid water (Morrone and Car 2008; Chen et al. 2003; Habershon et al. 2009; Fritsch et al. 2014). As AIMD simulations of liquid water including dispersion corrections yield pair distribution function very close to the experiment (Forster-Tonigold and Groß 2014; Sakong et al. 2016), it is still unclear whether the incorporation of nuclear quantum effects is really necessary.

We will now discuss the structure of water adsorbed on metal electrodes using Fig. 6 as an illustration where the structure of a hexagonal icelike structure on a $Pt_1Ru_2/Pt(111)$ surface alloy is depicted (Fischer et al. 2016). Isolated water monomers typically bind in an almost flat configuration via their oxygen atom in an on-top configuration to metal surfaces with the hydrogen and oxygen atoms of the water molecules almost at the same height (Michaelides 2006; Schnur and Groß 2009, 2011), similar to the water molecules adsorbed on the Ru atoms of the $Pt_2Ru_1/Pt(111)$ surface alloy (Fischer et al. 2016) shown Fig. 6. The adsorption energy of water monomers on metals is relatively weak, ranging from -0.1 to -0.4 eV. Their binding energy on typical late transition metal surfaces is ordered
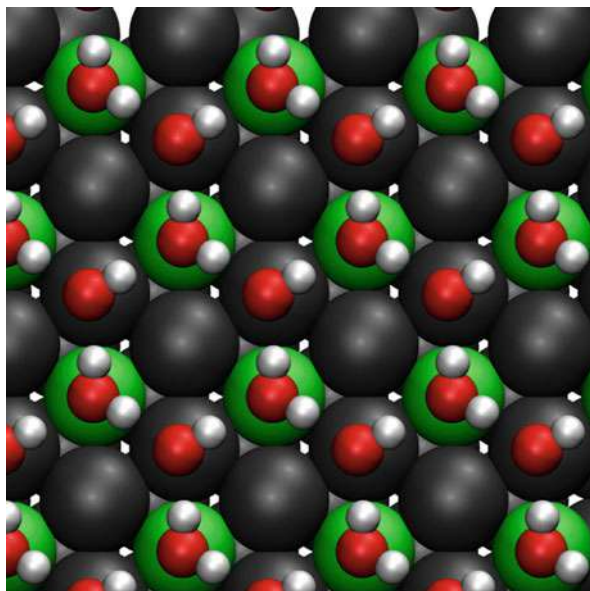
according to Au<Ag<Cu<Pd<Pt<Ru<Rh (Michaelides 2006). The weak interaction of water with metal surfaces is also reflected by the fact that the electronic local density of states of the metal atoms at the surfaces is not substantially altered upon water adsorption (Gohda et al. 2008; Schnur and Groß 2009). As a consequence, the binding energy of specifically adsorbed species to late transition metal electrode surfaces is hardly influenced by the presence of water (Roudgar and Groß 2005a, b; Sakong et al. 2015).

It is important to note that the rather nondirectional water-metal interaction is of the same magnitude than the directional water-water interaction mediated by hydrogen bonds. Hence the minimum energy structure of water layers adsorbed on metal electrodes corresponds to a compromise between the attractive water-metal and water-water interaction (Groß et al. 2014). As for close-packed hexagonal metal surfaces, icelike hexagonal rings fit rather nicely on top of them. However, in order to build a hexagonal icelike network, not all water molecules can bind in the optimum configuration through their oxygen atom to the metal atoms. In fact, there are two different possible configurations for the water molecules not bound via their O atom to the metal, namely, the so-called H-down and H-up structures with one hydrogen atom either pointing toward or away from the surface. Figure 6 depicts an H-down structure; the H-down hydrogen atom is actually not visible in the top view. The H-down and H-up structures are rather close in energy, and both structures can be stable depending on the particular choice of the metal: the H-up structure is more stable on Ni(111), Cu(111), and Ru(0001), whereas on Rh(111), Ag(111), Pt(111), Pd(111) (Michaelides 2006), and Pd/Au(111) (Roudgar and Groß 2005b), the H-down structure is energetically favorable. As far as the hexagonal

icelike layer shown in Fig. 6 is concerned, it is obvious that the water molecules interacting more strongly through their O atom with the metal are located above the Ru atom which reflects the stronger interaction of Ru with water monomers compared to Pt mentioned above.

The interaction of water, in particular, with noble metal surfaces such as Ag(111) and Au(111) is so weak that in fact it is more stable to form a three-dimensional ice cluster (Carrasco et al. 2011; Tonigold and Groß 2012), i.e., water does not wet these surfaces. In contrast, on Pt(111) and Pd(111), wetting layers form.

In spite of the weak interaction of intact water molecules with metal surfaces, water can still adsorb dissociatively on metals (Feibelman 2002; Michaelides 2006). On Ru(0001) a half-dissociated water layer should be preferred according to DFT calculations (Feibelman 2002) where every second water molecule is dissociated to OH. Ru is not the only metal where half-dissociated water layers should form; also on Rh and Ni, half-dissociated water layers should be energetically favorable (Michaelides 2006). Still, the water dissociation is associated with relatively high barriers so that the formation of the half-dissociated layer is kinetically hindered.

Up to now we have only considered minimum energy structures of water on close-packed metal electrodes. Still the question remains whether at room temperatures these water structures are really crystalline or rather liquid-like. In order to address this question, a systematic AIMD study of the structure of two water layers on Ag(111), Au(111), Pt(111), Pd/Au(111), and Ru(0001) at room temperature was performed (Schnur and Groß 2009). The simulations were started with the icelike bilayer as the initial configuration and then run for 10 ps. Figure 7 shows snapshots of the AIMD trajectories at 8 ps.

According to Fig. 7 for water on the noble metal surfaces Ag(111) and Au(111), there is almost no indication of the initial hexagonal geometry anymore after 8 ps. Also on Pd/Au(111) that is even more strongly interacting with adsorbates than Pd(111) (Roudgar and Groß 2003a, b) but has a larger lattice constant, the water structure is broken. On Pt(111) and Ru(0001), there is still some hexagonal water structure discernible which might be an artifact of the relatively small $2\sqrt{3} \times 2\sqrt{3}$ unit cell chosen. However, the orientation of the water molecules seems to be rather random; there is no indication of either the H-up or the H-down structure. Only the half-dissociated water structure on Ru(0001) stays rather fixed which is due to the fact that the hydroxyl groups strongly interact with the metal substrates and thus pin the water network.

Figure 7 suggests that at room temperature, water layers on close-packed metal electrodes are disordered, at least as far as the orientation of the water molecules is concerned. Experimentally, at the moment it is not really possible to directly monitor the orientation of water molecules in a water film on electrode surfaces. However, the work function change of electrodes upon water adsorption sensitively depends on the orientation of the water molecules. Because of opposite dipole orientation, the H-up and H-down water structures cause work function changes that differ by about 2 eV (Schnur and Groß 2009, 2011; Filhol and Bocquet 2007), as is shown in Fig. 8. The comparison with the experimentally measured work function changes
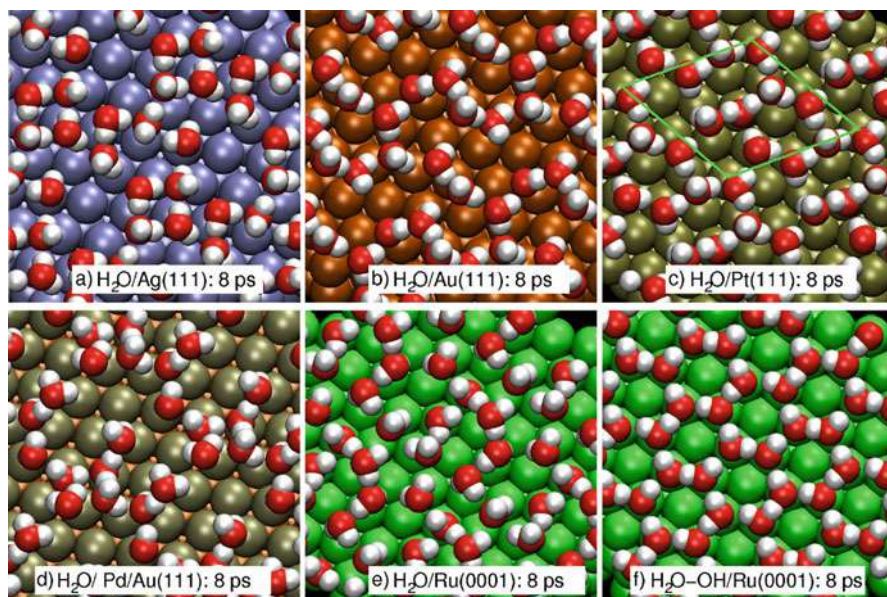
Fig. 7 Snapshots of AIMD simulations of two water layers on (**a**) Ag(111), (**b**) Ag(111), (**c**) Pt(111), (**d**) Pd/Au(111), and (**e**) Ru(0001) at 300 K. Only the first water layer is shown (Schnur and Groß 2009). In addition, a snapshot of a half-dissociated water layer on Ru(0001) is shown (panel f). The size of the surface unit cell used in the simulations is shown in panel c. (Reprinted from Schnur and Groß (2009) ©IOP Publishing Ltd and Deutsche Physikalische Gesellschaft. Reproduced by permission of IOP Publishing. CC BY-NC-SA)

for Au(111) (Heras and Viscido 1980), Pt(111) (Langenbach et al. 1984; Kiskinova et al. 1985), or Ru(0001) (Hoffmann and Benndorf 1997; Lilach et al. 2001) also shown in Fig. 8 reveals a large discrepancy. In fact, the experimentally observed work functions for these systems lie in between those calculated for the H-up and H-down structures.

In Fig. 8, also calculated work function changes averaged over AIMD runs for water on Au(111), Pt(111), and Ru(0001) at 150 and 300 K are shown. In fact, the discrepancy between measured and calculated work function changes is strongly reduced when the thermal disorder in the orientation of the water molecules is taken into account through AIMD simulations (Schnur and Groß 2009, 2011) because then the preferential orientation of the water molecules and thus of their dipole moments is significantly reduced. This yields strong evidence that indeed water layers at close-packed fcc(111) electrodes are disordered at room temperature and not aligned, as assumed in Fig. 1.

The determination of the work function is not only important to derive the mean orientation of adsorbed water molecules, it is also directly related to the electrode potential (Trasatti 1995) in electrochemical systems. For example, the so-called potential of zero charge (pzc) is given by the work function of an electrode
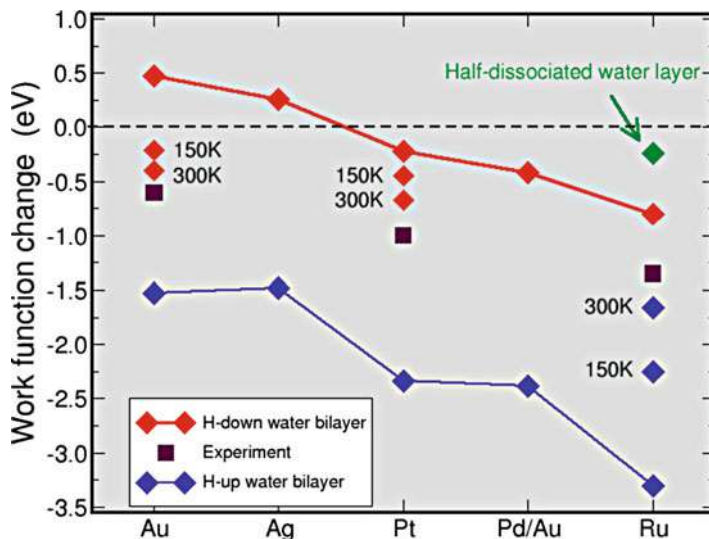
Fig. 8 Work function change induced by the presence of H-up and H-down water bilayers on Au(111), Ag(111), Pt(111), Pd/Au(111) and Ru(0001) according to DFT calculations (Schnur and Groß 2009). For Ru, also the work function change induced by the half-dissociated water layer is included. Furthermore, the average work function change determined in AIMD runs with at 150 and 300 K is compared to experimental results for Au(111) (Heras and Viscido 1980), Pt(111) (Langenbach et al. 1984; Kiskinova et al. 1985) and Ru(0001) (Hoffmann and Benndorf 1997; Lilach et al. 2001). (Reprinted from Schnur and Groß (2011), with permission from Elsevier)

covered by a pure, ion-free water film (Trasatti 1986, 1991; Tripkovic et al. 2011). In order to evaluate the pzc of Pt(111), ab initio molecular dynamic simulations of six water layers on a five-layer Pt(111) slab within a 3×3 unit cell were run for more than 30 ps (Sakong et al. 2016). In Fig. 9, the work function calculated along this run for the first 30 ps is plotted. On short time scales, the work function can vary by ± 1 eV because of the still rather small size of the surface unit cell. The red line in Fig. 9 corresponds to smoothened values averaged over 5 ps. The overall distribution of the work function along the AIMD run is plotted in the inset of Fig. 9. A necessary but not sufficient condition for performing thermal averages in the canonical ensemble is that the quantities of interest exhibit a Gaussian distribution which is satisfactorily fulfilled in this case. A Gaussian fit to the simulation results yields a mean value of the work function of 4.88 V which is in rather good agreement with the experimentally observed absolute pzc of 4.9 V measured using the CO adsorption technique (Trasatti 1991; Weaver 1998; Cuesta 2004; Petrii 2013). This also lends credibility to the simulation results. Assuming a value of 4.44 V for the absolute standard hydrogen electrode (SHE) potential as suggested by both theory and experiments (Tripkovic et al. 2011; Trasatti 1983, 1986; Reiss and Heller 1985), a calculated pzc of 0.44 V versus SHE is obtained. In Fig. 9, still an overall drift in the work function is observable. Yet, running the
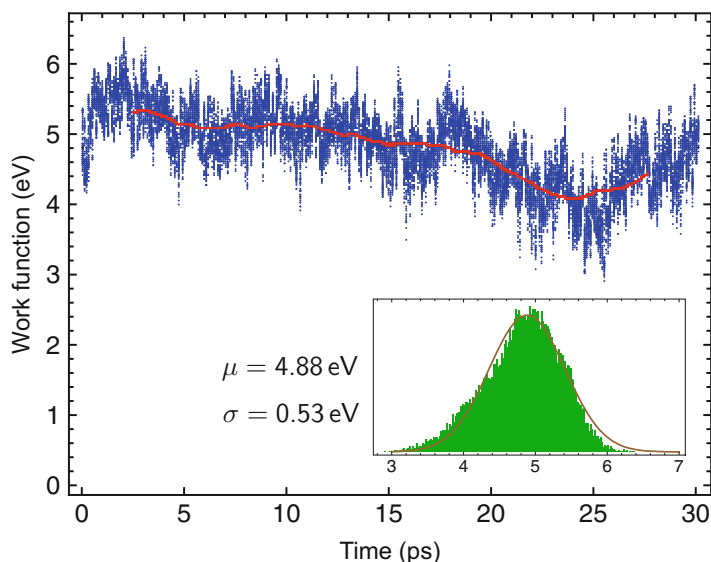
Fig. 9 Calculated work function of the water-covered Pt(111) slab along an AIMD trajectory of six explicit water layers on Pt(111) (Sakong et al. 2016). The red line corresponds to values each averaged over 5 ps. The distribution in the work function is shown together with a Gaussian fit that is presented in the inset. (Courtesy of Dr. Sung Sakong)

simulations for further 10 ps does not influence the results significantly (Sakong et al. 2016).

The AIMD simulations can also be used to analyze the water structure as a function of the distance from the Pt(111) electrode (Sakong et al. 2016). The distributions of the oxygen and the hydrogen atoms of the water molecules in the AIMD simulations are shown in Fig. 10a, b, respectively. In the middle of the water film, the mean spatial distribution becomes already rather flat as in bulk water. This suggests that the water film is thick enough to include a region that is bulk water-like and that water above Pt(111) becomes liquid-like from the third water layer on.

Up to now we only considered clean metal electrodes. However, as shown in Sect. 2, metal electrodes such as Pt(111) are typically covered by hydrogen at low electrode potentials and by anions at high electrode potentials. This will also have a consequence as far as the water structure at these electrode-electrolyte interfaces is concerned. AIMD simulations of water layers on hydrogen-covered Pt(111) (Roman and Groß 2013) have shown that the presence of hydrogen on Pt(111) weakens the interaction between water and Pt. As a consequence, the lowest water layer is about 1 Å farther away from the surface than in the case of the clean electrode. Thus one could say that the adsorption of hydrogen on Pt(111) leads to a passivation of the Pt-electrode.

Furthermore, the weakening of the water-metal interaction also leads to a strengthening of the water-water interaction which can be explained invoking the

Fig. 10 Averaged
distribution of the oxygen
atoms (**a**) and hydrogen
atoms (**b**) of water molecules
along an AIMD simulation of
six explicit water layers on
Pt(111) for more than
30 ps (Sakong et al. 2016).
The blue short-dashed lines
show a deconvolution of the
distributions in terms of
Gaussian functions, and the
long-dashed black lines
correspond to integrated
distributions. (Courtesy of Dr.
Sung Sakong)

(a) Distribution of oxygen atoms

Gaussian
peaks
*p*1: 12.4
*p*2: 15.3
*p*3: 18.9
*p*4: 21.7
*p*5: 24.9
*p*6: 28.5
*p*7: 29.2

(b) Distribution of hydrogen atoms

Gaussian
peaks
*p*1: 12.0
*p*2: 13.8
*p*3: 15.7
*p*4: 18.4
*p*5: 21.9
*p*6: 24.8
*p*7: 28.5
*p*8: 28.7
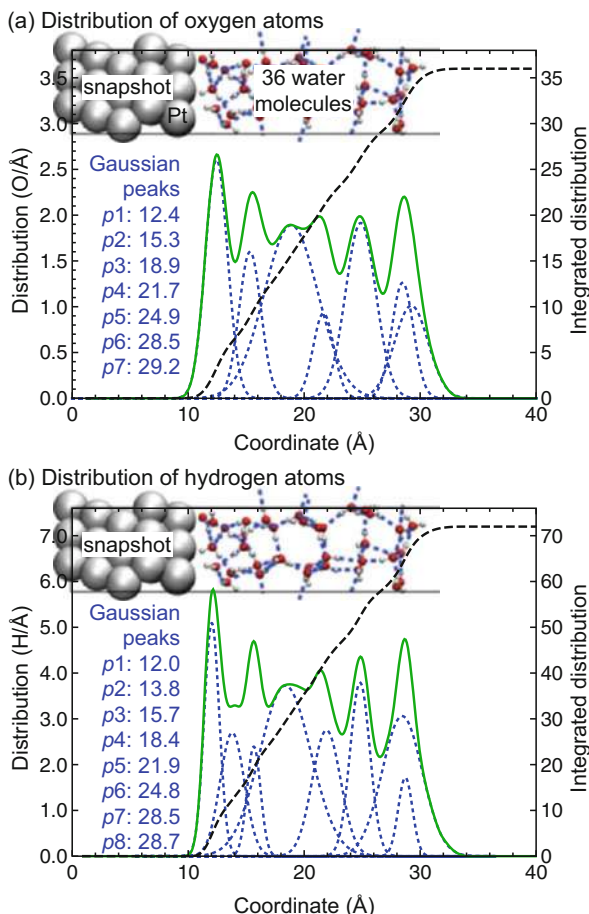
concept of bond-order conservation (Groß et al. 2014). This leads to a more
structured distribution of the water molecules in the first two layers above the
hydrogen-covered Pt(111) electrode (Roman and Groß 2013).

Besides metal electrodes, oxide electrodes are also of strong interest in electro-
chemistry and electrocatalysis, in particular as catalysts for the oxygen evolution
reaction (Halck et al. 2014; Mom et al. 2014). However, in contrast to metal
electrodes, oxide electrodes often have a more complex surface structure due to
the covalent nature of the bonding (Groß 2009). In particular, many oxide surfaces
become hydroxylated in the presence of water which then has a significant influence
on the bonding of water to these electrodes. In a DFT study, it was found that out-of-
plane hydroxyl groups act as H-bond donors, whereas in-plane hydroxyl groups are
H-bond acceptors (Gaigeot et al. 2012) and thus define a patterning of the interfacial
water structure.
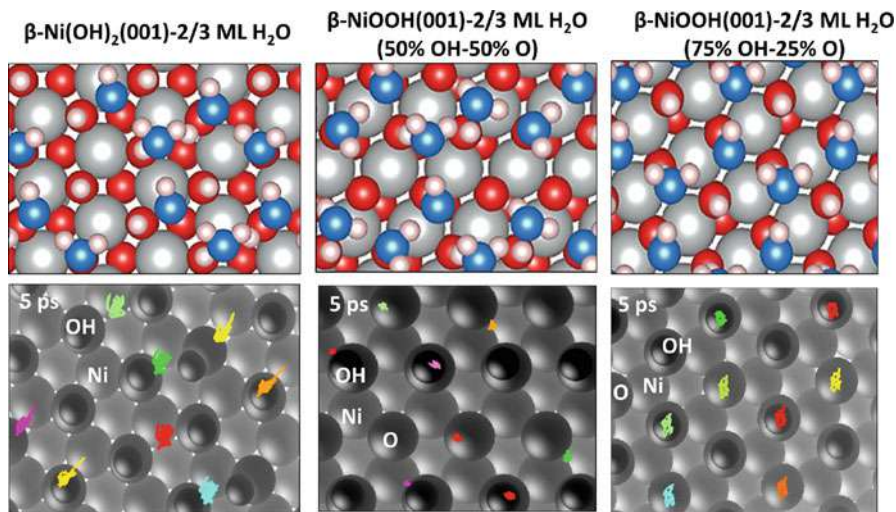
Fig. 11 Side and top views of the structure of water bilayer taken from the AIMD simulations of water at 140 K on β-Ni(OH)$_2$(001), β-NiOOH(001) terminated with 50% OH and 50% O, and β-NiOOH(001) terminated with 75% OH and 25% O. Lower panel: trajectories of the oxygen atoms of the water molecules along an AIMD run at 140 K. The water coverages correspond to 2/3 ML. (Reproduced from Eslamibidgoli et al. (2017) with permission from the PCCP Owner Societies)

In another DFT study, the adsorption of water layers on nickel(oxy)hydroxides was addressed (Eslamibidgoli et al. 2017). Figure 11 shows the top and side views of water bilayers at a coverage of 2/3 monolayers (ML). On β-Ni(OH)$_2$(001) and β-NiOOH$_{75/25}$(001), i.e., a surface terminated with 75% OH and 25%O, hexagonal water layers form. In contrast, on β-NiOOH$_{50/50}$(001), only an incomplete hydrogen bond network is stable (Eslamibidgoli et al. 2017). However, the lattice constant of β-Ni(OH)$_2$(001) is about 7.5% larger than the one of the β-NiOOH(001) surfaces, resulting in a weaker water-water interaction on β-Ni(OH)$_2$(001). In addition, on β-Ni(OH)$_2$(001) also the surface-water interactions are weaker as due to the surface structure, no favorable adsorption configuration with the water molecule bound through the oxygen to the surface can be realized. Consequently, a temperature of 140 K leads to a disordered structure, as the lower panel of Fig. 11 demonstrates.

As shown above, for selected systems the statistical averages necessary to determine the properties of systems including liquid electrolytes can nowadays be obtained from first principles using AIMD simulations. Still, in spite of the development of efficient algorithms and the ongoing progress in computer power, AIMD simulations are numerically rather costly. This represents a severe obstacle for the first-principles modeling of electrode-electrolyte interfaces, in particular, when many different configurations are needed to be calculated, as, for example, in the description of electrocatalytic reaction networks.

As a computationally attractive alternative, the electrolyte can be represented in an implicit solvent model (Mathew et al. 2014; Andreussi et al. 2012; Andreussi and Marzari 2014; Sakong et al. 2015), i.e., the solvent is treated as a continuous dielectricum. This approach is justified when the averaged behavior of many highly dynamic solvent molecules can be approximated by a potential of mean force (Leach 2001), but critical issues are a suitable short-range cavity parameterization and an appropriate treatment of the nonlinearity in the ionic response (Sundararaman and Schwarz 2017). Still, qualitative trends can be well reproduced by using implicit solvent models, as was recently shown for the methanol electrooxidation on Pt(111) (Sakong and Groß 2016, 2017).

Experimentally, it was shown that on a Pt(111) electrode in an electrochemical environment, methanol is initially predominantly converted to hydroxymethyl (H_2COH) involving the breaking of one of the C-H bonds of the methyl group of methanol (Franaszczuk et al. 1992). In contrast, at the Pt-vacuum interface, methanol also decays to methoxy (CH_3O) associated with the breaking of the O-H bond (Franaszczuk et al. 1992). Calculations with and without the presence of an implicit solvent show that the presence of the implicit solvent favors isomers with a hydrophilic hydroxyl group. This also increases the barrier for the O-H bond breaking in the dehydrogenation from methanol to methoxy in the presence of the solvent as the molecule loses its hydrophilic hydroxyl group along the reaction path (see Fig. 12). Thus the selectivity toward the initial C-H bond breaking with regard to the initial O-H bond breaking is increased by three orders of magnitude in the
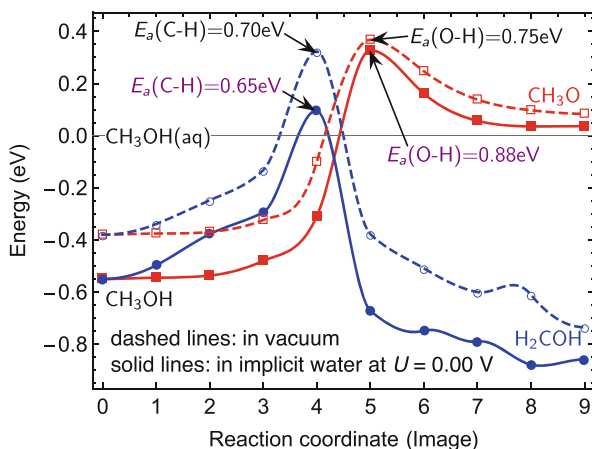
Fig. 12 Energy minimum curves of the initial methanol dehydrogenation on Pt(111) in vacuum (dashed lines) and in implicit water (solid lines) along a reaction path coordinate. The blue lines represent the formation of H_2COH involving the breaking of a C-H bond, while the red lines correspond to the creation of CH_3O involving the breaking of an O-H bond. (Reprinted with permission from Sakong and Groß (2016) Sung Sakong and Axel Groß, The importance of the electrochemical environment in the electro-oxidation of methanol on Pt(111), ACS Catal. 2016, 6, 5575. Copyright 2016 American Chemical Society)

presence of a solvent (Sakong and Groß 2016, 2017), confirming the experimental findings (Franaszczuk et al. 1992).

5 Explicit Consideration of Varying Electrode Potentials

It was already shown in Sect. 2 how varying electrode potentials can be implicitly included in the grand-canonical description of electrode-electrolyte interfaces through the dependence of the electrochemical potentials on the electrode potential. However, fundamentally this is not a satisfactory approach as typically the explicit dependence of the properties on the electrode potential is not taken into account.

Let us have a look at how electrochemical experiments are performed which is illustrated in Fig. 13a. Typically the electrode potential of a working electrode W with respect to a reference electrode R is specified. When processes occur at the working electrode that would change the charge distribution at the interface and thus the specific electrode potential, a charge is flowing between working and reference electrode that keeps the electrode potential constant. This leads to different excess charges on the electrodes at a given constant potential which for slabs of finite thickness in a periodic DFT, ansatz can be achieved by varying the number of electrons and keeping the chemical potential μ of the electrons constant, i.e., the grand-canonical formulation of DFT (Mermin 1965) needs to be invoked. This requires a self-consistent field procedure in which the number of electrons is allowed to vary such that the Fermi level stays at a preset value (Lozovoi et al. 2001). However, a straightforward implementation of a feedback loop leads to oscillatory instabilities (Bonnet et al. 2012; Schneider and Auer 2014).
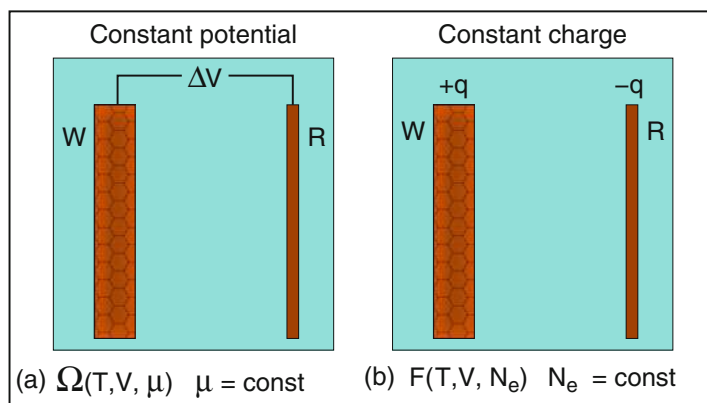
Fig. 13 Illustration of (**a**) the constant chemical potential ($\mu = const$) and (**b**) the constant charge ($N_e = const$) mode to treat electrochemical cells within periodic DFT calculations. The corresponding thermodynamic potentials to describe a slab together with a reference electrode are the grand potential Ω and the Helmholtz free energy F, respectively

These problems can be avoided by connecting the system to a fictitious potentiostat (Bonnet et al. 2012). The potentiostat functions analogously to a thermostat (Nosé 1984). The total electronic charge of the system becomes a dynamical variable through the exchange of electronic charge with an external potentiostat at a specified potential. This leads to dynamical equations for the total electronic charge such that the calculated trajectory samples the grand-canonical distribution at a given electrode potential.

Another constant potential scheme was developed for a nonperiodic finite DFT setup (Schneider and Auer 2014) which avoids the introduction of compensating countercharges. Still, a straightforward implementation of a grand-canonical DFT scheme leads to numerical instabilities also in this finite setup so that a numerical evaluation of $d\mu/dN$ is employed in order to estimate the correct number of electrons N.

It is fair to say that both of these constant potential implementations have so far not been routinely employed to describe structures and processes at electrochemical solid-electrolyte interfaces, probably because of their high numerical demand. In fact, most of the first-principles studies performed for different electrode potentials so far have been performed in the constant charge mode depicted in Fig. 13b with the corresponding electrode potential derived in an a posteriori fashion.

In contrast to calculations for a finite setup, in a periodic setup, only charge-neutral supercells can be employed because otherwise the electrostatic energy diverges. Hence any excess charge on an electrode has to be compensated by an oppositely charged counter electrode. This counter electrode can, for example, be realized, as illustrated in Fig. 14, by a localized planar charge distribution (Fu and Ho 1989; Bohnen and Kolb 1998; Lozovoi and Alavi 2003; Che and Chan 2003; Zhao et al. 2007; Schnur and Groß 2011) in the form of a Gaussian profile perpendicular to the surface:

$$\rho_{ce}(\mathbf{r}) = \frac{q}{\sqrt{2\pi}\alpha} \exp\left\{ \left(-\frac{z - z_0}{2\alpha} \right)^2 \right\}, \tag{11}$$

where q is the total charge of the counter electrode and z_0 corresponds to the position of the counter electrode.

In Fig. 14, laterally averaged one-electron potentials are plotted for different excess electron numbers per unit cell for a symmetric slab covered by one water layer on each side. Variations in the electron number cause a shift in the Fermi energy whose range is indicated $\Delta\varepsilon_F$. The counter electrode has to be positively charged for negatively charged slabs which in fact makes it attractive for electrons. This causes artificial electron transfer from the slabs to the counter electrode when the one-electron potential in the counter electrode drops below the Fermi energy for a sufficiently large positive countercharge. Consequently the range toward negative charges on the slabs corresponding to low electrode potentials is limited.

Note that a countercharge needs to be introduced in order to compensate the excess charges on the electrodes, but it has no physical relevance. Nevertheless there
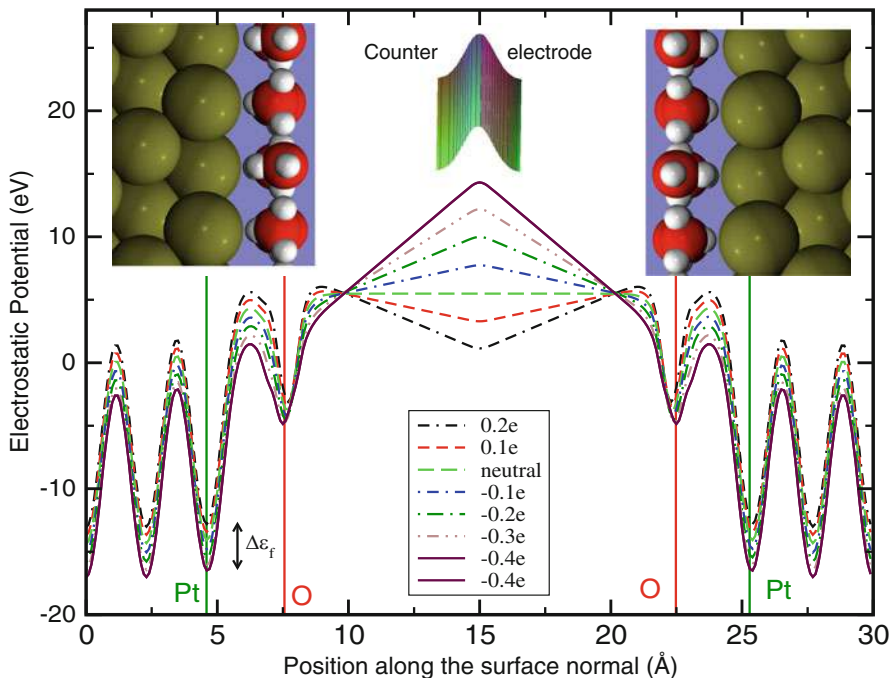
Fig. 14 Illustration of the implementation of an explicit Gaussian-shaped counter electrode in a periodic DFT setup. The laterally averaged one-electron potential of a symmetrically constructed metal-water slab along the surface normal is plotted for various charges of the slab (Schnur and Groß 2011). (Reprinted and adapted from Schnur and Groß (2011), with permission from Elsevier)

is an additional direct electrostatic interaction between the electrons and the ion cores with the counter electrode which needs to be corrected for by integrating the electrochemical potential μ over the applied charge (Taylor et al. 2006), i.e.:

$$E = \int_0^q \mu \, dQ = E_{\text{DFT}} + \int_0^q \left[\int \frac{V_{\text{tot}}}{\Omega} d^3 x \right] dQ. \tag{12}$$

Furthermore, the total grand-canonical free energy of the electrons has to be expressed as

$$E_{\text{free}} = E + \mu q, \tag{13}$$

in order to take the varying number of electrons in each system into account. Here E is the total energy given by Eq. 12, i.e., the energy E_{DFT} from the DFT calculations corrected for the electrostatic interaction with the counter electrode, and q is the total charge of the explicit electron-ion system.

In order to derive the electrode potential related to a certain charge state of the electrode, in principle the corresponding work function needs to be evalu-

ated (Lozovoi et al. 2001; Taylor et al. 2006; Schnur and Groß 2011). This requires to determine the vacuum level of the one-electron energy. However, as Fig. 14 demonstrates, for $q \neq 0$ there is no field-free region in the supercell, so that the work function and hence the electrode potential cannot be specified. One might relate the charge to the electrode potential via experimentally derived capacities (Bohnen and Kolb 1998); however, this introduces an empirical component to the approach.

Still, following the procedure of Lozovoi et al. (2001), the work function can be derived by determining the one-electron potential v at a point z' where the potential is linear using:

$$\Phi = v\left(z'\right) - (z' - z_{\text{ref}}) \left.\frac{dv}{dz}\right|_{z=z'} - \varepsilon_F, \qquad (14)$$

where z_{ref} is a point where the vacuum level in the one-electron potential has been reached in the field-free case. As Fig. 14 illustrates, there is a very convenient reference point z_{ref} for this particular system where all one-electron potentials cross and where the electrode potential in the neutral case already assumes its vacuum level. Choosing $z' = z_{\text{ref}}$, the work function can be derived using

$$\Phi = v\left(z_{\text{ref}}\right) - \varepsilon_F \qquad (15)$$

In fact, defining the counter electrode as a localized planar distribution still requires to modify the electronic structure codes. However, when no explicit counter electrode is implemented, then upon charging the system, i.e., when the number of electrons does not equal the total positive charge of all ion cores, the excess charge is automatically compensated by a uniform charge background. This is because in the Fourier transform of the electrostatic energy, the monopole term is omitted as it would lead to a diverging energy. This approach has in fact also been used to realize the counter electrode in periodic DFT calculations, addressing electrochemical electrode-electrode interfaces (Filhol and Neurock 2006; Taylor et al. 2006). In this method it is even more complex to derive the electrode potential a posteriori because in the vacuum region, the one electrode potential exhibits a quadratic behavior for charged situations (Taylor et al. 2006). This also leads to erroneous field effects at the electrode surface. They do not appear when the vacuum region is filled with explicit water molecules (Taylor et al. 2006), but still no work function can directly be derived.

As a solution of this problem, the so-called double-reference method was developed (Filhol and Neurock 2006; Taylor et al. 2006). In a first step, a DFT calculation with a well-defined vacuum region is performed by constructing a solvated slab with a vacuum region introduced in the middle of the unit cell between the slabs. Note that electric fields do not penetrate into perfect metallic conductors. Hence the potential inside of the metal slab representing the metal electrode taken with respect to the vacuum level is taken as the first reference point.

Then a second set of calculations for varying charges q is performed in which the regions between the metal slabs are filled with explicit water. The water

configuration in a region far from the electrode is fixed in its $q = 0$ arrangement, and the one-electron potential $\phi_0(w)$ at this site is used as the second reference point. For $q \neq 0$, the remaining system is relaxed under the influence of the excess charge, and the potential at all other positions is shifted with respect to the second reference point. The absolute electrode potential is then taken, for example, with respect to the standard hydrogen electrode (SHE).

We will use Fig. 15 to illustrate how results as a function of the electrode potential can be derived from constant charge calculations. In Fig. 15a, results with respect to the oxygen dissociation on Pt(111) obtained with the double-reference method are shown (Wasileski and Janik 2008). As a first step, initial, transition, and final state of the O_2 dissociation on Pt(111) in the explicit presence of water molecules were calculated for five different charge states ranging from $-1e$ to $+1e$ per supercell. For each configuration, the corresponding electrode potential was determined a posteriori, and the energies were plotted as a function of the electrode potential. Points that are roughly above each other have been obtained for the same charge. In fact, if the electrode potential had been constant along the dissociation path for each particular charge, the points would have been exactly above each other. The variation shows how much the electrode potential varies along the reaction path at a given charge.

The dissociation barrier is given by the difference between the transition state and the initial state, which corresponds to the O_2 molecule adsorbed on Pt(111). This difference has to be taken at the same potential. As the points are not exactly above each other, the energies as a function of the electrode potential had to be interpolated, in this case by a quadratic spline. Then for any given electrode potential, the dissociation barrier can be read off from the quadratic fits. These barriers are plotted for constant potential (μ constant) as the red symbols in Fig. 15b. Not only results for pure water but also with co-adsorbed Na atoms are shown.

In addition, the barriers are also plotted derived from the differences in the energies at constant charge (q constant). For low electrode potentials, where the corresponding points in Fig. 15a are almost directly aligned above each other, there is not a significant difference between the barriers at constant charge and at constant potential. However, for positive electrode potentials, the corresponding points are no longer above each other, and the difference between the barriers obtained at constant charge and at constant potential becomes larger than 0.2 eV, which is in fact a consequence of the small unit cells that have to be used in periodic DFT calculations. Thus the change of the dipole distribution along a reaction path causes significant changes in the work function which then alter the corresponding electrode potential. If for a large surface unit cell the studied reaction takes place only on a small portion of the surface unit cell, then the associated work function change, which corresponds to an average over the whole surface unit cell, will only be small, even if locally the dipole moments change substantially.

In the discussion so far, we only had focused on electronic effects upon changing the electrode potential. However, when the electrode potential changes, also the electric double layer will rearrange, water molecules might adjust their preferential orientation, and the density of solvated ions in the double layer can be altered. Hence
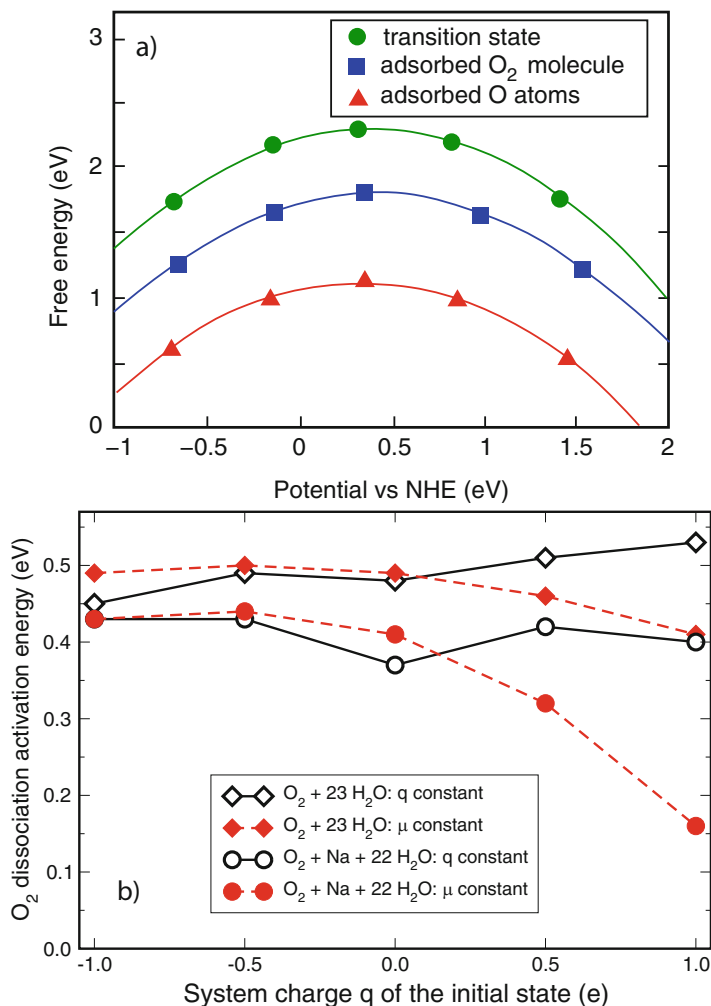
Fig. 15 (**a**) Calculated free energies as a function of potential for the initial, transition, and final state of O_2 dissociation on solvated Pt(111) obtained with the double-reference method (Taylor et al. 2006). The symbols correspond to the results obtained for different charge states of the Pt electrode, whereas the solid curves are quadratic fits to the results. (**b**) Dissociation barrier of O_2 on solvated Pt(111) without and with Na co-adsorption, respectively, for various constant system charges q of the system and for constant potential, kept at the corresponding value of the initial state. (Adapted from Wasileski and Janik 2008)

in fact the change of the electrode potential can also be represented by a change of the concentration of the ions at the electrode surface which modifies the dipole distribution and consequently the work function and also the electrode potential.

These considerations have been the basis for the model illustrated in Fig. 16 (Skúlason et al. 2007; Rossmeisl et al. 2008). The electrode potential is
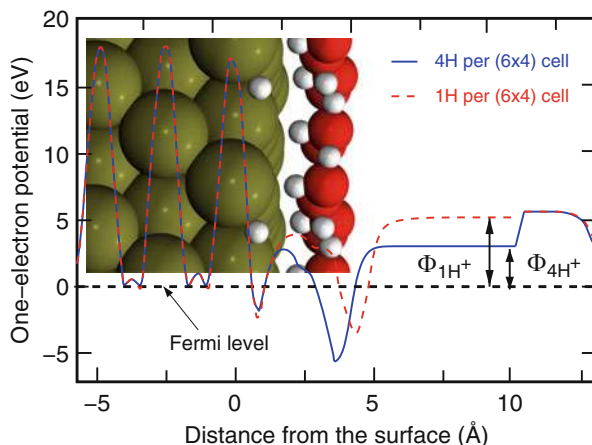
Fig. 16 Illustration of how the work function and thus the electrode potential is changed by varying the number of counterions at the electrode-electrolyte interface. The one-electron potential is changed by varying the number of hydrogen atoms in the water layer at Pt(111) within a (6×4) geometry. In the vacuum layer, there is a potential drop due to the presence of a dipole layer. (Adapted from Skúlason et al. 2007). The inset illustrates the structure of the interface with additional protons

varied by introducing hydrogen atoms in the water layer close to the metal electrode. The added hydrogen atoms change the work function of the electrode-electrolyte system and thus the corresponding electrode potential, as demonstrated for one and four hydrogen atoms per (6×4) surface unit cell in Fig. 16. Within this approach, the whole unit cell remains neutral so that the introduction of a counter electrode is not needed.

The hydrogen evolution reaction on Pt(111) has been addressed using this method (Skúlason et al. 2007). As only integer numbers of hydrogen atoms can be added, the electrode potential cannot be changed continuously; however, an extrapolation method to infinite unit cell sizes has been developed which fixed this problem (Skúlason et al. 2007). As far as the calculation of reaction barriers is concerned, this ansatz faces the same problem as constant charge simulations, namely, that the configurations along the reaction path correspond to different dipole moments and thus to different electrode potentials.

However, in principle also calculations in the constant potential mode should take into account that the work function cannot only be adjusted by changing the number of electrons in the systems but also by rearranging the structure of the electrical double layer. Hence any realistic method addressing electrochemical electrode-electrolyte interfaces for varying electrode potentials should take into account both changes in the excess charge of the electrode and the structural rearrangement of the electrolyte within the double layer.

The fact that every surface configuration along a AIMD run in the constant charge mode corresponds in general to a different electrode potential can actually

be used as an advantage. By running long trajectories and monitoring the work function, different configurations can be binned according to their work function. This can also be done for different numbers of ions and adsorbates, thus creating a grand-canonical ensemble of states for a certain electrode potential range (Rossmeisl et al. 2013; Hansen et al. 2017). This so-called generalized computational hydrogen electrode approach was, for example, used to derive atomic-scale quantities of the Au(111)/water interface (Hansen and Rossmeisl 2016). However, this method still suffers from the problem that the sampling is rather inefficient.

6   Conclusions

There is a strong focus on electrochemical electrode-electrolyte interfaces at the moment as they are crucial for our understanding of structures and processes in electrochemical energy conversion and storage. Still, the theoretical description of these interfaces is quite challenging and demanding, as described in this chapter, because, in particular, for liquid electrolytes a sufficient statistical sampling is required, and furthermore a proper treatment of the effect of the electrode potential is necessary. Thus in spite of the complexity of the systems, quantum chemical approaches are needed that properly take electronic structure effects into account which on the other hand make numerical studies computationally very demanding.

There have been many theoretical and numerical efforts to describe electrode-electrolyte interfaces, but still it is fair to say that this field is far from being mature. There is, for example, no complete atomistic picture of the electric double layer even for simple systems. As far as the theoretical methods are concerned, there is no common approach that is widely used except for the grand-canonical ansatz within the concept of the computational hydrogen electrode. Applications of this concept typically use rather severe approximations, but the reliability of these approximations cannot easily be assessed as there are hardly any true benchmark calculations. There are attempts to perform quantum chemical electronic structure calculations of electrode-electrolyte interfaces in the constant potential mode, but obviously there is a certain disillusion with respect to these methods as there are no widespread applications of them.

In spite of these reservations, there has recently been significant progress in the theoretical description of solid-electrolyte interfaces, and it is now clear that the electrochemical environment has to be properly taken into account in theoretical and numerical studies, as illustrated in this contribution. Researchers are aware of all the pitfalls of present approaches and are actively pursuing strategies to overcome the limitations of the methods. Progress in this field is certainly required because of the technological importance of electrochemical energy conversion and storage for our future sustainable energy supply and because of severe problems that still need to be overcome. In addition, there are many interesting scientific problems on a fundamental level that need to be addressed.

The still incomplete status of our knowledge about the microscopic nature of structures and processes at the electrode-electrolyte interface and in the double

layer on a quantitative level makes this research field rather demanding, sometimes frustrating, but at the same time it also makes it exciting and rewarding with lots of opportunity for further progress, both as far as methodological development and applications of numerically efficient schemes that will deepen our knowledge about electrode-electrolyte interfaces are concerned. Hence this is certainly still a promising and emergent field that offers great opportunities.

References

Andreussi O, Marzari N (2014) Electrostatics of solvated systems in periodic boundary conditions. Phys Rev B 90:245101. https://doi.org/10.1103/PhysRevB.90.245101

Andreussi O, Dabo I, Marzari N (2012) Revised self-consistent continuum solvation in electronic-structure calculations. J Chem Phys 136:064102. https://doi.org/10.1063/1.3676407

Behler J (2014) Representing potential energy surfaces by high-dimensional neural network potentials. J Phys Condens Matter 26:183001

Bohnen KP, Kolb DM (1998) Charge- versus adsorbate-induced lifting of the Au(100)-(hex) reconstruction in an electrochemical environment. Surf Sci 407:L629–L632. https://doi.org/10.1016/S0039-6028(98)00232-5

Bonnet N, Morishita T, Sugino O, Otani M (2012) First-principles molecular dynamics at a constant electrode potential. Phys Rev Lett 109:266101. https://doi.org/10.1103/PhysRevLett.109.266101

Bonthuis DJ, Netz RR (2013) Beyond the continuum: how molecular solvent structure affects electrostatics and hydrodynamics at solid–electrolyte interfaces. J Phys Chem B 117:11397–11413. https://doi.org/10.1021/jp402482q

Bukowski R, Szalewicz K, Groenenboom GC, van der Avoird A (2007) Predictions of the properties of water from first principles. Science 315:1249

Carrasco J, Santra B, Klimeš J, Michaelides A (2011) To wet or not to wet? Dispersion forces tip the balance for water ice on metals. Phys Rev Lett 106:026101. https://doi.org/10.1103/PhysRevLett.106.026101

Carrasco J, Klimeš J, Michaelides A (2013) The role of van der Waals forces in water adsorption on metals. J Chem Phys 138:024708. https://doi.org/10.1063/1.4773901

Chapman DL (1913) A contribution to the theory of electrocapillarity. PhilMag 25:475

Che JG, Chan CT (2003) Charge transfer, surface charging, and overlayer-induced faceting. Phys Rev B 67:125411. https://doi.org/10.1103/PhysRevB.67.125411

Chen B, Ivanov I, Klein ML, Parrinello M (2003) Hydrogen bonding in water. Phys Rev Lett 91:215503. https://doi.org/10.1103/PhysRevLett.91.215503

Cheng J, Sprik M (2014) The electric double layer at a rutile TiO_2 water interface modelled using density functional theory based molecular dynamics simulation. J Phys Condens Matter 26:244108

Crozier PS, Rowley RL, Henderson D (2000) Molecular dynamics calculations of the electrochemical properties of electrolyte systems between charged electrodes. J Chem Phys 113:9202

Cuesta A (2004) Measurement of the surface charge density of CO-saturated Pt(111) electrodes as a function of potential: the potential of zero charge of Pt(111). Surf Sci 572:11–22. https://doi.org/10.1016/j.susc.2004.08.014

Daw MS, Baskes MI (1984) Embedded-atom method: derivation and applications to impurities, surfaces, and other defects in metals. Phys Rev B 29:6443

Daw MS, Foiles SM, Baskes MI (1993) The embedded-atom method – a review of theory and applications. Mater Sci Rep 9:252

Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density functional for general geometries. Phys Rev Lett 92:246401

Do H, Besley NA (2012) Structural optimization of molecular clusters with density functional theory combined with basin hopping. J Chem Phys 137:134106. https://doi.org/10.1063/1.4755994

Eslamibidgoli MJ, Groß A, Eikerling M (2017) Surface configuration and wettability of nickel(oxy)hydroxides: a first-principles investigation. Phys Chem Chem Phys 19:22659–22669. https://doi.org/10.1039/C7CP03396F

Fantauzzi D, Bandlow J, Sabo L, Mueller JE, van Duin ACT, Jacob T (2014) Development of a reaxff potential for Pt-O systems describing the energetics and dynamics of Pt-Oxide formation. Phys Chem Chem Phys 16:23118–23133. https://doi.org/10.1039/C4CP03111C

Feibelman PJ (2002) Partial dissociation of water on Ru(0001). Science 295:99

Fernández-Serra MV, Artacho E (2004) Network equilibration and first-principles liquid water. J Chem Phys 121:11136–11144. https://doi.org/10.1063/1.1813431

Filhol JS, Bocquet ML (2007) Charge control of the water monolayer/pd interface. Chem Phys Lett 438:203

Filhol JS, Neurock M (2006) Elucidation of the electrochemical activation of water over pd by first principles. Angew Chem Int Ed 45:402

Fischer JM, Mahlberg D, Roman T, Groß A (2016) Water adsorption on bimetallic PtRu/Pt(111) surface alloys. Proc R Soc A 472:20160618. https://doi.org/10.1098/rspa.2016.0618

Forster-Tonigold K, Groß A (2014) Dispersion corrected RPBE studies of liquid water. J Chem Phys 141:064501. https://doi.org/10.1063/1.4892400

Franaszczuk K, Herrero E, Zelenay P, Wieckowski A, Wang J, Masel RI (1992) A comparison of electrochemical and gas-phase decomposition of methanol on platinum surfaces. J Phys Chem 96:8509–8516. https://doi.org/10.1021/j100200a056

Fritsch S, Potestio R, Donadio D, Kremer K (2014) Nuclear quantum effects in water: a multiscale study. J Chem Theory Comput 10:816–824. https://doi.org/10.1021/ct4010504

Fu CL, Ho KM (1989) External-charge-induced surface reconstruction on Ag(110). Phys Rev Lett 63:1617–1620. https://doi.org/10.1103/PhysRevLett.63.1617

Gaigeot MP, Sprik M, Sulpizi M (2012) Oxide/water interfaces: how the surface chemistry modifies interfacial water properties. J Phys Condens Matter 24:124106

Garcia-Araez N, Koper MTM (2010) A sublattice-model isotherm for the competitive coadsorption of hydrogen and bromide on a Pt(100) electrode. Phys Chem Chem Phys 12:143–148. https://doi.org/10.1039/B912091B

Garcia-Araez N, Climent V, Herrero E, Feliu J, Lipkowski J (2005a) Thermodynamic studies of chloride adsorption at the pt(111) electrode surface from 0.1 m {HClO4} solution. J Electroanal Chem 576:33–41. https://doi.org/10.1016/j.jelechem.2004.10.003

Garcia-Araez N, Climent V, Herrero E, Feliu JM, Lipkowski J (2005b) Determination of the Gibbs excess of H adsorbed at a Pt(111) electrode surface in the presence of co-adsorbed chloride. J Electroanal Chem 582:76–84. https://doi.org/10.1016/j.jelechem.2005.01.031

Garcia-Araez N, Climent V, Herrero E, Feliu J, Lipkowski J (2006) Thermodynamic studies of bromide adsorption at the Pt(111) electrode surface perchloric acid solutions: comparison with other anions. J Electroanal Chem 591:149–158. https://doi.org/10.1016/j.jelechem.2006.04.008

Goddard WA III, Merinov BV, van Duin ACT, Jacob T, Blanco M, Molinero V, Jang SS, Jang YH (2003) Multi-paradigm multi-scale simulations for fuel cell catalysts. Mol Simul 32:251

Gohda Y, Schnur S, Groß A (2008) Influence of water on elementary reaction steps in electrocatalysis. Faraday Discuss 140:233–244

Gossenberger F, Roman T, Forster-Tonigold K, Groß A (2014) Work function change of platinum electrodes induced by halide adsorption. Beilstein J Nanotechnol 5:152–161

Gossenberger F, Roman T, Groß A (2015) Equilibrium coverage of halides on metal electrodes. Surf Sci 631:17–22. https://doi.org/10.1016/j.susc.2014.01.021

Gossenberger F, Roman T, Groß A (2016) Hydrogen and halide co-adsorption on Pt(111) in an electrochemical environment: a computational perspective. Electrochim Acta 216:152–159. https://doi.org/10.1016/j.electacta.2016.08.117

Gouy LG (1910) Sur la constitution de la charge électrique à la surface d'un électrolyte. J Phys Theor Appl 9:457

Grimme S (2004) Accurate description of van der waals complexes by density functional theory including empirical corrections. J Comput Chem 25:1463

Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132:154104

Groß A (2002) The virtual chemistry lab for reactions at surfaces: is it possible? Will it be helpful? Surf Sci 500:347

Groß A (2009) Theoretical surface science – a microscopic perspective, 2nd edn. Springer, Berlin

Groß A, Gossenberger F, Lin X, Naderian M, Sakong S, Roman T (2014) Water structures at metal electrodes studied by ab initio molecular dynamics simulations. J Electrochem Soc 161:E3015–E3020. https://doi.org/10.1149/2.003408jes

Habershon S, Markland TE, Manolopoulos DE (2009) Competing quantum effects in the dynamics of a flexible water model. J Chem Phys 131:024501. https://doi.org/10.1063/1.3167790

Hafner J, Wolverton C, Ceder G (2006) Toward computational material design: the impact of density functional theory on materials research. MRS Bull 31:659

Halck NB, Petrykin V, Krtil P, Rossmeisl J (2014) Beyond the volcano limitations in electrocatalysis – oxygen evolution reaction. Phys Chem Chem Phys 16:13682–13688. https://doi.org/10.1039/C4CP00571F

Hammer B, Hansen LB, Nørskov JK (1999) Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. Phys Rev B 59:7413

Hansen HA, Man IC, Studt F, Abild-Pedersen F, Bligaard T, Rossmeisl J (2010) Electrochemical chlorine evolution at rutile oxide (110) surfaces. Phys Chem Chem Phys 12:283–290. https://doi.org/10.1039/B917459A

Hansen MH, Rossmeisl J (2016) pH in grand canonical statistics of an electrochemical interface. J Phys Chem C 120:29135–29143. https://doi.org/10.1021/acs.jpcc.6b09019

Hansen MH, Nilsson A, Rossmeisl J (2017) Modelling pH and potential in dynamic structures of the water/Pt(111) interface on the atomic scale. Phys Chem Chem Phys 19:23505–23514. https://doi.org/10.1039/C7CP03576D

Helmholtz H (1853) Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. Pogg Ann 89:211

Heras JM, Viscido L (1980) Work function changes upon water contamination of metal surfaces. Appl Surf Sci 4:238–241. https://doi.org/10.1016/0378-5963(80)90133-6

Hoffmann W, Benndorf C (1997) Investigations on the influence of substrate geometry of flat and stepped ruthenium surfaces Ru(0001) and Ru(10$\bar{1}$8) on the adsorption kinetics of H$_2$O and D$_2$O. Surf Sci 377–379:681–686. https://doi.org/10.1016/S0039-6028(96)01468-9

Imoto S, Forbert H, Marx D (2015) Water structure and solvation of osmolytes at high hydrostatic pressure: pure water and TMAO solutions at 10 kbar versus 1 bar. Phys Chem Chem Phys 17:24224–24237. https://doi.org/10.1039/C5CP03069B

Jinnouchi R, Kodama K, Morimoto Y (2014) DFT calculations on H, OH and O adsorbate formations on Pt(111) and Pt(332) electrodes. J Electroanal Chem 716:31–44. https://doi.org/10.1016/j.jelechem.2013.09.031

Keith JA, Jerkiewicz G, Jacob T (2010) Theoretical investigations of the oxygen reduction reaction on Pt(111). Chem Phys Chem 11:2779–2794. https://doi.org/10.1002/cphc.201000286

Kiskinova M, Pirug G, Bonzel H (1985) Adsorption and decomposition of H$_2$O on a k-covered Pt(111) surface. Surf Sci 150:319–338. https://doi.org/10.1016/0039-6028(85)90649-1

Kolb DM (2002) An atomistic view of electrochemistry. Surf Sci 500:722

Langenbach E, Spitzer A, Lüth H (1984) The adsorption of water on Pt(111) studied by irreflection and UV-photoemission spectroscopy. Surf Sci 147:179–190. https://doi.org/10.1016/0039-6028(84)90174-2

Leach AR (2001) Molecular modelling: principles and applications, 2nd edn. Pearson, Harlow

Lilach Y, Romm L, Livneh T, Asscher M (2001) The first layers of water on ru(001). J Phys Chem B 105:2736–2742. https://doi.org/10.1021/jp003511g

Lin IC, Seitsonen AP, Tavernelli I, Rothlisberger U (2012) Structure and dynamics of liquid water from ab initio molecular dynamics–comparison of BLYP, PBE, and revPBE density functionals

with and without van der Waals corrections. J Chem Theory Comput 8:3902–3910. https://doi.org/10.1021/ct3001848

Lin X, Gossenberger F, Groß A (2016) Ionic adsorbate structures on metal electrodes calculated from first-principles. Ind Eng Chem Res 55(42):11107–11113. https://doi.org/10.1021/acs.iecr.6b03087

Liu LM, Krack M, Michaelides A (2009) Interfacial water: a first principles molecular dynamics study of a nanoscale water film on salt. J Chem Phys 130:234702. https://doi.org/10.1063/1.3152845

Lorenz S, Scheffler M, Groß A (2006) Descriptions of surface chemical reactions using a neural network representation. Phys Rev B 73:115431

Lozovoi AY, Alavi A (2003) Reconstruction of charged surfaces: general trends and a case study of Pt(110) and Au(110). Phys Rev B 68:245416. https://doi.org/10.1103/PhysRevB.68.245416

Lozovoi AY, Alavi A, Kohanoff J, Lynden-Bell RM (2001) Ab initio simulation of charged slabs at constant chemical potential. J Chem Phys 115:1661

Marx D, Hutter J (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Cambridge

Mathew K, Sundararaman R, Letchworth-Weaver K, Arias TA, Hennig RG (2014) Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways. J Chem Phys 140:084106. https://doi.org/10.1063/1.4865107

Mermin ND (1965) Thermal properties of inhomogeneous electron gas. Phys Rev 137:A 1441

Michaelides A (2006) Density functional theory simulations of water-metal interfaces: waltzing waters, a novel 2D ice phase, and more. Appl Phys A 85:415

Mom RV, Cheng J, Koper MTM, Sprik M (2014) Modeling the oxygen evolution reaction on metal oxides: the influence of unrestricted DFT calculations. J Phys Chem C 118:4095–4102. https://doi.org/10.1021/jp409373c

Morawietz T, Behler J (2013) A density-functional theory-based neural network potential for water clusters including van der Waals corrections. J Phys Chem A 117:7356–7366

Morawietz T, Singraber A, Dellago C, Behler J (2016) How van der Waals interactions determine the unique properties of water. Proc Natl Acad Sci 113:8368–8373

Morrone JA, Car R (2008) Nuclear quantum effects in water. Phys Rev Lett 101:017801. https://doi.org/10.1103/PhysRevLett.101.017801

Natarajan SK, Behler J (2016) Neural network molecular dynamics simulations of solid-liquid interfaces: water at low-index copper surfaces. Phys Chem Chem Phys 18:28704–28725. https://doi.org/10.1039/C6CP05711J

Nørskov JK, Rossmeisl J, Logadottir A, Lindqvist L, Kitchin JR, Bligaard T, Jónsson H (2004) Origin of the overpotential for oxygen reduction at a fuel-cell cathode. J Phys Chem B 108:17886–17892. https://doi.org/10.1021/jp047349j

Nørskov JK, Bligaard T, Logadottir A, Kitchin JR, Chen JG, Pandelov S, Stimming U (2005) Trends in the exchange current for hydrogen evolution. J Electrochem Soc 152:J23

Nosé S (1984) A unified formulation for the constant temperature molecular dynamics method. J Chem Phys 81:511

Paier J, Marsman M, Kresse G (2007) Why does the B3LYP hybrid functional fail for metals? J Chem Phys 127:024103

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865

Petrii OA (2013) Zero charge potentials of platinum metals and electron work functions (review). Russ J Electrochem 49:401–422. https://doi.org/10.1134/S1023193513050145

Raghavan K, Foster K, Motakabbir K, Berkowitz M (1991) Structure and dynamics of water at the Pt(111) interface: molecular dynamics study. J Chem Phys 94:2110–2117

Reiss H, Heller A (1985) The absolute potential of the standard hydrogen electrode: a new estimate. J Phys Chem 89:4207–4213. https://doi.org/10.1021/j100266a013

Reuter K, Scheffler M (2001) Composition, structure, and stability of RuO_2(110) as a function of oxygen pressure. Phys Rev B 65:035406

Rogal J, Reuter K, Scheffler M (2007) CO oxidation at Pd(100): a first-principles constrained thermodynamics study. Phys Rev B 75:205433. https://doi.org/10.1103/PhysRevB.75.205433

Roman T, Groß A (2013) Structure of water layers on hydrogen-covered Pt electrodes. Catal Today 202:183–190

Roman T, Gossenberger F, Forster-Tonigold K, Groß A (2014) Halide adsorption on close-packed metal electrodes. Phys Chem Chem Phys 16:13630–13634. https://doi.org/10.1039/C4CP00237G

Rossmeisl J, Nørskov JK, Taylor CD, Janik MJ, Neurock M (2006) Calculated phase diagrams for the electrochemical oxidation and reduction of water over Pt(111). J Phys Chem B 110:21833–21839. https://doi.org/10.1021/jp0631735

Rossmeisl J, Skúlason E, Björketun MJ, Tripkovic V, Nørskov JK (2008) Modeling the electrified solid-liquid interface. Chem Phys Lett 466:68

Rossmeisl J, Chan K, Ahmed R, Tripkovic V, Bjorketun ME (2013) pH in atomic scale simulations of electrochemical interfaces. Phys Chem Chem Phys 15:10321–10325. https://doi.org/10.1039/C3CP51083B

Roudgar A, Groß A (2003a) Local reactivity of metal overlayers: density functional theory calculations of Pd on Au. Phys Rev B 67:033409

Roudgar A, Groß A (2003b) Local reactivity of thin Pd overlayers on Au single crystals. J Electroanal Chem 548:121

Roudgar A, Groß A (2005a) Hydrogen adsorption energies on bimetallic overlayer systems at the solid-vacuum and the solid-liquid interface. Surf Sci 597:42

Roudgar A, Groß A (2005b) Water bilayer on the Pd/Au(111) overlayer system: coadsorption and electric field effects. Chem Phys Lett 409:157

Sakong S, Groß A (2016) The importance of the electrochemical environment in the electro-oxidation of methanol on Pt(111). ACS Catal 6:5575–5586. https://doi.org/10.1021/acscatal.6b00931

Sakong S, Groß A (2017) Methanol oxidation on Pt(111) from first-principles in heterogeneous and electrocatalysis. Electrocatal 8:577–586. https://doi.org/10.1007/s12678-017-0370-1

Sakong S, Naderian M, Mathew K, Hennig RG, Groß A (2015) Density functional theory study of the electrochemical interface between a Pt electrode and an aqueous electrolyte using an implicit solvent method. J Chem Phys 142:234107. https://doi.org/10.1063/1.4922615

Sakong S, Forster-Tonigold K, Groß A (2016) The structure of water at a Pt(111) electrode and the potential of zero charge studied from first principles. J Chem Phys 144:194701. https://doi.org/10.1063/1.4922615

Santra B, Michaelides A, Scheffler M (2007) On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters: benchmarks approaching the complete basis set limit. J Chem Phys 127:184104

Santra B, Michaelides A, Fuchs M, Tkatchenko A, Filippi C, Scheffler M (2008) On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters. II. The water hexamer and van der Waals interactions. J Chem Phys 129:194111. https://doi.org/10.1063/1.3012573

Schlögl R (2010) The role of chemistry in the energy challenge. ChemSusChem 3:209–222

Schmickler W, Santos E (2010) Interfacial electrochemistry, 2nd edn. Springer, Berlin

Schneider WB, Auer AA (2014) Constant chemical potential approach for quantum chemical calculations in electrocatalysis. Beilstein J Nanotechnol 5:668–676. https://doi.org/10.3762/bjnano.5.79

Schnur S, Groß A (2009) Properties of metal-water interfaces studied from first principles. New J Phys 11:125003

Schnur S, Groß A (2011) Challenges in the first-principles description of reactions in electrocatalysis. Catal Today 165:129–137. https://doi.org/10.1016/j.cattod.2010.11.071

Siepmann JI, Sprik M (1995) Influence of surface topology and electrostatic potential on water/electrode systems. J Chem Phys 102:511–524. https://doi.org/10.1063/1.469429

Skúlason E, Karlberg GS, Rossmeisl J, Bligaard T, Greeley J, Jónsson H, Nørskov JK (2007) Density functional theory calculations for the hydrogen evolution reaction in an electrochemical double layer on the Pt(111) electrode. Phys Chem Chem Phys 9:3241

Soper AK (2007) Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water. J Phys Condens Matter 19:335206

Spohr E (1989) Computer simulations of the water/platinum interface. J Phys Chem 93:6171

Spohr E (1997) Effect of electrostatic boundary conditions and system size on the interfacial properties of water and aqueous solutions. J Chem Phys 107:6342–6348

Stern O (1924) Zur Theorie der Elektrolytischen Doppelschicht. Z Elektrochem 30:508

Sundararaman R, Schwarz K (2017) Evaluating continuum solvation models for the electrode-electrolyte interface: challenges and strategies for improvement. J Chem Phys 146:084111. https://doi.org/10.1063/1.4976971

Taylor CD, Wasileski SA, Filhol JS, Neurock M (2006) First principles reaction modeling of the electrochemical interface: consideration and calculation of a tunable surface potential from atomic and electronic structure. Phys Rev B 73:165402

Tkatchenko A, Scheffler M (2009) Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. Phys Rev Lett 102:073005. https://doi.org/10.1103/PhysRevLett.102.073005

Todorova M, Lundgren E, Blum V, Mikkelsen A, Gray S, Gustafson J, Borg M, Rogal J, Reuter K, Andersen JN, Scheffler M (2003) The Pd(100)-$(\sqrt{5} \times \sqrt{5})R27°$-O surface oxide revisited. Surf Sci 541:101

Tonigold K, GroßA (2012) Dispersive interactions in water bilayers at metallic surfaces: a comparison of the PBE and RPBE functional including semi-empirical dispersion corrections. J Comput Chem 33:695–701. https://doi.org/10.1002/jcc.22900

Trasatti S (1983) Structuring of the solvent at metal/solution interfaces and components of the electrode potential. J Electroanal Chem 150:1–15. https://doi.org/10.1016/S0022-0728(83)80183-1

Trasatti S (1986) The absolute electrode potential: an explanatory note. Pure Appl Chem 58:955–966. https://doi.org/10.1351/pac198658070955

Trasatti S (1991) Physical electrochemistry of ceramic oxides. Electrochim Acta 36:225–241. https://doi.org/10.1016/0013-4686(91)85244-2

Trasatti S (1995) Surface science and electrochemistry: concepts and problems. Surf Sci 335:1

Tripkovic V, Björketun ME, Skúlason E, Rossmeisl J (2011) Standard hydrogen electrode and potential of zero charge in density functional calculations. Phys Rev B 84:115452. https://doi.org/10.1103/PhysRevB.84.115452

VandeVondele J, Mohamed F, Krack M, Hutter J, Sprik M, Parrinello M (2005) The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. J Chem Phys 122:014515. https://doi.org/10.1063/1.1828433

Wang HF, Liu ZP (2009) Formic acid oxidation at pt/h₂o interface from periodic dft calculations integrated with a continuum solvation model. J Phys Chem C 113:17502–17508. https://doi.org/10.1021/jp9059888

Wasileski SA, Janik MJ (2008) A first-principles study of molecular oxygen dissociation at an electrode surface: a comparison of potential variation and coadsorption effects. Phys Chem Chem Phys 10:3613

Weaver MJ (1998) Potentials of zero charge for platinum(111)-aqueous interfaces: a combined assessment from in-situ and ultrahigh-vacuum measurements. Langmuir 14:3932–3936. https://doi.org/10.1021/la9801054

Xia X, Berkowitz ML (1995) Electric-field induced restructuring of water at a platinum-water interface: a molecular dynamics computer simulation. Phys Rev Lett 74:3193–3196

Zhang C, Wu J, Galli G, Gygi F (2011) Structural and vibrational properties of liquid water from van der Waals density functionals. J Chem Theory Comput 7(10):3054–3061. https://doi.org/10.1021/ct200329e

Zhao J, Chan CT, Che JG (2007) Effects of an electric field on a water bilayer on Ag(111). Phys Rev B 75:085435. https://doi.org/10.1103/PhysRevB.75.085435

Fundamental Atomic Insight in Electrocatalysis

60

Alexander Bagger, Ivano E. Castelli, Martin Hangaard Hansen, and Jan Rossmeisl

Contents

Abstract

Electrochemical energy conversion reactions depend on the atomic structure of the interface between the electrode and the electrolyte. In order to make

---

A. Bagger · J. Rossmeisl (✉)
Department of Chemistry, University of Copenhagen, Copenhagen, Denmark
e-mail: alexander@chem.ku.dk; jan.rossmeisl@chem.ku.dk

I. E. Castelli
Department of Energy Conversion and Storage, Technical University of Denmark, Lyngby, Denmark

M. H. Hansen
SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, California, CA, USA

advances in technology, atomic-scale simulations are needed to provide insight and fundamental understanding of the electrocatalytic reactions. Thus electronic structure calculations relevant for electrocatalysis have attracted a lot of attention in the last decade. However, it is not straightforward to translate state-of-the-art simulations into electrocatalysis. These simulations normally are done at constant number of ions and electrons rather than at constant potential and constant pH, which represent the real physical conditions. In fact, due to this, the electrochemical interface presents one of the frontiers for electronic structure simulations.

In this chapter we describe how standard simulations can provide atomic-scale understanding of electrocatalytic reactions. We introduce the computational version of reference electrodes, which are key in the interpretation of simulations. Furthermore, the reference electrodes are used to create phase diagrams and reaction free energy diagrams for electrocatalytic reactions. The chapter will focus on simulations which can be done without any special implementation of the electronic structure method. This means that we focus on explicit solvent and charge-neutral interfaces. The connection to the electrode potential is introduced in the analysis of the simulations rather than in the simulations themselves.

1   Introduction

Electrochemistry holds the promise to transform society from being based on petro-fuel and petro-chemical to being based on renewable energy. Chemicals and fuels are today by far mostly derived from fossil molecules, oil and gas. Oil is however a limited resource, and in a not too distant future, the chemical industry will have to transform. The transformation toward electrification is further undertaken as the price of electricity from renewable sources is decreasing and in some cases it is actually the most economically viable choice.

Electrochemical energy conversion can convert stable molecules into chemicals with a high-energy content. The source of energy is electricity (direct current) and the feedstock can be abundant molecules from air or water. Thus, chemicals and fuels could be produced using safe, cheap, environmental friendly, and more abundant reactants than today. Furthermore, the products could be provided on demand at the place where they are needed, reducing expensive and hazardous transport of chemicals. Electrocatalysis is, however, not yet established for chemical and fuel production, even though it has the potential to be a disruptive technology. Today, only a few chemicals are produced electrochemically on a huge industrial scale. The electrochemical chlor-alkali process is one of the largest chemical reactions worldwide. Another example is production of aluminum, which is a huge and very energy intense industry based on electrochemical reduction via the Hall–Héroult process (Charles 1889).

Electrochemical energy storage and conversion is related to charge transfer reactions between ion-conducting electrolytes and electron-conducting electrodes. It is at this charged interface between the electrode and the electrolyte that the

energy conversion takes place. These interfaces are ubiquitous in electrochemical devices, and the lifetime and efficiency of the electrochemical devices are often related to the performance of the interfaces. The complexity of the interface provides a severe impediment against its experimental elucidation, and currently there are no atomic-scale simulations or analyses that capture all the essential parts of the electrochemical interface. Fundamental atomic-scale insight is needed to understand and improve the interfaces for energy conversion and storage, and the charged solid/liquid interface represents one of the frontiers of atomic-scale simulations.

The aim of this chapter is to provide an overview of the current state of the art for ab initio quantum mechanical simulations of the electrochemical interface. Implicit electrochemical interface models may require special implementation (Lozovoi et al. 2001; Hamada et al. 2013; Jinnouchi and Anderson 2008a, b; Andreussi et al. 2012; Dupont et al. 2013; Sakong and Gross 2016). The focus here is on methods and analysis, which can be performed without any special requirements for the electronic structure code. Therefore, we focus on explicit water and charge-neutral simulations, with the idea that the analysis here can be performed by people with experience with electronic structure simulations. Many of the results shown here have been obtained using the open-source GPAW and ASE codes (Mortensen et al. 2005; Larsen et al. 2009, 2017; Enkovaara et al. 2010).

2 The Electrochemical Cell

A typical electrochemical cell is sketched in Fig. 1. Overall a cell consists of the electrolyte, at least two electrodes, the ion conductor, the electron conductor, and the potential control. In this case, the ion conductor is shown to conduct protons; however, it could also conduct OH^- groups or other ions. The reactions on the electrodes are determined by the potential applied, and the name *Anode* is given to the oxidizing electrode and *Cathode* is the reducing electrode, and these labels can be switched around by tuning the potential to a different range.

In Fig. 1, BH and A are the reactants and B and AH are the products of a redox reaction. The cathode reduces reactants and the anode oxidizes reactants, and the generalized reaction scheme for an electrochemical cell is:

$$\text{Overall: A} + \text{BH} \longleftrightarrow \text{AH} + \text{B} \tag{1}$$

$$\text{Cathode: A} + (H^+ + e^-) \longleftrightarrow \text{AH} \tag{2}$$

$$\text{Anode: BH} \longleftrightarrow \text{B} + (H^+ + e^-). \tag{3}$$
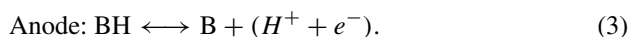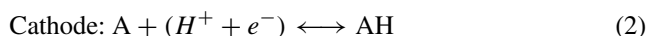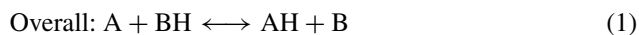
As the above is a generalized reaction scheme for an electrochemical cell, the proton-electron transfer can be utilized for different reactions. As an example, the relevant reaction for the hydrogen evolution reaction (HER) or the hydrogen oxidation reaction (HOR) (Nørskov et al. 2005; Strmcnik et al. 2013), the oxygen reduction reaction (ORR) or the oxygen evolution reaction (OER)
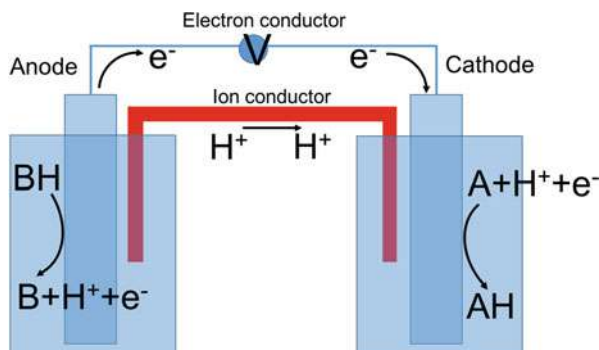
Fig. 1 Illustration of an electrochemical cell with separated electron and ion conductor to emphasize that the two electrodes can in principle be separated at arbitrary distance. The potential control sets up the potential difference to create the oxidizing anode and reducing cathode. BH indicates a reactant to be oxidized (e.g., H_2, methanol), and A indicates a reactant to be reduced (e.g., O_2, CO_2, CO). Note that reactions can be reversed by changing the potential

(Marković et al. 2001; Man et al. 2011; Busch et al. 2016), and the CO_2 reduction reaction (CO_2RR) (Hori et al. 1989; Peterson et al. 2010; Bagger et al. 2017a) or CO oxidation reaction can be written as reduction reaction schemes:

$$2(H^+ + e^-) \longleftrightarrow H_2, \tag{4}$$

$$O_2 + 4(H^+ + e^-) \longleftrightarrow 2H_2O, \tag{5}$$

$$CO_2 + 2(H^+ + e^-) \longleftrightarrow CO + H_2O. \tag{6}$$
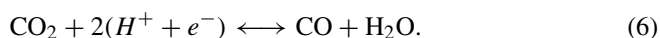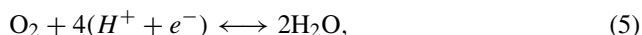
These different reactions require multiple proton-electron transfers, and this indicates that intermediate steps between the reactant and the product exist.

2.1   Reaction Intermediates

Different reaction intermediates which have different atomic composition bind in different ways or adsorb on different sites on the electrode surface. The general reduction scheme of the reactant A can be considered through the intermediate A*, where * indicates a surface site and A* denotes that A binds to the electrode surface:

$$\text{Proton-Electron reaction: } {}^* + A \longleftrightarrow A^* + (H^+ + e^-) \longleftrightarrow AH^*. \tag{7}$$

When a proton and an electron are transferred to the adsorbate A* simultaneously, the reaction is called a proton-electron coupled reaction.

The overall reactions (4), (5), and (6) are proton-electron coupled; however, the intermediate steps in (7) can be observed as coupled, uncoupled, or anywhere in between. Determining the transfer of the electron(s) is difficult as partial charge
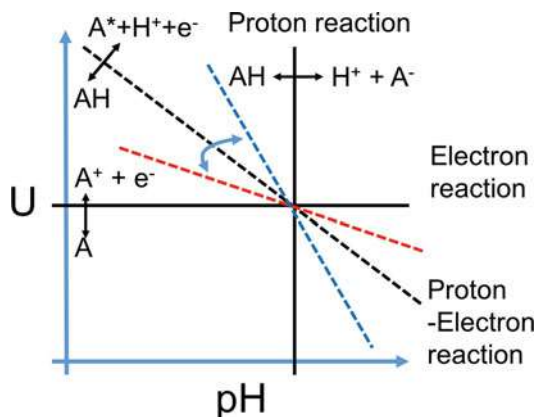
Fig. 2 The Pourbaix diagram shows the three different types of reaction: proton reaction (8), electron reaction (9), and proton-electron reaction (7). The proton reaction, as a typical buffer or acid reaction, only depends on pH (pK_a). The electron reaction, as a typical ionization reaction, depends only on the applied potential. The proton-electron reaction is constant on U_{RHE} scale and thus has a constant slope on absolute scale (U). The stability of the intermediate can be influenced by the interface region (shown by red and blue dashed lines)

of an intermediate is not a physical observable. Instead the continuum of electron transfer states can be accessed quantum mechanically by measuring the dipole of the interface or the work function.

The reactions which only involve a proton (acid-base reaction) or only the electron (ionization reaction) can be written as:

$$\text{Proton reaction: AH} \longleftrightarrow A^- + H^+ \qquad (8)$$

$$\text{Electron reaction: A} \longleftrightarrow A^+ + e^-. \qquad (9)$$
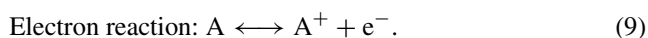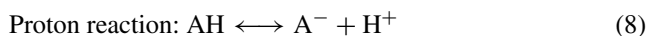
These reactions can be illustrated in a Pourbaix diagram (Pourbaix 1974) as shown in Fig. 2. Here the absolute potential (U) is plotted versus the pH. It shows how the electron and proton reactions (8) and (9) are only dependent on the potential and pH, respectively, while the proton-electron reaction (7) is dependent on both. As it is indicated by the blue and red dashed lines, the electrochemical reduction reaction of A* can be dominated mostly by protons (blue line) or mostly by electrons (red line).

One has to consider the interface between the electrolyte and the electrode as the reaction intermediate can extend into the electrolyte, as is shown in Fig. 3a. The distance from the electrode has been divided into three different regions of interest: the surface region, the interface region, and the electrolyte region.

In the surface region, which is close to the working electrode, the energy of the reaction intermediate is almost unaffected by the electrolyte environment. This would mean that the reaction would be a complete proton-electron reaction

(a) Regions from the electrode.

(b) Electric field effect on intermediates.
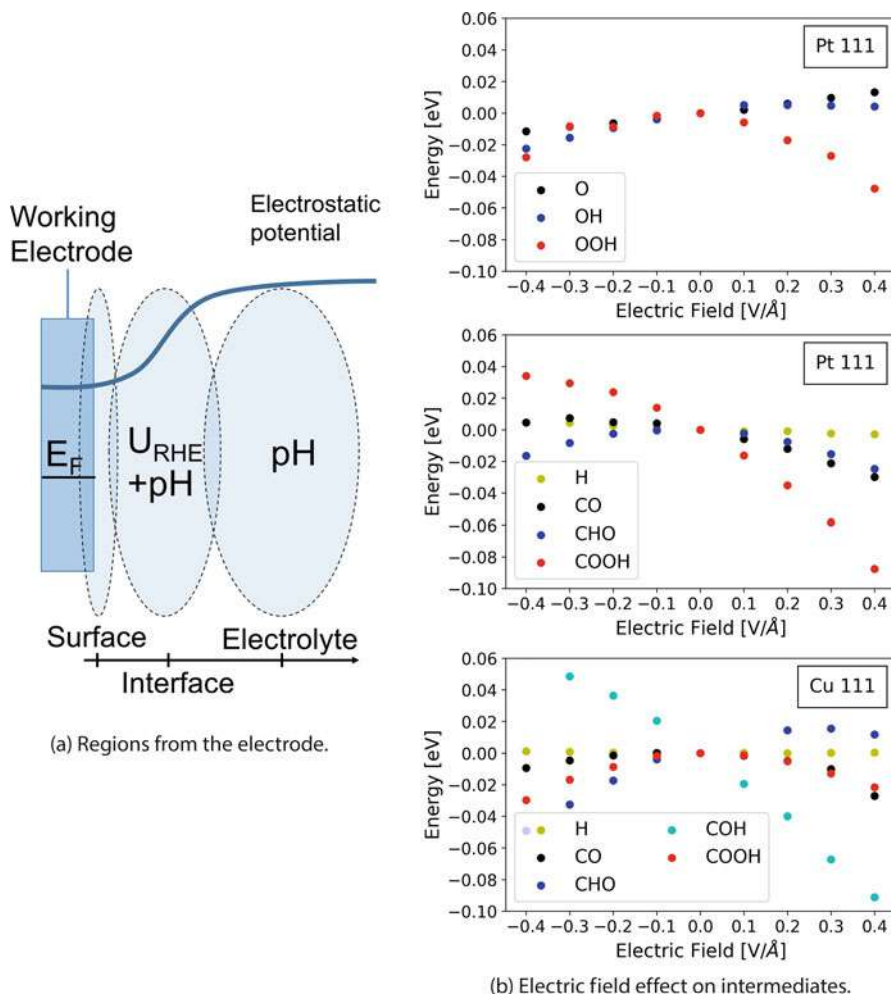
Fig. 3 (**a**) shows the three regions on the direction normal to the electrode: the surface, the interface, and the electrolyte. Reactions in the surface region are dominated by the potential, while far away from the electrode, the conventional pH equilibrium dominates. The interface region requires a control of both the potential and pH. (**b**) shows small, medium, and large intermediates as a function of electric field. Negligible affect of electric field is observed on small intermediates, while medium and larger intermediates, sticking out into the interface region and having a large dipole, are affected more by the electric field

following the black dashed line in Fig. 2, no matter the ions in the electrolyte and the pH.

In the interface region, the intermediates that are affected by the environment are either larger or adsorbs in a more electrolyte exposed position on the surface. Hence the intermediate extends into the interface region, and reducing or oxidizing

this intermediate may deviate from the ideal proton-electron transfer reaction. The electrostatic potential is determined by the absolute potential and an intermediate extending into the interface region with a dipole will interact with the electrostatic field in this region (see Fig. 3a). Lastly, far away from the electrode, in the electrolyte region, acid-base reactions are dominated by the pH only, as the average electrostatic potential is constant regardless of the potential of the working electrode.

The binding energy of an adsorbate is obtained from DFT calculations as:

$$\Delta E_A = E_{slab,A*} - E_{slab} - E_{A(g)}, \tag{10}$$

where the ΔE_A is the binding energy of the intermediate A, $E_{slab,A*}$ is the energy of the slab with A adsorbed, E_{slab} is the bare slab energy, and $E_{A(g)}$ is the energy of the chosen gas-phase reference. Choosing a good reference without functional dependent errors can be challenging; however, methods to remove systematic reference errors can be used (Christensen et al. 2015).

From the DFT-calculated binding energy, ΔE_A, the Gibbs free energy for a given reaction can also be calculated (Reuter and Scheffler 2001, 2003):

$$\Delta G_A = \Delta E_A - T\Delta S + \int_0^T C_p dT + \Delta ZPE \tag{11}$$

where ΔS is the entropy, $\int_0^T C_p dT$ is the heat capacity, and ΔZPE is the zero-point energy correction of the reaction. The adsorption free energy upon different metals for different reactions can be calculated with this equation. But the electrochemical reactions which also involve a proton-electron reaction require additional treatment as given in Sect. 3.

The electrostatic potential can affect the adsorption energies. This can be estimated by adding simple electric field to the binding energy calculation (Karlberg et al. 2007a). The electric double-layer thickness is around $\sim 3\,\text{Å}$, and the electric field effect due to the electrode is approximately on the order of $-0.2\,\text{V/Å}$(Rossmeisl et al. 2006).

Figure 3b shows relevant intermediate binding energies calculated for ORR/OER, HER, and CO2RR as a function of electric field strength on Pt (111) and Cu (111) facets. The graphs show how the intermediate binding energies depend on the electric field. The effect due to the electric field is shown to be very small. However, a solvated ion in the interface have been shown to possibly give local enhancements of the electric field effect up to $-1.0\,\text{V/Å}$ (Chen et al. 2016), and defected structures may also allow for a stronger dependency of the electric field for the binding energy (Mistry et al. 2017).

We note that there is not a direct mapping from the electrochemical potential and the electric field added in these calculations. The field will not be zero, when the potential is zero and different adsorbates does not need to have zero potential at similar fields.

2.2 Reference Electrodes

A requirement to compare the performances of catalysts is that a proper reference electrode is used. The standard hydrogen electrode (SHE) reference electrode is defined as 0 V at pH = 0:

$$(H^+ + e^-) \longleftrightarrow \tfrac{1}{2} H_2(g). \tag{12}$$
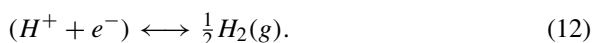
The SHE reference electrode, which is independent on the electrode material, has a potential of $U_{SHE} = -4.44$ eV with respect to the absolute potential, which is defined as electrons at rest in vacuum (Trasatti 1986). Experimentally, the reference electrode used is typically Pt; however, it could easily also be all other kinds of materials. The main advantage of using the Pt as reference electrode is simply that the kinetics of the proton reaction is very fast and equilibrium of (12) is obtained quickly.

A problem when using the U_{SHE} scale is that the activity of the protons change with pH (left side of Eq. (12)). Comparing catalysts in two different electrolytes and pH requires a shift with respect to the change in concentration of protons. The reversible hydrogen electrode (RHE) potential does not change with pH. The RHE scale is simply normalized by the activity of the protons, and the SHE and RHE scale are linked together by:

$$U_{RHE} = U_{SHE} - \frac{k_B T}{ne} \log\left(\frac{a_{H^+}}{\sqrt{p(H_2)}}\right) = U_{SHE} - 0.059V \cdot pH \tag{13}$$

where k_B is the Boltzmann constant, T is the temperature, n is the number of electrons transferred in the reaction, a_{H^+} is the proton activity, and $p(H_2)$ is the pressure of hydrogen at 1 bar. At room temperature the absolute potential change becomes 0.059 V per pH value.

3 The Computational Hydrogen Electrode (CHE)

In electrochemical reactions protons and electrons are transferred. The reaction energy will depend on the chemical potential of the protons and electrons. In the computational hydrogen electrode (CHE) (Rossmeisl et al. 2006; Hansen et al. 2008), the simplification of a combined chemical potential of protons and electrons is made. The CHE is based upon three assumptions:

(1) The bulk of the electrolyte is in equilibrium.
(2) The surface where the reactions take place is in equilibrium with the electrodes and the bulk of the electrolyte.
(3) The adsorption energy is independent on the electrostatic field.

The two first assumptions are used directly when deriving the CHE, and the last is a result of the minor electrostatic effect observed in the dependence between the adsorption energy and field shown in Fig. 3b. This implicitly assumes a proton-electron coupled reaction, and adding or subtracting hydrogen atoms in an electrochemical reaction can be carried out by considering the combined chemical potential of the proton-electron pair. The chemical potential of the proton depends on the pH, and the chemical potential of the electron depends on the potential of the cell. The free energy in an electrochemical reaction of adding and subtracting electrons and protons is:

$$\Delta G = \Delta E - T\Delta S + \int_0^T C_p dT + \Delta ZPE - n \cdot (\mu_{H^+} + \mu_{e^-}), \qquad (14)$$

where μ_{H^+} and μ_{e^-} are the chemical potentials of the protons in solution and the electrons at the Fermi level of the working electrode. These chemical potentials will depend on electrolyte and potential applied for the cell, and a reference for the theoretical calculated electrochemical reaction energy is needed. The hydrogen reference electrode, shown to the left in Fig. 4, is in equilibrium with hydrogen at standard conditions and with electrons and protons. This equilibrium condition can be written as:

$$\tfrac{1}{2}\mu_{H_2} = \mu_{H^+}^{ref} + \mu_{e^-}^{ref} \qquad (15)$$

where μ_{H_2} is the chemical potential of hydrogen gas at standard conditions and superscript *ref* denotes chemical potential of electrons and protons at the reference electrode. Figure 4 illustrates that μ_{e^-} is related to the reference electrode potential at arbitrary pH as:
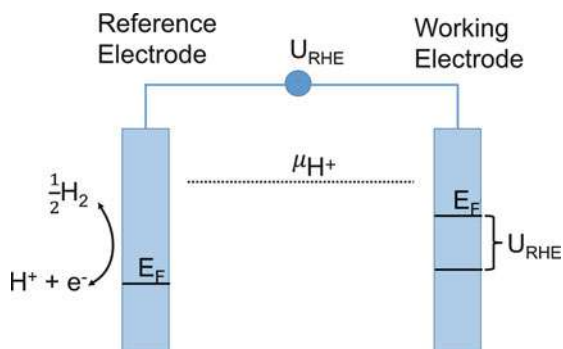
Fig. 4 Illustration of the computational hydrogen electrode (CHE). The hydrogen reaction is in equilibrium with the reference electrode. The protons for both the working and the reference electrode have the same chemical potential, while the electrons at the working electrode are affected by the applied potential (U_{RHE}). In this case U_{RHE} is negative

$$\mu_{e^-}^{ref} = \mu_{e^-} + eU_{RHE} \qquad (16)$$

where U_{RHE} is the potential difference between the working and reference electrode, in this case the reversible hydrogen electrode. Further, we can make the assumption that protons at the reference electrode are in equilibrium with protons everywhere in the electrolyte:

$$\mu_{H^+}^{ref} = \mu_{H^+}. \qquad (17)$$

Computationally, hydrogen can easily be calculated and then referenced back to the chemical potentials of the electrons and protons in solutions. One simply needs to calculate the hydrogen molecule energy and then add the potential U_{RHE} to the equation as follows for the adsorption total energy ΔE_{CHE} and free energy ΔG_{CHE}:

$$\Delta E_{CHE}(n, U_{RHE}) = E(n) - E(n = 0) - n \cdot (\tfrac{1}{2}\mu_{H_2} - eU_{RHE}), \qquad (18)$$

$$\Delta G_{CHE}(n, U_{RHE}) = \Delta E - T\Delta S$$
$$+ \int_0^T C_p dT + \Delta ZPE - n \cdot (\tfrac{1}{2}\mu_{H_2} - eU_{RHE}), \qquad (19)$$

where n is the number of proton-electron transferred and ΔG_{CHE} and ΔE_{CHE} are then a function of both potential and proton-electron pair transfers. The thermodynamic contributions to ΔG_{CHE} from the metal are typically neglected, and only the contributions from the gas phase and vibrational entropy of the adsorbate are included.

The simplicity of the computational hydrogen electrode is obtained by the choice of references and the assumption of electrolyte and electrode in equilibrium. As will be shown in the next sections, the CHE is applicable for calculating many interesting properties in electrochemical reactions, such as surface Pourbaix diagrams, isotherms, and free energy diagrams. Later on, the generalized computational hydrogen electrode (GCHE), where the chemical potentials of the electrons and protons are separated, will be introduced.

3.1   Surface Pourbaix Diagrams

Many factors have to be considered when calculating reaction energies in electrocatalysis. The potential has the effect of changing the coverage on the surface. Reactions can, in fact, be very different if they occur on a clean metal or on a surface with species preadsorbed. The surface Pourbaix diagrams give information on the
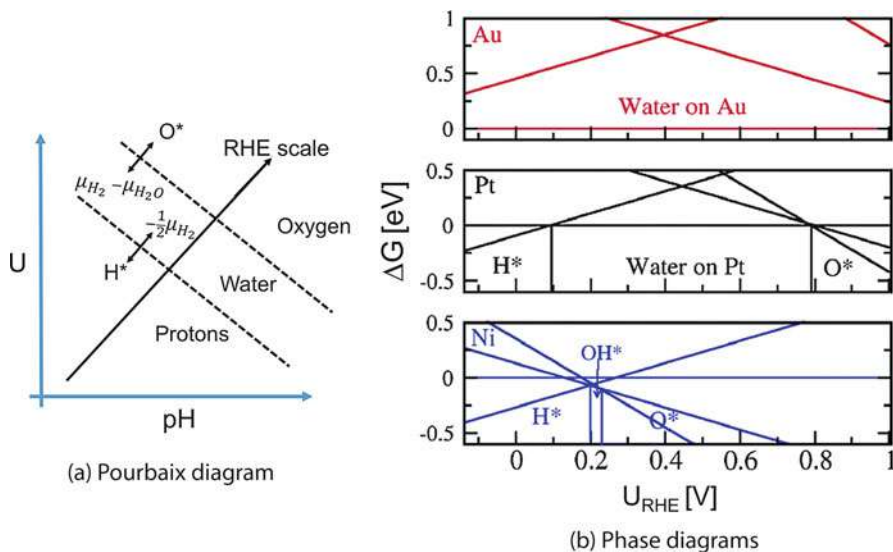
Fig. 5 (**a**) shows a sketch of a Pourbaix diagram with the RHE scale drawn. Where the slope of the RHE scale is 0.059 V/pH at room temperature. (**b**) shows phase diagrams of water on Au(111), Pt(111), and Ni(111) on RHE scale. The thermochemically most stable phase corresponds to the lowest line. (Figure from Rossmeisl et al. 2006)

coverage of a surface in a similar way as the classical Pourbaix diagrams provide an understanding in the phase transition in bulk systems (Persson et al. 2012; Castelli et al. 2014).

The starting point, to construct a Pourbaix diagram, is to consider all possible states that a surface can have. In the example chosen here, four states have been considered: water-, hydrogen-, hydroxyl-, and oxygen-covered surfaces. Where the coverage energy is calculated from the CHE as a function of potential and shown in Fig. 5. The Pourbaix diagrams show the coverages on Au(111), Pt(111), and Ni(111). Au is the noblest of the metals, and therefore it binds adsorbates only at very high or low potentials. Pt is more reactive, and it has hydrogen adsorbed at low potential (typically called hydrogen under potential deposition H$_{upd}$), oxygen at high potential, and a water layer (clean surface) in between. Ni is very reactive and always has adsorbates covering the surface.

This approach is very simple and in particular it can be applied to screening projects and to study trends. On the other side, it fails when large dipole moments are involved because ΔE is influenced by electric fields in the out of place direction of the surface. This issue is addressed by the generalized computational hydrogen electrode (GCHE) in Sect. 4. Further, the surface Pourbaix diagrams assume that no reaction occur, i.e., hydrogen coverage becomes limited if hydrogen is over potential deposited on metals such as Au, Ag, or Cu as hydrogen is evolved.

3.2   Integrated Cyclic Voltammograms

The surface Pourbaix diagrams derived by the use of the CHE depicts when intermediate species are stable on the surface. The presence of intermediates on the metal surface from a proton-electron coupled reaction is a quantity that can be measured experimentally: i.e., the under potential deposited hydrogen (H_{upd}) H* intermediate from the reaction $H^+ + e^- \rightarrow H^*$ or the creation of an OH* from the reaction $H_2O \rightarrow OH^* + H^+ + e^-$ is proportional to the current measure as a function of potential. Experimentally the current for the reactions can be measured by Cyclic Voltammetry (CV), and the integration of the CV over time corresponds to the coverage of charge on the surface, which is related to adsorption isotherms.

The CHE can be used to calculate the differential energy of adding multiple H* or OH* on the surface, where the differential adsorption energy will include the adsorbate-adsorbate interaction. A more simple assumption for intermediate coverages on the surface is to include the differential configurable entropy of non-interaction particles given as:

$$-T\Delta S_{conf} = -k_b T \ln\left(\frac{1-\theta}{\theta}\right) \tag{20}$$

where θ is defined as the coverage on the surface. Adding this entropy term to the CHE, it becomes a function of both potential and coverage:

$$\Delta G(\theta, U_{RHE}) = \Delta G_{CHE}(n=1, U_{RHE}) - k_b T \ln\left(\frac{1-\theta}{\theta}\right) \tag{21}$$

where $\Delta G_{CHE}(n = 1, U_{RHE})$ is the calculated free energy of H* or OH*. Assuming that the coverage on the surface is in equilibrium at all potentials, the coverage can be found as a function of potential:

$$\theta(U_{RHE}) = \frac{1}{1 + \exp\left(\frac{\Delta G_{CHE}(n=1, U_{RHE})}{k_b T}\right)} \tag{22}$$

The coverage of none interacting intermediates on the surface is now defined as a function of potential, which is the Langmuir isotherm expression.

The calculated coverage found by the Langmuir isotherm can be directly compared with the experimental integrated CV (see Fig. 6). In Fig. 6a the isotherms for H* on Pt (111) and Pt (100) are given in the top, and the calculated coverages are given in the bottom. Specifically in these calculations, a mean field interaction between the H* has been included corresponding to a Frumkin isotherm (see Karlberg et al. 2007b). In Fig. 6b similar treatment for the OH* intermediate is shown on Pt and Pt_3Ni for experiment (dashed lines) and for calculation (solid lines).
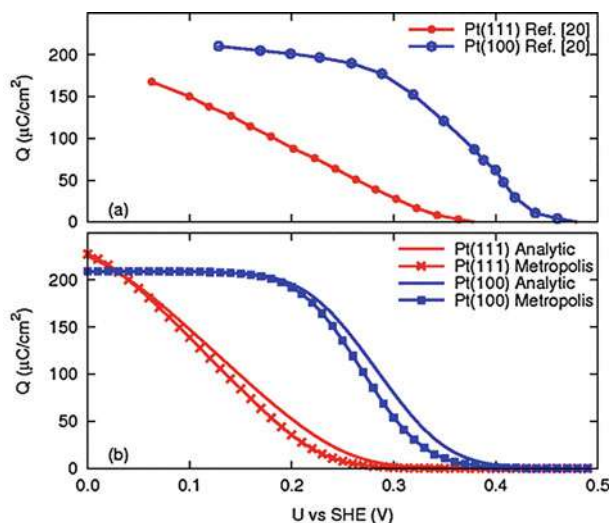
(a) Pt (111) & Pt (100) integrated H* CV
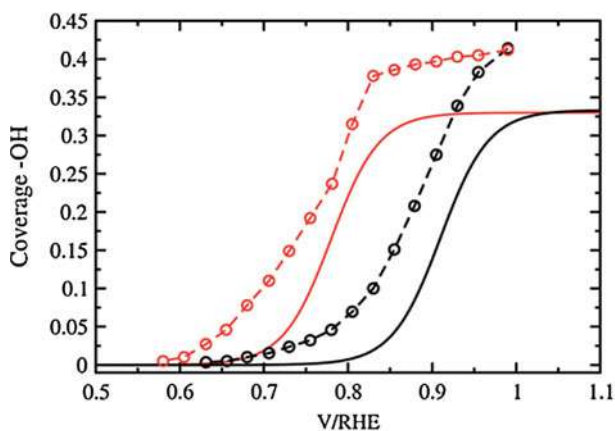
(b) Pt (111) integrated OH* CV

Fig. 6 (**a**) Experimental integrated CV on Pt (111) and Pt (100) for H* (top) compared with the calculated coverage/isotherm (bottom). (**b**) Experimental integrated CV on Pt (red dashed) and Pt$_3$Ni (black dashed) compared with the calculated coverage/isotherm from the 1/3 monolayer OH* binding energy for the two structures (solid lines). (Figures from Karlberg et al. 2007b, Rossmeisl et al. 2009)

For both H* adsorption and OH* adsorption, the calculated coverages compare well with the integrate CVs in the low coverage regime. Since the H* and OH* are reaction intermediates, only the low overpotentials can be compared, since at too high driving forces H$_2$ or O$_2$ start to evolve.

3.3   Understanding Electrocatalytic Reactions and Trends

Another comparison with experiments is the thermodynamically calculated free energy diagram. The free energy diagram allows to interpret different reaction pathways, by calculating the reaction intermediates, and most importantly it gives a "theoretical onset potential." The theoretical onset potential is defined as the potential at which all reaction steps become exergonic, and it allows for a comparison of the overpotential, $\eta = |U^{\Theta} - U^{Theoretical}|$, with experiments. Even if, in experiments, the onset potential is not well defined, the trends in calculated and measured potentials can be compared. It is important to note that the comparison has to be carried out at low currents, as the model assumes equilibrium and this is only a valid approximation at low current densities. Furthermore, the free energies calculated for the free energy diagram can be used to find scaling relations and construct Sabatier activity volcanoes as a function of reactivity descriptors. This framework will be explained here from an OER and ORR perspective.

Free Energy Diagrams

Two key reactions are the four coupled electron proton reactions OER and ORR, which are two reversed sequence reactions. The OER reaction mechanism proceeds through the OH*, O*, and OOH* intermediates:

$$* + H_2O \rightarrow OH^* + H^+ + e^- \tag{23}$$

$$OH^* \rightarrow O^* + H^+ + e^- \tag{24}$$

$$O^* + H_2O \rightarrow OOH^* + H^+ + e^- \tag{25}$$

$$OOH^* \rightarrow O_2 + H^+ + e^- \tag{26}$$
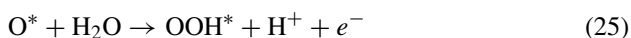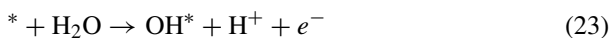
The CHE can be used to calculate the energy of each step of the reaction, with respect to the previous step. Thus, the calculation can be given for only one proton-electron transfer for each calculated ΔG as shown by Eq. (19):

$$\Delta G_1(1, U_{RHE}) = G_{OH^*} - G_{H_2O^*} + \left(\tfrac{1}{2}\mu_{H_2} + eU_{RHE}\right), \tag{27}$$

$$\Delta G_2(1, U_{RHE}) = G_{O^*} - G_{OH^*} + \left(\tfrac{1}{2}\mu_{H_2} + eU_{RHE}\right), \tag{28}$$

$$\Delta G_3(1, U_{RHE}) = G_{OOH^*} - G_{O^*} - G_{H_2O^*} + \left(\tfrac{1}{2}\mu_{H_2} + eU_{RHE}\right), \tag{29}$$

$$\Delta G_4(1, U_{RHE}) = G_{O_2} - G_{OOH^*} + \left(\tfrac{1}{2}\mu_{H_2} + eU_{RHE}\right). \tag{30}$$

The calculated ΔGs can now be plotted in a free energy diagram, as shown in Fig. 7a. For the ideal catalyst, the proton-electron can take place at the equilibrium potential of $U_{RHE} = 1.23\,\text{eV}$ for OER and ORR. This means that the ideal catalyst should bind OH*, O*, and OOH* such that $\Delta G_1 = \Delta G_2 = \Delta G_3 = \Delta G_4 =$
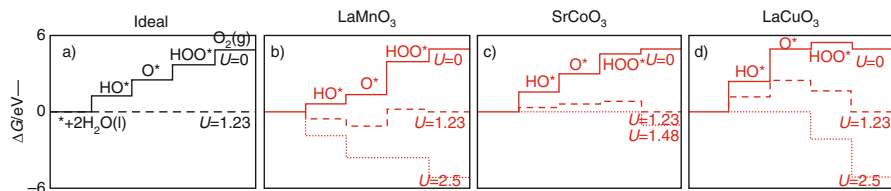
Fig. 7 Free energy diagram for the oxygen evolution reaction at zero potential ($U_{RHE} = 0$ eV), at equilibrium potential ($U_{RHE} = 1.23$ eV) and at the potential where all steps are favorable. (**a**) is the ideal catalyst, and (**b**), (**c**), and (**d**) are the strong, intermediate, and weak binding catalyst. (Figure from Man et al. 2011)

1.23 eV. However, the free binding energy for three perovskite materials shown in (b), (c), and (d) indicates that either the catalyst binds to strongly, weakly, or even on the catalysts with intermediate binding; it is not possible to obtain equal values of each step.

Scaling Relations

The constrains found for the free energy diagram in Fig. 7 are due to the linear relations between the OOH* and OH* as shown in Fig. 8a for various systems (perovskites, rutiles, anatase, Mn_xO_y, Co_3O_4, and NiO oxides). The linear relations between adsorbates are called *scaling relations* and have been found between C-, N-, S-, and O-derived adsorbates on catalyst surfaces (Fernández et al. 2008). The scaling relations limit reactions which have intermediates that bind through the same atom and on the same site (on top, bridge, or hollow). This indicates a limitation in achieving better catalysts. As the scaling relation limits the properties of catalyst, *breaking the scaling* is a hot topic, and different suggestions have been proposed to circumvent this (Halck et al. 2014; Montoya et al. 2015; Doyle et al. 2015). The particular difficult scaling to *break* is the one for OER and ORR, where OOH* and OH* both binds to the on top site and through the oxygen and the two energies scale as:

$$\Delta G_2 + \Delta G_3 = 3.2 \pm \sigma, \tag{31}$$

where σ notes that there are some minor spread around the scaling relation shown in Fig. 8a.

On the other hand, one way of using the scaling relations, without breaking them, is to utilize the fact that instead of having a catalyst with all three types of sites (on top, bridge, or hollow), porphyrin structures (single site catalyst) may have a structural advantages of simply not having the hollow site (Bagger et al. 2017b). This can be utilized to control the ORR mechanics toward H_2O_2 by removing the hollow oxygen binding possibility (Siahrostami et al. 2013). Or in CO_2RR, the competitive HER can be limited by not having the hollow site (Bagger et al. 2017b; Ju et al. 2017).
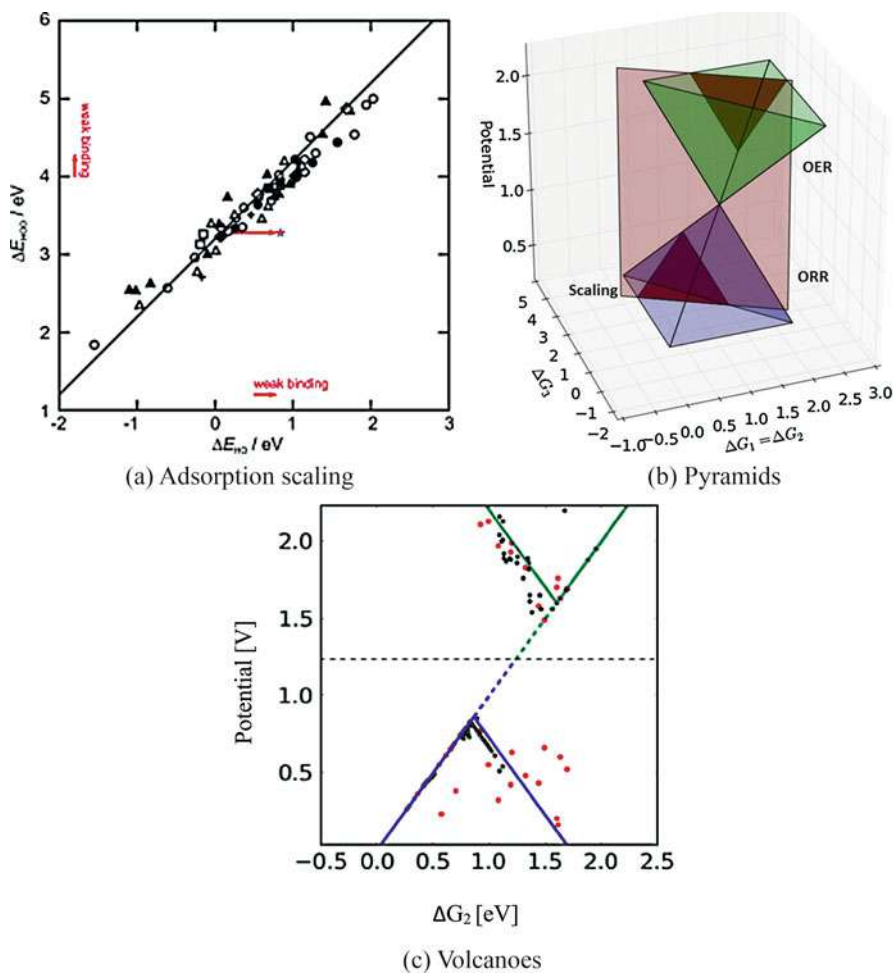
Fig. 8 (**a**) shows scaling relation between OOH* and OH*, (**b**) shows the three-dimensional pyramids of the OER/ORR parameter space, and (**c**) shows calculated transition metal and nitrogen-doped carbon catalyst materials to follow the scaling relation constrain and have different performance as OER or ORR catalyst. Each surface will have ΔG_1 and ΔG_2 values, which will correspond to a theoretical overpotential for ORR and OER read of the pyramids in (**b**). For normal surfaces the ΔG_1 and ΔG_2 values are related via scaling in (**a**), which means that most catalyst are constrained to the red plane in (**b**) constructing the volcanoes in (**c**). (Figures from Man et al. 2011; Busch et al. 2016)

Volcano Plots

The catalytic volcano plots are based on the Sabatier principle (Nørskov et al. 2014), which states that the interaction between the catalyst surface and the adsorbate/intermediate should be just right. The adsorbate should bind strongly enough to be brought to the surface, but not too strongly so that the product cannot

get off the catalyst and consequently an optimal catalyst binds just right. With the scaling relations the Sabatier principle can now be put on a quantitative form, and for electrochemistry reactions typically a binding free energy of intermediates is used for the x-axis. The volcano illustrates the theoretical overpotential that has to be paid to make the reaction run with all reaction steps being exergonic.

The ideal theoretical Sabatier volcano for OER and ORR is shown as three-dimensional pyramids in Fig. 8b. The most promising OER and ORR catalyst would be at the intersect of the two volcanoes. The scaling relation gives the constraint between the OOH* and OH*. This constraint will restrict catalysts to lie in the red plane. The red plane shows that obtaining the ideal catalyst at the intersect of the two pyramids ($U_{RHE} = 1.23$ eV) is not possible when fulfilling the scaling. Furthermore the sliced 2D plane is shown in Fig. 8c with calculated data (Man et al. 2011; Viswanathan et al. 2012). The two volcanoes show that if one material is good for OER, then it will be poor for ORR and vice versa. Furthermore, none of the two volcanoes reach the thermodynamic limit, and a large over potential has to be paid for all OER and ORR catalysts, which also has been observed experimentally (Seh et al. 2017).

3.4   Electrochemical Barriers

The comparisons of free energies and volcano plots with experiments are based on thermodynamics. To obtain kinetics of reactions, the barrier between reaction steps must be estimated. For heterogeneous catalysis the theoretical barriers can be calculated using the nudged elastic band (NEB) method (Jónsson et al. 2011; Henkelman et al. 2000; Henkelman and Jonsson 2000). Although similarities clearly exist between heterogeneous barriers and electrochemical barriers, the latter is not straightforward to obtain and to add to the electrochemical free energy diagram.

Correct electrochemical barriers are notoriously difficult to calculate, since the DFT-calculated unit cell has a constant charge (and changing potential), while in real experiments the potential is constant (and current/charge changes). This is shown in Fig. 9, where barriers are calculated for the Heyrovsky reaction on three different unit cells with different area. The Heyrovsky reaction is the reaction of a hydrogen on the surface (H*) and a proton (H^+) in the double layer creating H_2. Due to the translation of the proton in the double layer, the reaction involves a transfer of charge. From the NEB calculations, the energetic barriers (black) and the changes in dipole (red) can be plotted against the reaction coordinate. As the charges move, the measured dipole changes from initial state to final state by more than 3 V for the small cell and for about 0.5 V at the largest unit cell shown here. Since the dipole is directly proportional to the work function which is the potential of the electrode, the potential changes during the calculation of the reaction path. If the area of the interface gets bigger, the change in work function related to transferring one charge becomes smaller. In the limit of a very large area, the change in work function becomes neglectable. However, this is not possible, simply due to the practical size of the unit cell and a finite size problem present. The issue is that a charge transfer
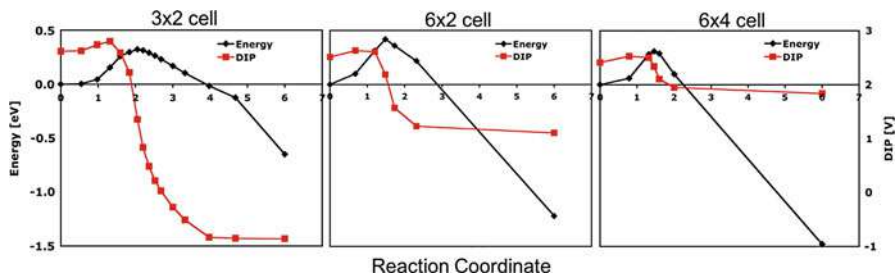
Fig. 9 Barriers calculated for the Heyrovsky reaction $H^+ + e^- + H^* \rightarrow H_2$ in three different unit cell sizes of Pt (111). The energy is plotted together with the calculated dipole of the slab, where the dipole is directly proportional to the work function. The changes on the dipole along the reaction coordinate illustrate the problem of constant charge in the calculation. (Figures from Skulason et al. 2009)
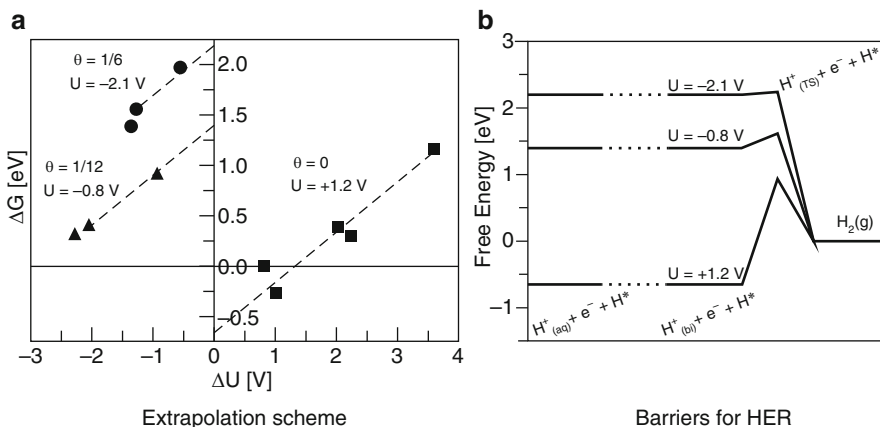
Fig. 10 (**a**) The free energy barriers of the reaction Heyrovsky reaction are plotted versus the change in electrode potential and for minimum three different unit cell sizes; the barriers are extrapolated unto zero potential change. (**b**) Free energy diagram with electrochemical barriers added at three different potential for Pt(111). (Figures from Rossmeisl et al. 2008)

step changes the potential as the charge species interact with the potential setup by itself and thereby effecting the energy of itself.

In order to handle the finite size problem and obtain an electrochemical barrier at a constant potential, an extrapolation scheme has been developed (Skulason et al. 2007; Rossmeisl et al. 2008). This scheme applies the observed results from Fig. 9, in which the same reaction is calculated for different unit cell sizes and this allows to obtain barriers as a function of changing electrode potential (Skulason et al. 2010). An extrapolation to $\Delta U = 0$ is possible by plotting these barriers versus the changing electrode potential, and the barriers can be estimated as shown in Fig. 10.

Recently, advances have been made to calculate electrochemical barriers in a simple way, by using only one NEB calculation for each barrier (Chan and

Nørskov 2015). This new scheme uses the assumption of a decoupled chemical and electrostatic contribution to the calculation. This has been used to calculate barriers in particularly for the CO_2RR (Shi et al. 2016; Liu et al. 2017).

Furthermore, a proton shuttling scheme has be made to calculated the barriers. This is carried out by transferring a proton from the surface to the electrolyte, while transferring a proton from the electrolyte to the surface, which could then ensure an approximate constant potential in the calculation. This method has also been used to investigate the kinetics of the CO_2RR (Nie et al. 2014).

These schemes allow the dipole of the interface to stay constant, and the barriers may be considered at constant electrochemical conditions: pH and potential. However, it is still limited by only being referred to the RHE scale. Even if the barriers can give some insight, the structure of the interface may not be the relevant one at realistic conditions. Finding the relevant structure of the electrochemical interface region may be important to obtain more accurate electrochemical barriers. Determining stable structures at given pH and potential is the subject in the next section.

4 The Generalized Computational Hydrogen Electrode (GCHE)

The CHE was derived by applying three assumptions (see Sect. 3). The assumption (3) *the adsorption energy is independent of the electrostatic field* does not apply if intermediates have a strong dipole or interacts strongly with the electrolyte. In that case, the simulation of the intermediate binding energy must include the interface region (Fig. 3a), and the assumption (3) can be replaced by:

(4) The interface region is charge neutral and large enough to screen all charges (Hansen and Rossmeisl 2016).

For the CHE model, the U_{RHE} was defined as the potential difference between the reference and the working electrode no matter the chemical potential of the protons. To create the generalized computational hydrogen electrode (GCHE) (Rossmeisl et al. 2013), the proton and electron chemical potentials have to be referred to two reference electrodes using the Born-Haber cycle by two paths (Fig. 11): (1) the RHE reference through $\mu_{H^++e^-}$ (eU_{RHE}) or (2) through the

Fig. 11 The Born-Haber cycle, which allows two pathways to obtain the chemical potential of protons and the electrode potential
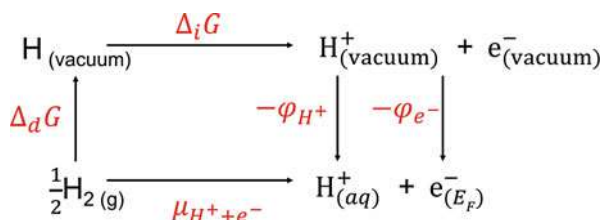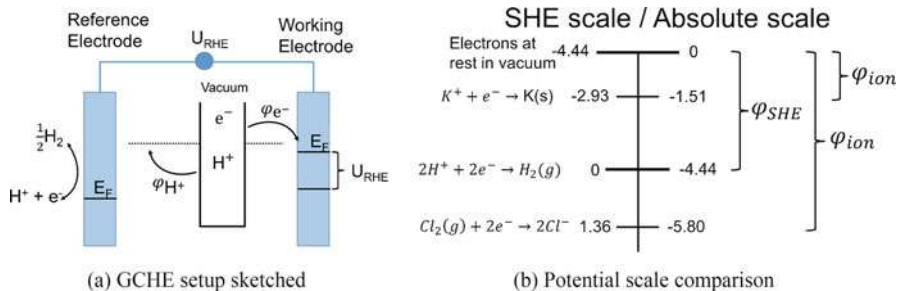
Fig. 12 (**a**) Illustration of the generalized computational hydrogen electrode (GCHE), which includes both the proton work function (ϕ_{H+}) and the electron work function (ϕ_{e+}). (**b**) The standard hydrogen electrode (SHE) scale and the absolute potential scale as comparison, to understand how to include ions into the interface region.

vacuum state of the proton H^+ and the electron e^-. Now two reference points can be obtained: the reference electrode and the near field vacuum electron work function (Fig. 12a). From these two references, the electrode potential and pH can be measured separately, which is shown next.

Using the Born-Haber cycle, we can obtain U_{RHE} in the following way:

$$eU_{RHE} = \mu_{H^+ + e^-} = \Delta_d G + \Delta_i G - \phi_{H^+} - \phi_{e^-}, \tag{32}$$

where $\Delta_d G$ is the hydrogen molecule dissociation energy, $\Delta_i G$ is the hydrogen atom ionization energy, ϕ_{H^+} is the proton work function, and ϕ_{e^-} is the electron work function (defined as $\phi_{e^-} = E_{e^-} - E_{E_F}$). For pH = 0 and $U_{RHE} = 0$ eV, this expression is equal to zero:

$$0 = \Delta_d G + \Delta_i G - \phi_{H^+}^0 - \phi_{SHE}, \tag{33}$$

where $\phi_{H^+}^0$ is the energy difference between H^+ at rest in vacuum and H^+(aq) at pH = 0. Further, ϕ_{e^-} has changed to ϕ_{SHE}, which is an experimental value between 4.44 and 4.85 eV (Trasatti 1986; Kötz et al. 1986). The pH-dependent proton work function can be obtained by the relation:

$$\phi_{H^+} - \phi_{H^+}^0 = 2.3 k_B T p H. \tag{34}$$

Combining these equations, the eU_{RHE} in the CHE can be replaced to include terms from both the electrode potential and pH. The free energy for the GCHE for adding or subtracting hydrogens from the interface is obtained:

$$\Delta G_{GCHE}(n, \phi_{e^-}, pH) = E(n, \phi_{e^-}) - E(n = 0) - n\frac{1}{2}\mu_{H_2}^0$$

$$- ne \overbrace{(\phi_{SHE} - \phi_{e^-} - 2.3 k_B T p H)}^{U_{RHE}}, \tag{35}$$

Fig. 13 shows the ΔG_{GCHE} energy as a function of work function and eU_{RHE}, with planes representing specific pH values (gray planes) and a plane for adding a proton (yellow plane) or creating a hydroxy (blue plane) and water (red plane)
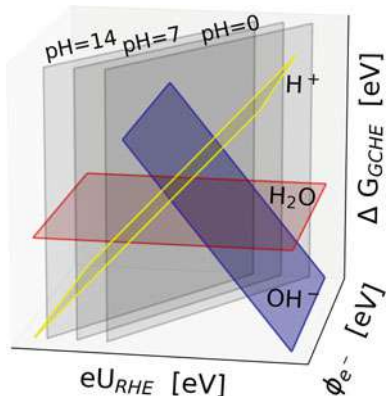
Fig. 14 shows the GCHE setup in a side view of a DFT unit cell with four Au metals layers and three waters layers to the right of the Au slab. From this setup the work function ϕ can be measured together with the energy. (Figures from Hansen and Rossmeisl 2016)
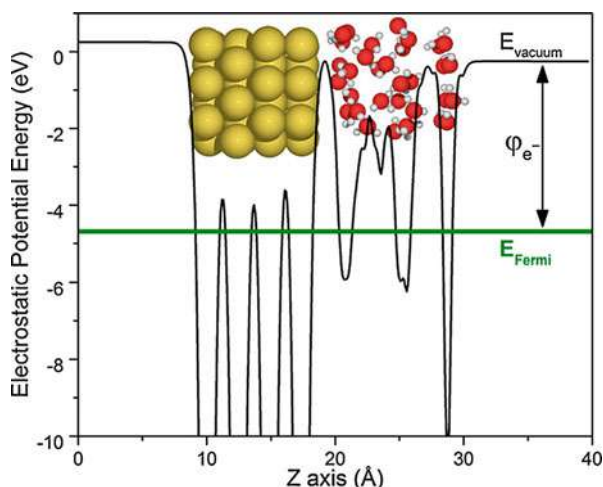
where $E(n, \phi_{e^-})$ can be obtained from DFT with added or subtracted number of hydrogens, $E(n = 0)$ is a reference energy, $\mu_{H_2}^0$ is the hydrogen energy including ΔZPC and $-T\Delta S$ for the hydrogen molecule, ϕ_{SHE} is the SHE work function (Fig. 12b), ϕ_{e^-} is the calculated work function (Fig. 14), and the last term defines the energy with respect to the pH of the electrolyte.

It can be very illustrative to think of the GCHE in terms of a three-dimensional plot as shown in Fig. 13. This figure shows how the ΔG_{GCHE} energy is related to eU_{RHE} and the work function ϕ_{e^-}. In the three-dimensional plot, the pH planes of pH = 0, 7, and 14 are drawn (gray vertical planes) together with a plane for when a hydrogen is added (yellow) or removed (blue) leaving an OH$^-$ and a pure water simulation (red). The intermediate planes are displayed linearly (following RHE); however, they could also have a curvature (interface region). Note, for these planes, that our usual surface Pourbaix diagram corresponds to a projection into the ΔG_{GCHE} and eU_{RHE} plane.

To simulate the correct electrochemical environment, one may need to include charged ions. For this purpose the GCHE equation (35) needs to be modified from the proton reaction ($2H^+ + 2e^- \rightarrow H_2(g)$) to the ionization reaction of the charged ions. The ionization potentials for (positive) potassium and for (negative) chlorine ions are shown as an example on the SHE scale and absolute scale in Fig. 12b. Here ϕ_{ion} relates the ionization potential to the absolute potential scale, and the GCHE equation becomes:

$$\Delta G_{GCIE}(n_{ion}, \phi_{e^-}, [A]) = E(n_{ion}, \phi_{e^-}) - E(n = 0) - n_{ion}\mu_{ion}^0$$
$$- n_{ion}e\left(\phi_{ion} - \phi_{e^-} - 2.3k_B T ln\left(\frac{[A]}{[A]^0}\right)\right) \quad (36)$$

where $[A]$ is the reference ion concentration in the electrolyte, $[A]^0$ is the standard condition of 1 M, and G_{GCIE} refers to the fact that this is now a generalized computational ion electrode. Furthermore, it should be noted that in Eq. (36) the last term changes sign for the chlorine ionization reaction path compared to the proton and potassium reaction in Fig. 12b.

4.1   Practical and Technical Aspects of the GCHE

The CHE method has usually been applied to find one binding energy for each intermediate. The GCHE method measures the energy and the corresponding work function of the specific structure of the interface. For one structure, only one energy and work function is available, which for the GCHE means that interface energy is only found at one absolute potential. Hence we should sample the interface region of the electrochemical cell with the electrolyte (water molecules, ions, adsorbates, etc.) at some fixed temperature and generate an ensemble of states with different work functions to obtain an ensemble of energies in the wanted potential range. This sampling can be carried out by molecular dynamics (MD) simulations, which exist at various different levels of treating the electrolyte and metal (Spohr 1989; Sergeev et al. 2017; Merlet et al. 2014; Liang et al. 2014; Yeh et al. 2013; Larentzos and Criscenti 2008). Ab initio MD (AIMD) (Schnur and Groß 2009; Herron et al. 2016) uses electronic structure codes, and the values of interest can be directly obtained from this.

An example of ab initio MD simulation (i.e., where DFT is combined with MD) to measure the energy and work function is shown for Au (111) with four water layers in Fig. 14. The three water layers closer to the Au slab are allowed to move, while the outermost layer is fixed in hexagonal structure to keep the water density constant. The energy and the ϕ_{e^-} can be measured as the MD simulation is kept at room temperature, and adding or subtracting hydrogens allows to find the H* or OH* coverage on Au (111) as function of pH and potential.

This methodology is a simple brute force approach, which is impractical. Currently this can be carried out only on the proof of concept level due to the high computational cost required to sample enough structures. Future advances in algorithms and resources may be needed to leverage further advances from the methodology.

4.2   Electrolyte and pH

Here we show how the water interface structure, which can exist at a specific potential and pH, can be found in the case of pure water and with the limitations of the DFT setup (functional, PAW setup, etc.). To do this the GCHE is applied to a set of water structures obtained from MD with the calculation environment shown in Fig. 14. The pure water is sampled together with systems adding or removing up to six extra hydrogen atoms. The subtraction or addition can be written as n/N, where N = 12 is the number of surface atoms in the unit cell. In total this sampling requires 13 temperature equilibrated MD simulations, which is very computational consuming (one equilibrations currently takes approximately 5000 cpu hours).

From the MD structures, the energies, work functions, and number of protons are saved. In this way the interface energy can be calculated by using the GCHE equation (35), and the result are plotted in Fig. 15a for pH = 0 (left) and pH = 14 (right). These plots are equivalent to Fig. 5, but here each point is at a particular potential instead of extrapolating one point to all potentials. In the plots each line corresponds to a trajectory with n/N protons. To obtain the water structure at a given potential, one can manually select the minimum energy states from the plot or apply a Monte Carlo algorithm, which has been used to find the structures at finite temperature as shown in Fig. 15b.

Similar GCHE, MD ensembles, and Monte Carlo treatment have been carried out on Pt (111) (Hansen et al. 2017), where the Pt (111) interface model is of particular interest in the electrocatalytic proton exchange reactions. A difference from the previous CHE calculated integrated CV, and the experimental integrated CV is shown in Fig. 6b. The binding energy used was obtained from a coverage of 1/3 monolayer of OH. By using the GCHE framework, a stable 5/12 monolayer OH* structure was found and shown in Fig. 16. This shows that combining GCHE with MD ensembles and Monte Carlo, we may access unknown states and put them on an energy scale to find the most stable interface structure as a function of pH and potential.

5   Additional Considerations

Until now, we have described how to calculate various electrochemical properties on metallic surfaces. For the metallic surfaces, the calculations of the absolute level of the electronic states have been avoided, since the binding on metals is much better described with, e.g., DFT than the absolute positions of the electronic states.
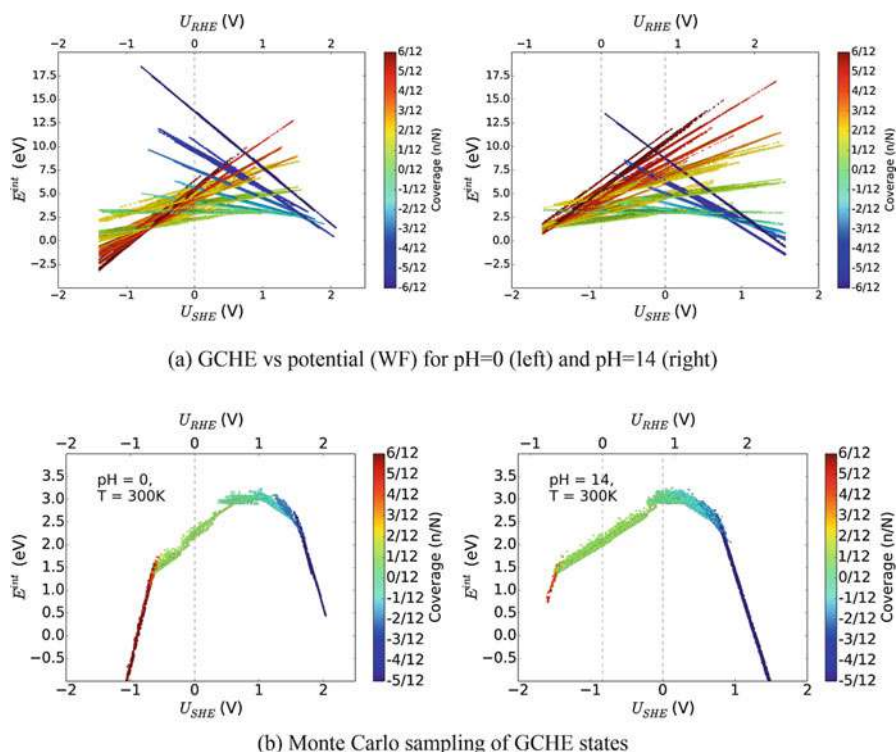
(a) GCHE vs potential (WF) for pH=0 (left) and pH=14 (right)

(b) Monte Carlo sampling of GCHE states

Fig. 15 Molecular dynamics water sampling of the interface region setup. (**a**) shows the GCHE energy as a function of potential and pH (left and right), while (**b**) shows the most stable state found by applying a Monte Carlo algorithm to the states in (**a**). (Figures from Hansen and Rossmeisl 2016)

Fig. 16 OH* adsorption isotherms calculated using the GCHE method with liquid water structures from molecular dynamics on Pt (111) surface. The inset shows a snapshot of the 5/12 ML OH* chemisorbed layer structure corresponding to the experimental integrated cyclic voltammogram from Fig. 6b. (Figure from Hansen et al. 2017)

(a) Energy diagram of metal and electrolyte levels.

(b) Three regions of the interface.
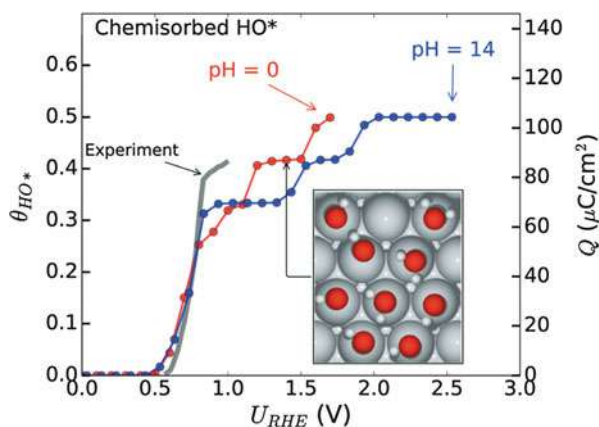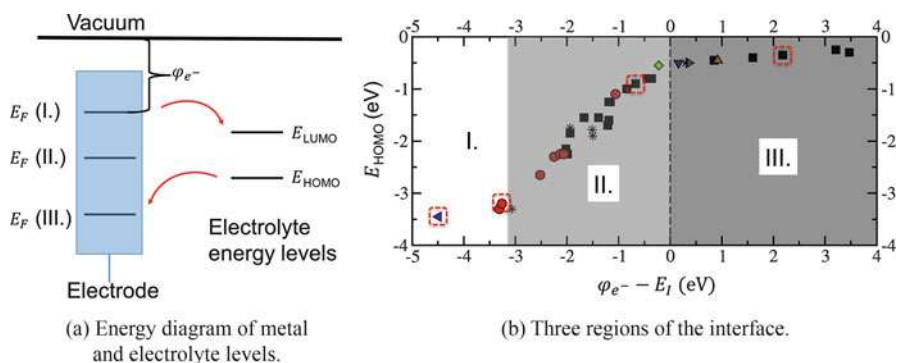
Fig. 17 (**a**) Illustration of how the metal Fermi level should be aligned between the electrolyte HOMO and LUMO levels (II.). When this is the case, the pinning of the HOMO energy level is avoided as shown in (**b**). (Figures from Björketun et al. 2013)

However, we will give two examples where the energy of the electronic states will play a role. This represents situations where the previous described approach may not be sufficient: (1) interface alignment of electronic states and (2) calculation of binding energies on semiconductors.

A critical issue when calculating interface models is the alignment of the electrode and molecular water electronic energy levels (Cheng and Sprik 2010; Björketun et al. 2013), as shown in Fig. 17a. The issue occurs when the work function of the metal is either above the lowest unoccupied molecular level (LUMO) of the electrolyte, labeled (I.), or below the electrolyte highest occupied molecular level (HOMO), labeled (III.). In both these cases, the Fermi level will be pinned, by either electron transfer from the Fermi energy to the LUMO in case (I.) or electron transfer from the HOMO to the Fermi energy in case (III.). Only if the Fermi energy of the electrode is in between the electrolyte HOMO and LUMO energy levels, proper potentials can be obtained (none-pinned). Calculations of the HOMO level as a function of the difference between the metal work function and electrolyte ionization potential is presented in Fig. 17b. These results directly show how the HOMO level becomes pinned with respect to the metal Fermi energy in region (I.) and (III.), while a proper interface region can be simulated in (II.). As seen in the pinned regions, the work function will not change even if more charge is added to the electrode.

The adsorption properties on semiconductors carry similar problems, where the positions of the bands of the semiconductor force a charge transfer to or from the adsorbate. This is shown by the densities of states calculation on MgO in Fig. 18b. In an intrinsic semiconductor, the Fermi level is placed in the middle of the band gap (black line and labeled as *). When an electron donor is adsorbed (red lines), an electron charge transfers to the conduction band of the slab and the Fermi level shifts up. In the opposite case, if an electron acceptor is adsorbed (blue lines), the
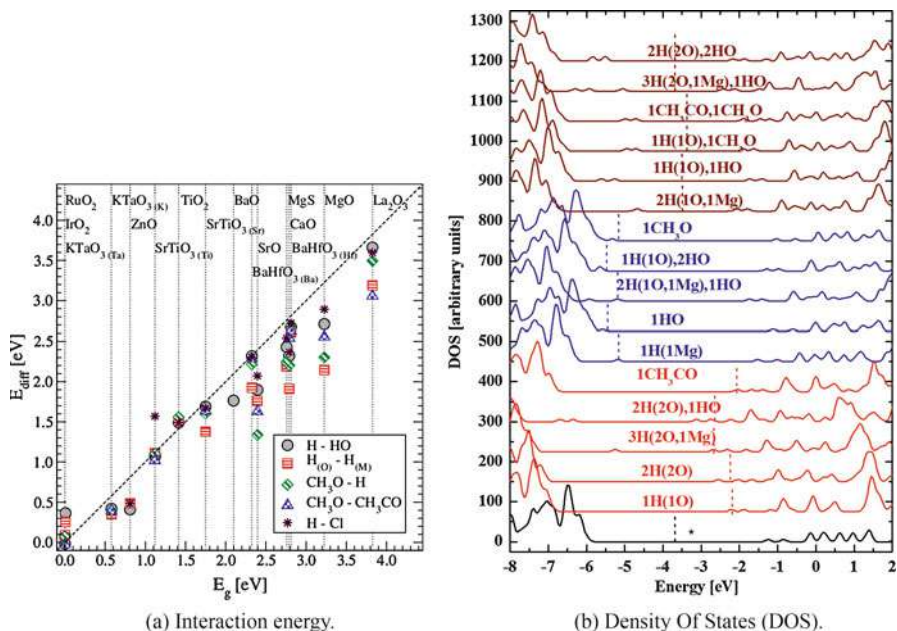
Fig. 18 (**a**) Interaction energy, E_{diff}, as a function of the band gap, E_g, for different fragments on various surfaces. The one-to-one relation between the interaction energy and the band gap is indicated by a dashed line. (**b**) DOS of various fragments on MgO. The electron donors are indicated in red, the electron acceptors in blue, and an electron acceptor-donor pair in brown. (Figures from Castelli et al. 2017)

charge moves from the valence band to the fragment and the Fermi level shifts down. When an electron donor-acceptor pair is formed (brown lines), the shift of the Fermi level is close to zero because the charge donated from one fragment and is accepted by the other and this happens through the surface. This phenomenon is called *cooperative adsorption* or *interaction energy* and has been shown to happen on semiconductors (Broqvist et al. 2002; Ménétrey et al. 2007; Metiu et al. 2012; Hu and Metiu 2012; Li and Metiu 2012). The band gap of the clean surface has been identified as a simple descriptor for the interaction energy. As shown in Fig. 18a for various fragments and surfaces, there is a one-to-one relation between the band gap of the clean surface and the interaction energy. The binding energy of an electron donor-acceptor pair, $\Delta E_{\text{A,D}}$, is equal to:

$$\Delta E_{\text{A,D}} = \Delta E_{\text{A}} + \Delta E_{\text{D}} + \Delta E_g \tag{37}$$

where ΔE_{A} and ΔE_{D} are the binding energies of the A and D fragments calculated alone in the unit cell (Castelli et al. 2017).

6 Summary

In this chapter first principle models of the electrochemical cell have been elucidated. The electrochemical environment has been included by posterior analysis of ab initio calculations.

The computational hydrogen electrode (CHE) has been derived and the assumptions of it explained. Key results obtained by applying the CHE together with state-of-the-art DFT calculations have allowed to derive surface Pourbaix diagrams, integrated CVs, free energy diagrams, Sabatier volcanoes, and electrochemical barriers. Trends in these properties can be compared with trends in experiments.

To decouple the combined chemical potential of protons and electrons, the generalized computational hydrogen electrode (GCHE) has been derived. This allows one to obtain the most stable states with a decoupled dependency on pH and potential. Because of this the GCHE relies on measuring the energy and work function together, allowing only one energy per calculated system and one to obtain an ensemble of states. This generalized setup can be used to derive many interesting properties of the electrocatalytic interface region such isotherms, specific capacities calculated from the differential isotherms, potential of maximum entropy, atomic distance distributions to the surface and atomic pair correlation functions of the interface, and more.

Finally, additional considerations have been made upon systems that are not straightforward to calculate and how correct band alignment must be existing to obtain correct energetics of the interface.

References

Andreussi O, Dabo I, Marzari N (2012) Revised self-consistent continuum solvation in electronic-structure calculations. J Chem Phys 136(6):064102

Bagger A, Ju W, Varela AS, Strasser P, Rossmeisl J (2017a) Electrochemical CO_2 reduction: a classification problem. ChemPhysChem 18:3266–3273

Bagger A, Ju W, Varela AS, Strasser P, Rossmeisl J (2017b) Single site porphyrin-like structures advantages over metals for selective electrochemical CO_2 reduction. Catal Today 288:74–78

Björketun ME, Zeng Z, Ahmed R, Tripkovic V, Thygesen KS, Rossmeisl J (2013) Avoiding pitfalls in the modeling of electrochemical interfaces. Chem Phys Lett 555(Supplement C):145–148

Broqvist P, Panas I, Fridell E, Persson H (2002) NO_x storage on BaO(100) surface from first principles: a two channel scenario. J Phys Chem B 106(1):137–145

Busch M, Halck NB, Kramm UI, Siahrostami S, Krtil P, Rossmeisl J (2016) Beyond the top of the volcano? A unified approach to electrocatalytic oxygen reduction and oxygen evolution. Nano Energy 29:126–135. Electrocatalysis

Castelli IE, Thygesen KS, Jacobsen KW (2014) Calculated Pourbaix diagrams of cubic perovskites for water splitting: stability against corrosion. Top Catal 57(1):265–272

Castelli IE, Man I-C, Soriga S-G, Parvulescu V, Halck NB, Rossmeisl J (2017) Role of the band gap for the interaction energy of coadsorbed fragments. J Phys Chem C 121(34):18608–18614

Chan K, Nørskov JK (2015) Electrochemical barriers made simple. J Phys Chem Lett 6(14):2663–2668

Charles MH (1889) Process of reducing aluminium from its fluoride salts by electrolysis, 2 Apr 1889. US Patent 400,664

Chen LD, Urushihara M, Chan K, Nørskov JK (2016) Electric field effects in electrochemical CO_2 reduction. ACS Catal 6(10):7133–7139

Cheng J, Sprik M (2010) Aligning electronic energy levels at the TiO_2/H_2O interface. Phys Rev B 82:081406

Christensen R, Hansen HA, Vegge T (2015) Identifying systematic DFT errors in catalytic reactions. Catal Sci Technol 5:4946–4949

Doyle AD, Montoya JH, Vojvodic A (2015) Improving oxygen electrochemistry through nanoscopic confinement. ChemCatChem 7(5):738–742

Dupont C, Andreussi O, Marzari N (2013) Self-consistent continuum solvation (SCCS): the case of charged systems. J Chem Phys 139(21):214110

Enkovaara J, Rostgaard C, Mortensen JJ, Chen J, Dułak M, Ferrighi L, Gavnholt J, Glinsvad C, Haikola V, Hansen HA, Kristoffersen HH, Kuisma M, Larsen AH, Lehtovaara L, Ljungberg M, Lopez-Acevedo O, Moses PG, Ojanen J, Olsen T, Petzold V, Romero NA, Stausholm-Møller J, Strange M, Tritsaris GA, Vanin M, Walter M, Hammer B, Hakkinen H, Madsen GKH, Nieminen RM, Nørskov JK, Puska M, Rantala TT, Schiøtz J, Thygesen KS, Jacobsen KW (2010) Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. J Phys Condens Matter 22(25):253202

Fernández EM, Moses PG, Toftelund A, Hansen HA, Martínez JI, Abild-Pedersen F, Kleis J, Hinnemann B, Rossmeisl J, Bligaard T, Nørskov JK (2008) Scaling relationships for adsorption energies on transition metal oxide, sulfide, and nitride surfaces. Angew Chem Int Ed 47(25):4683–4686

Halck NB, Petrykin V, Krtil P, Rossmeisl J (2014) Beyond the volcano limitations in electrocatalysis – oxygen evolution reaction. Phys Chem Chem Phys 16:13682–13688

Hamada I, Sugino O, Bonnet N, Otani M (2013) Improved modeling of electrified interfaces using the effective screening medium method. Phys Rev B 88:155427

Hansen MH, Rossmeisl J (2016) pH in grand canonical statistics of an electrochemical interface. J Phys Chem C 120(51):29135–29143

Hansen HA, Rossmeisl J, Nørskov JK (2008) Surface Pourbaix diagrams and oxygen reduction activity of Pt, Ag and Ni(111) surfaces studied by DFT. Phys Chem Chem Phys 10(25):3722

Hansen MH, Nilsson A, Rossmeisl J (2017) Modelling pH and potential in dynamic structures of the water/Pt(111) interface on the atomic scale. Phys Chem Chem Phys 19:23505–23514

Henkelman G, Jonsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. J Chem Phys 113(22):9978–9985

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113(22):9901–9904

Herron JA, Morikawa Y, Mavrikakis M (2016) Ab initio molecular dynamics of solvation effects on reactivity at electrified interfaces. Proc Natl Acad Sci USA 113:E4937–E4945

Hori Y, Murata A, Takahashi R (1989) Formation of hydrocarbons in the electrochemical reduction of carbon dioxide at a copper electrode in aqueous solution. J Chem Soc Faraday Trans 85:2309–2326

Hu Z, Metiu H (2012) Halogen adsorption on CeO_2: the role of Lewis acid-base pairing. J Phys Chem C 116(11):6664–6671

Jinnouchi R, Anderson AB (2008a) Aqueous and surface redox potentials from self-consistently determined Gibbs energies. J Phys Chem C 112(24):8747–8750

Jinnouchi R, Anderson AB (2008b) Electronic structure calculations of liquid-solid interfaces: combination of density functional theory and modified Poisson-Boltzmann theory. Phys Rev B 77:245417

Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In Classical and Quantum Dynamics in Condensed Phase Simulations, Berne BJ, Ciccotti G, Coker DF (eds). World Scientific, pp 385–404

Ju W, Bagger A, Hao G-P, Varela AS, Sinev I, Bon V, Roldan Cuenya B, Kaskel S, Rossmeisl J Strasser P, (2017) Understanding activity and selectivity of metal-nitrogen-doped carbon catalysts for electrochemical reduction of CO_2. Nat Commun, 8(944):2041–1723

Karlberg GS, Rossmeisl J, Norskov JK (2007a) Estimations of electric field effects on the oxygen reduction reaction based on the density functional theory. Phys Chem Chem Phys 9:5158–5161

Karlberg GS, Jaramillo TF, Skúlason E, Rossmeisl J, Bligaard T, Nørskov JK (2007b) Cyclic voltammograms for H on Pt(111) and Pt(100) from first principles. Phys Rev Lett 99:126101

Kötz ER, Neff H, Müller K (1986) A UPS, XPS and work function study of emersed silver, platinum and gold electrodes. J Electroanal Chem Interfacial Electrochem 215(1):331–344

Larentzos JP, Criscenti LJ (2008) A molecular dynamics study of alkaline earth metal-chloride complexation in aqueous solution. J Phys Chem B 112(45):14243–14250

Larsen AH, Vanin M, Mortensen JJ, Thygesen KS, Jacobsen KW (2009) Localized atomic basis set in the projector augmented wave method. Phys Rev B 80(19):195112

Larsen AH, Mortensen JJ, Blomqvist J, Castelli IE, Christensen R, Dulak M, Friis J, Groves MN, Hammer B, Hargus C, Hermes ED, Jennings PC, Jensen PB, Kermode J, Kitchin JR, Kolsbjerg EL, Kubal J, Kaasbjerg K, Lysgaard S, Maronsson JB, Maxson T, Olsen T, Pastewka L, Peterson A, Rostgaard C, Schiøtz J, Schütt O, Strange M, Thygesen KS, Vegge T, Vilhelmsen L, Walter M, Zeng Z, Jacobsen KW (2017) The atomic simulation environment—a Python library for working with atoms. J Phys Condens Matter 29(27):273002

Li B, Metiu H (2012) Does halogen adsorption activate the oxygen atom on an oxide surface? I. A study of Br_2 and HBr adsorption on La_2O_3 and La_2O_3 doped with Mg or Zr. J Phys Chem C 116(6):4137–4148

Liang T, Cheng Y-T, Nie X, Luo W, Asthagiri A, Janik MJ, Andrews E, Flake J, Sinnott SB (2014) Molecular dynamics simulations of CO_2 reduction on Cu(111) and Cu/ZnO(1010) using charge optimized many body potentials. Catal Commun 52(Supplement C):84–87

Liu X, Xiao J, Peng H, Hong X, Chan K, Nørskov JK (2017) Understanding trends in electrochemical carbon dioxide reduction rates. Nat Commun 8:15438

Lozovoi AY, Alavi A, Kohanoff J, Lynden-Bell RM (2001) Ab initio simulation of charged slabs at constant chemical potential. J Chem Phys 115(4):1661–1669

Man IC, Su H-Y, Calle-Vallejo F, Hansen HA, Martinez JI, Inoglu NG, Kitchin J, Jaramillo TF, Nørskov JK, Rossmeisl J (2011) Universality in oxygen evolution electrocatalysis on oxide surfaces. ChemCatChem 3(7):1159–1165

Marković NM, Schmidt TJ, Stamenković V, Ross PN (2001) Oxygen reduction reaction on Pt and Pt bimetallic surfaces: a selective review. Fuel Cells 1(2):105–116

Ménétrey M, Markovits A, Minot C (2007) Adsorption of chlorine and oxygen atoms on clean and defective rutile-TiO_2 (110) and MgO (100) surfaces. J Mol Struct Theochem 808(1–3):71–79

Merlet C, Limmer DT, Salanne M, van Roij R, Madden PA, Chandler D, Rotenberg B (2014) The electric double layer has a life of its own. J Phys Chem C 118(32):18291–18298

Metiu H, Chrétien S, Hu Z, Li B, Sun X (2012) Chemistry of Lewis acid-base pairs on oxide surfaces. J Phys Chem C 116(19):10439–10450

Mistry H, Choi Y-W, Bagger A, Scholten F, Bonifacio C, Sinev I, Divins NJ, Zegkinoglou I, Jeon HS, Kisslinger K, Stach EA, Yang JC, Rossmeisl J, Cuenya BR (2017) Enhanced carbon dioxide electroreduction to carbon monoxide over defect rich plasma-activated silver catalysts. Angew Chem Int Ed Engl 56(38):11394–11398

Montoya JH, Tsai C, Vojvodic A, Nørskov JK (2015) The challenge of electrochemical ammonia synthesis: a new perspective on the role of nitrogen scaling relations. ChemSusChem 8(13):2180–2186

Mortensen JJ, Hansen LB, Jacobsen KW (2005) Real-space grid implementation of the projector augmented wave method. Phys Rev B 71(3):035109

Nie X, Luo W, Janik MJ, Asthagiri A (2014) Reaction mechanisms of CO_2 electrochemical reduction on Cu(111) determined with density functional theory. J Catal 312(Supplement C):108–122

Nørskov JK, Bligaard T, Logadottir A, Kitchin JR, Chen JG, Pandelov S, Stimming U (2005) Trends in the exchange current for hydrogen evolution. Electrochem Soc J 152(2):J23–J26

Nørskov JK, Studt F, Abild-Pedersen F, Bligaard T (2014) Fundamental concepts in heterogeneous catalysis. Wiley, Hoboken

Persson KA, Waldwick B, Lazic P, Ceder G (2012) Prediction of solid-aqueous equilibria: scheme to combine first-principles calculations of solids with experimental aqueous states. Phys Rev B 85:235438

Peterson AA, Abild-Pedersen F, Studt F, Rossmeisl J, Nørskov JK (2010) How copper catalyzes the electroreduction of carbon dioxide into hydrocarbon fuels. Energy Environ Sci 3:1311–1315

Pourbaix M (1974) Atlas of electrochemical equilibria in aqueous solutions. National Association of Corrosion Engineers, Pergamon Press: Oxford, London, Edinburgh, New York, Toronto, Paris, Frankfurt

Reuter K, Scheffler M (2001) Composition, structure, and stability of RuO_2 (110) as a function of oxygen pressure. Phys Rev B 65:035406

Reuter K, Scheffler M (2003) First-principles atomistic thermodynamics for oxidation catalysis: surface phase diagrams and catalytically interesting regions. Phys Rev Lett 90:046103

Rossmeisl J, Nørskov JK, Taylor CD, Janik MJ, Neurock M (2006) Calculated phase diagrams for the electrochemical oxidation and reduction of water over Pt(111). J Phys Chem B 110(43):21833–21839

Rossmeisl J, Skulason E, Björketun ME, Tripkovic V, Nørskov JK (2008) Modeling the electrified solid-liquid interface. Chem Phys Lett 466(1):68–71

Rossmeisl J, Karlberg GS, Jaramillo T, Nørskov JK (2009) Steady state oxygen reduction and cyclic voltammetry. Faraday Discuss 140:337–346

Rossmeisl J, Chan K, Ahmed R, Tripkovic V, Björketun ME (2013) pH in atomic scale simulations of electrochemical interfaces. Phys Chem Chem Phys 15:10321–10325

Sakong S, Gross A (2016) The importance of the electrochemical environment in the electro-oxidation of methanol on Pt(111). ACS Catal 6(8):5575–5586

Schnur S, Groß A (2009) Properties of metal-water interfaces studied from first principles. New J Phys 11(12):125003

Seh ZW, Kibsgaard J, Dickens CF, Chorkendorff I, Nørskov JK, Jaramillo TF (2017) Combining theory and experiment in electrocatalysis: insights into materials design. Science 355(6321):eaad4998

Sergeev AV, Chertovich AV, Itkis DM, Sen A, Gross A, Khokhlov AR (2017) Electrode/electrolyte interface in the Li-O_2 battery: insight from molecular dynamics study. J Phys Chem C 121(27):14463–14469

Shi C, Chan K, Yoo JS, Nørskov JK (2016) Barriers of electrochemical CO_2 reduction on transition metals. Org Process Res Dev 20(8):1424–1430

Siahrostami S, Verdaguer-Casadevall A, Karamad M, Deiana D, Malacrida P, Wickman B, Escudero-Escribano M, Paoli EA, Frydendal R, Hansen TW, Chorkendorff I, Stephens IEL, Rossmeisl J (2013) Enabling direct H_2O_2 production through rational electrocatalyst design. Nat Mater 12(12):1137–1143

Skulason E, Karlberg GS, Rossmeisl J, Bligaard T, Greeley J, Jonsson H, Norskov JK (2007) Density functional theory calculations for the hydrogen evolution reaction in an electrochemical double layer on the Pt(111) electrode. Phys Chem Chem Phys 9:3241–3250

Skulason E, Jonsson H, Nørskov JK, Bligaard T (2009) Electro-catalytic reactions using density functional theory calculations. PhD thesis, Technical University of Denmark (DTU), 10

Skulason E, Tripkovic V, Björketun ME, Gudmundsdottir S, Karlberg G, Rossmeisl J, Bligaard T, Jonsson H, Nørskov JK (2010) Modeling the electrochemical hydrogen oxidation and evolution reactions on the basis of density functional theory calculations. J Phys Chem C 114(42): 18182–18197

Spohr E (1989) Computer simulation of the water/platinum interface. J Phys Chem 93(16): 6171–6180

Strmcnik D, Uchimura M, Wang C, Subbaraman R, Danilovic N, van der Vliet D, Paulikas AP, Stamenkovic VR, Markovic NM (2013) Improving the hydrogen oxidation reaction rate by promotion of hydroxyl adsorption. Nat Chem 5:300–306

Trasatti S (1986) The absolute electrode potential: an explanatory note (recommendations 1986). J Electroanal Chem Interfacial Electrochem 209(2):417–428

Viswanathan V, Hansen HA, Rossmeisl J, Nørskov JK (2012) Universality in oxygen reduction electrocatalysis on metal surfaces. ACS Catal 2(8):1654–1660

Yeh K-Y, Janik MJ, Maranas JK (2013) Molecular dynamics simulations of an electrified water/Pt(111) interface using point charge dissociative water. Electrochim Acta 101(Supplement C):308–325

Electrocatalysis Beyond the Computational Hydrogen Electrode

61

Harald Oberhofer

Contents

Abstract

The computational hydrogen electrode approach allows the alignment of theoretical electrochemical potentials calculated with ab initio methods to those measured in experiment. It contributed greatly to opening up the fields of electrochemistry and photo-electrochemistry to theoretical treatment. Yet, virtually all practical implementations of the computational hydrogen electrode relied on a number of simplifications and approximations, which are not necessarily always justified. This chapter highlights three of these approximations as well as the challenges prompting them and gives a brief review of the computational methods available to overcome each. Specifically, it addresses the effects of the

H. Oberhofer (✉)
Chair for Theoretical Chemistry and Catalysis Research Center,
Technische Universität München, Garching, Germany
e-mail: harald.oberhofer@tum.de

electrolyte, the important choice of the model reactive site – including surface defects and co-catalysts – and the evaluation of kinetic barriers.

1   Introduction

The field of modeling electrochemical and photo-electrochemical processes made a tremendous leap forward with the introduction of the computational hydrogen electrode (CHE) approach in 2004 (Nørskov et al. 2004), with the name CHE coined later by Peterson et al. (2010). Thanks to its conceptual simplicity and computational efficiency, the CHE approach enabled both in-depth studies of reaction pathways and large-scale computational screening studies, dedicated to identifying optimal catalyst materials and facets. Thanks to its popularity, the CHE approach has received a number of updates and additions to account for effects not originally considered. This chapter is dedicated to illustrating the available computational approaches beyond the original implementations of the CHE. Thereby, many of the examples give focus on the water splitting reaction, one of the most ubiquitous and well-studied processes in electrocatalysis, yet all methods can equally be applied to other (photo-)electrochemical processes.

In essence, one needs to model the removal of two protons and electrons from a H_2O molecule which go on to form a hydrogen molecule. It is generally thought of in the form of two half-reactions, the water oxidation

$$H_2O \rightarrow \frac{1}{2}O_2 + 2H^+ + 2e^- \tag{1}$$
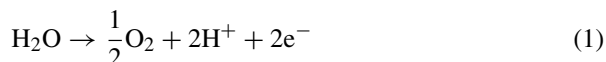
and the hydrogen evolution

$$2H^+ + 2e^- \rightarrow H_2, \tag{2}$$

which occur on anode and cathode, respectively.

Any simulation of complex, multistep reactions such as water splitting necessarily rest on an understanding of the involved reaction pathways. Although it is conceivable to explore the space of all possible reaction steps leading from an initial state (such as water) to a number of final products (molecular oxygen and hydrogen), such an approach is rarely feasible for more complex reactions and catalysts. Thus, studies more commonly rely on the assumption of certain predetermined reaction pathways going through one or more reaction intermediates. For the process of water oxidation, such intermediates could, for example, be OH^*, O^*, and OOH^*, where the asterisk denotes a species adsorbed on the catalyst surface. The task of the modeling scientist is then to determine how a reaction can proceed between those, given the associated energetic cost.

The CHE method, described in more detail in Sect. 1.1, is a way to assess the free energy profile of a (photo-)electrochemical reaction at a given catalyst surface. Yet, most of its practical realizations rely on a number of approximations and

simplifications which, as outlined in Sect. 1.2, are not necessarily justified in all cases. The first of these is the omission of any direct solvent effects on the binding free energies of reaction intermediates. Considering that these intermediates might be highly polar – in photo-electrochemical setups even charged – solvation could significantly alter the free energy profiles. In Sect. 2, we outline both explicit and implicit approaches to including solvation effects in (photo-)electrocatalysis and highlight their strengths and weaknesses.

Another hurdle in many simulations at the solid-liquid interface is the determination of the nature of the reactive site. As discussed in Sect. 3, realistic catalysts in reaction conditions are rarely pristine surfaces found, e.g., in vacuum conditions. Instead, they show all manners of non-periodicities such as defects, terraces, or edges, and computational methods to treat these are outlined in Sect. 3.1. In this context, we also examine co-catalysts, i.e., small molecules or metal or oxide clusters, usually employed in photo-electrocatalytic setups. Therefore, an overview over theoretical efforts to treat co-catalysts is given in Sect. 3.2.

Finally, most free energy profiles calculated with the CHE are composed of the relative free energy differences between the initial state, the intermediates, and the final state of a given reaction, respectively. While this gauges the ability of a reaction to proceed at given reaction conditions or get stuck along the way, it assumes that kinetic barriers between these states are negligible. In Sect. 4, we review the techniques and results of first calculations based on explicitly dynamical simulations, showing this to not always be the case. The chapter is finally rounded off by a summary and synopsis in Sect. 5.

1.1 The Computational Hydrogen Electrode

In analogy to earlier work on heterogeneous catalysis (Reuter and Scheffler 2001; Reuter 2016), the revolutionary CHE method of Nørskov and Rossmeisl renders a thermodynamic picture for (photo-)electrocatalysis (Nørskov et al. 2004; Valdes et al. 2008; Peterson et al. 2010). Given a reaction pathway proceeding through a number of intermediates, one can determine the free energy difference between them in order to gauge if a reaction can continue to the desired final state, deviate to another product, or get stuck along the way. The approach taken by Nørskov and Rossmeisl was to formulate each of these steps as the removal (or indeed addition) of a proton and an electron. In the case of the water oxidation half-reaction, the pathway commonly employed reads (Nørskov et al. 2004; Valdes et al. 2008):

$$A: H_2O + {}^* \rightarrow OH^* + H^+ + e^- \tag{3a}$$

$$B: OH^* \rightarrow O^* + H^+ + e^- \tag{3b}$$

$$C: H_2O + O^* \rightarrow OOH^* + H^+ + e^- \tag{3c}$$

$$D: OOH^* \rightarrow O_2 + {}^* + H^+ + e^-. \tag{3d}$$
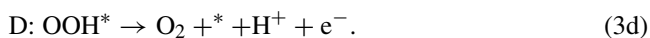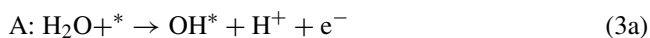
Here, and in the following, a single asterisk * denotes an empty reactive site on the catalyst. The nature of this site will be discussed in more detail in Sect. 3. At each step, a proton H^+ and an electron e^- are abstracted from the adsorbates, in principle to partake in the other half-reaction. As this half-reaction is omitted from the calculation for practical reasons – i.e., to keep system sizes computationally tractable – protons and electrons can be thought of being put into a reservoir. The free energy cost or gain ΔG_X for each step X in the oxidation reaction can then be expressed as

$$\Delta G_A = G_{OH*} - G_{H_2O} - G_* + \mu_{H^+} + \mu_{e^-} \tag{4a}$$

$$\Delta G_B = G_{O*} - G_{OH*} + \mu_{H^+} + \mu_{e^-} \tag{4b}$$

$$\Delta G_C = G_{OOH*} - G_{OOH*} - G_{H_2O} + \mu_{H^+} + \mu_{e^-} \tag{4c}$$

$$\Delta G_D = G_{O_2} + G_* - G_{OOH*} + \mu_{H^+} + \mu_{e^-}. \tag{4d}$$

The chemical potentials of the removed protons and electrons in their reservoirs are given as μ_{H^+} and μ_{e^-}, respectively. The Gibbs free energies of the involved species are written as G_X.

From a modeling point of view, the calculation of the free energies of surface adsorbed species is relatively straightforward with, for example, density functional theory (DFT) (Stampfl et al. 2002; Hu et al. 2007), while for the gas phase, they can often be found in chemical tables (Kandoi et al. 2004; Haynes 2014; Aktins and de Paula 2014). Each of its three main contributions, the energy E, the entropy S, and a zero point energy correction ZPE, has been studied in standard surface science calculations, with E taken as the total DFT energy of the system and S and ZPE determined from the system's phonon spectrum. The chemical potentials of proton and electron reservoirs are, on the other hand, much less straightforward to determine in ab initio calculations. Yet, especially for a comparison to experiment, calculated energies need to be aligned with respect to known quantities via the reservoir potentials. As a remedy to this dilemma, Rossmeisl and Nørskov put forward the transformative idea of following a route analogous to experiments. There, to allow a comparison of measurements between different experimental setups and electrodes, potentials are aligned with respect to a well-defined standard reaction, very often the formation of hydrogen

$$H^+ + e^- \leftrightarrows \frac{1}{2}H_2. \tag{5}$$

Depending on the reaction conditions, this is called the standard or the normal hydrogen electrode and provides a known potential to align measurements against. In computer simulations, the explicit inclusion of a further electrode into the simulation box is obviously not practical. Yet, observing that each step is formulated as simultaneous proton and electron transfers, an analogous alignment of free

energies can be achieved by identifying the chemical potentials of protons and electrons with the free energy G_{H_2} of a hydrogen molecule

$$\mu_{H^+} + \mu_{e^-} = \frac{1}{2} G_{H_2}. \tag{6}$$

This identity can directly be inserted into Eq. 4 for free energy differences comparable with experiment.

While this *computational* hydrogen electrode allows an alignment of potentials with experiment, there are still a number of contributions missing. First, the electron reservoir is the respective electrode the reaction is occurring on. As such, its chemical potential depends strongly on any external potentials U that could be applied to the electrode. This manifests as a term $-eU$ added to the free energy of each step which essentially provides the driving force for the reaction. Similar corrections apply in the case of photo-electrochemistry on semiconductors, where the electron reservoir is a photo-generated electron hole at the surface (Valdes et al. 2008).

The last component missing so far is the effect of pH on the reaction free energy due to changes in the concentration of unbound protons in solution. This contribution is generally taken as $k_B T pH \ln 10$ and simply added to the total free energy for each reaction step. This simple dependence is not necessarily always justified. Koper, for example, showed (Koper 2013a, b; Göttle and Koper 2017) that an uncoupling of proton and electron steps could lead to more complex dependence of the free energy on pH. Note that this is not a failure of the CHE approach per se, as the reservoir potentials could simply be defined with respect to the experimental *reversible* hydrogen electrode as demonstrated by Peterson et al. (2010). In this case, pH dependence trivially vanishes.

In any case, the restriction to simultaneous proton and electron abstraction steps seems like a severe one at a first glance. After all, both processes can occur on wildly different timescales and are influenced by different factors. Yet, for a purely thermodynamic viewpoint, the ordering of these processes is not important because of the properties of the free energy as a thermodynamic state function. This means that the exact path taken between two states is irrelevant for the free energy difference between them. In the (photo-)electrochemical case, this means that the free energies calculated with Eq. 4 hold, independent of the formulation of the step as a combined proton-coupled electron transfer or sequential process. As long as the true reaction pathway proceeds at some point through the same intermediates (cf. Fig. 1), ΔG_{A-D} can be a valuable estimates.

The sequence of free energy differences can, for example, be used to estimate the minimum potential U_{min} required to drive the reaction. Generally, the effect of the potential is thereby to lower ΔG_X by eU for each step X. At the minimum potential $U = U_{min}$, the free energy cost ΔG_X is either zero or negative such that the overall reaction can occur unhindered without additional thermodynamic barriers. For water oxidation to O_2, the overall energetic cost is known from experiments to be 4.92eV.
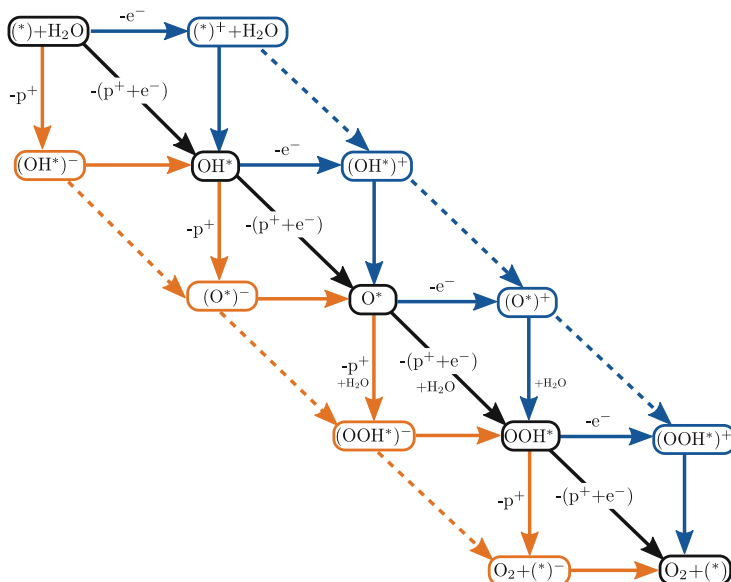
Fig. 1 Schematic diagram of the four steps of water oxidation including coupled (black and colored dashed lines) and uncoupled proton and electron pathways. Orange pathways proceed by first removing a proton and then an electron, while for pathways in blue first an electron is removed before the proton follows. Colored dashed lines depict coupled pathways that proceed solely through charged intermediates. The free energy being a state function, ΔG of each intermediate (boxes) is independent of the path taken to arrive there. In principle the reaction could proceed through any combination of the depicted pathways

An ideal catalyst would thereby distribute this cost over all four steps such that each of them only needs 1.23 eV to proceed, yielding an ideal catalyst potential of $U_{\text{ideal}} = 1.23\text{V}$. Realistic catalysts on the other hand differ from ideal behavior, leading to some steps needing considerable more energy than 1.23eV and others much less. The minimum necessary potential, though, needs to lower all ΔG_X to zero or less, resulting in a U_{min} significantly larger than U_{ideal}. The difference between U_{min} and U_{ideal} is known as the overpotential which essentially measures the energy lost due to less than ideal catalysts. Consequently, a lot of research to date, both experimental (Halter et al. 2018; Qian et al. 2017; Maayan et al. 2018; Thiyagarajan et al. 2018) and theoretical (Reda et al. 2018; Seh et al. 2017; Seitz et al. 2016; Zhang et al. 2016), has focused on designing catalysts with minimal overpotential.

Finally, the CHE approach can also serve as the basis of computational screening of novel catalyst materials. Utilizing the observation that binding energies of OH and OOH adsorbates are correlated at least in metallic systems – the so-called scaling relations (Abild-Pedersen et al. 2007) – allows the use of just one binding energy as a screening descriptor. A more detailed discussion of this approach can be found in ▶ Chap. 54, "Energy Trends in Adsorption at Surfaces" of this book.

1.2   Approximations Commonly Made in CHE-Based Calculations

While the CHE approach so far has undeniably led to great advances in our understanding of electrochemistry, most of its implementations so far are not without room for improvement. Foremost, standard applications of the CHE, being a posteriori corrections, where the effects of the applied potential and the pH are added to the energetics of potential-free calculations, cannot address their influence on the catalyst surface. This can lead to an erroneous representation of (photo-)electrochemistry in alkaline environments (Rossmeisl et al. 2013).

The problem of simulating pH in electrochemical environments has been tackled in recent work by Rossmeisl and co-workers (Nielsen et al. 2015; Hansen and Rossmeisl 2016; Hansen et al. 2017), where they presented a generalization of the computational hydrogen electrode (GCHE) which treats the overall electrochemical potential as a function of the applied electrostatic potential and the chemical potential of protons and electrons. As an input, this approach still only takes energetics but for different numbers of hydrogen atoms (protons and electrons) in the system. A more detailed account of the GCHE is given in ▶ Chap. 57, "Microkinetic Modeling of Surface Catalysis" of this book.

Furthermore, the assumed pathways do not necessarily represent the lowest energy ones, potentially leading to an overestimation of the overpotential (Koper 2013a, b). Also, especially in photo-electrochemistry on semiconducting electrodes, other pathways might proceed via charged intermediates (Göttle and Koper 2017) which to a large part are stabilized through the solvent response, often neglected in CHE-based calculations. The reason for this is twofold. First, the computational cost associated with the then necessary sampling of solvent degrees of freedom can easily render simulations of realistically sized systems infeasible. Second, the actual structure of the solid-liquid interface is not only strongly dependent on the choice of the catalyst material and facet but also on the applied potential which is not always readily controllable in computer simulations. To avoid both of these complications, a lot of recent effort was devoted to the development of sufficiently accurate effective simulation methods, so-called implicit solvation models.

On top of that, as with most computational studies, CHE-based calculations can depend strongly on the chosen model for the reactive site (Rao et al. 2017). On ideal flat surfaces, this is less of a problem as the sites there tend to be equal. Famous counterexamples to this do exist with e.g., the CO-adsorption "puzzle" at transition metal surfaces (Hu et al. 2007). Yet, under realistic conditions, no catalyst truly is an ideal single crystal surface. Rather, in order to maximize surface area, catalysts are present as nano-particles leading to a high number of defects, edges, and steps which can significantly alter the material's reactivity (Gambu et al. 2017). Additionally, experimental catalysts often employ small metallic nano-clusters as so-called co-catalysts to facilitate the reaction. Furthermore, oxide materials, which are commonly employed as photo-electrocatalysts, are prone to surface defect formation under reaction conditions (Berger et al. 2015), which can also influence predicted reaction free energies. Neglecting any of these properties of the catalyst can again lead to an erroneous estimate of the energy necessary to drive the reaction.

Recent studies therefore have begun to shed some light on the precise nature of the reactive site (Reuter et al. 2017).

Finally, CHE-based studies of (photo-)electrochemical reactions only consider overpotentials due to the thermodynamic free energy differences between intermediates, neglecting any kinetic barriers between them. These might in turn influence the respective minimum free energy pathway and thus lead to a completely different picture of the reaction. The main reason these have rarely been considered is again the computational cost associated with the dynamical calculations necessary to determine free energy profiles and thus barriers from first principles. With the advent of more elaborate simulation methodology and increased computing power, recently a number of studies of kinetic barriers in (photo-)electrochemistry have been put forward.

These three improvements over standard CHE calculations – i.e., solvation effects, nontrivial reactive sites, and free energy barriers – as well as the methodologies necessary to effectively study them, will be discussed in the following sections.

2 Solvation Effects

As already alluded to above, a simplification made in most early CHE calculations was to omit solvents and solvation effects from the DFT calculations of the energies of reaction intermediates (Nørskov et al. 2004; Valdes et al. 2008). The justification behind this omission is that due to the calculation of only potential-free neutral systems in a CHE-based calculation, solvent effects can be expected to play a minor role in the energetics of surface reactions. Also, up until recently even the basic structure of the solid-liquid interface was still under heavy debate (Feibelman 2002; Michaelides et al. 2004; Meng et al. 2004; Schnur and Groß 2009; Groß et al. 2014) and thus could not easily be tackled in computer simulations. An in-depth discussion of the solid-liquid interface can be found in ▸ Chap. 59, "Structure of Electrode-Electrolyte Interfaces, Modeling of Double Layer and Electrode Potential" of this book. Nowadays, though, the importance of an accurate description of the interface, also with a view on the structure of the reactive site discussed in Sect. 3, is undisputed (Filhol and Neurock 2006; Janik et al. 2007; Filhol and Doublet 2013; Siahrostami and Vojvodic 2015; Hansen and Rossmeisl 2016), and a number of methods have been put forward to tackle solvation, both explicitly and based on more approximate models. An account of both kinds of approaches, insofar as they are relevant for (photo-)electrochemical simulations, is given below.

2.1 Explicit Approaches

At a first glance, explicit simulation of solvents on a catalyst surface seems relatively straightforward. Considering, for example, metal catalysts, it has long been known that water layers closest to the metal surface tend to exhibit relatively ordered,

icelike structures due to strong surface-solvent interactions (Ogasawara et al. 2002; Schnur and Groß 2009; Groß et al. 2014). As a first improvement over vacuum calculations, many studies therefore employ a few layers of highly ordered icelike water (Skúlason et al. 2007, 2010; Chan and Nørskov 2015). This is achieved by a local geometry optimization of the water layer(s) for a given distribution of first-layer water molecules with hydrogen atoms either pointing toward or away from the surface. The exact distribution of "up" to "down" waters and thus the structure of the ice layer are thereby governed by the applied potential (Filhol and Doublet 2013). Potential-free calculations as performed in "standard" CHE calculations thus would not reproduce the correct behavior of the surface water layer. As a remedy, a number of approaches to simulate potentials in electrochemical calculations have been put forward in recent years. They fall, broadly, into two categories, one which directly includes an electrostatic potential dragging electrons from one side of the catalyst slab to the other (Otani et al. 2008; Otani and Sugino 2006) and one where the catalyst surface is charged by explicitly adding electrons to the simulation box. In this case, in order to keep the simulation box overall charge neutral, counter-charges have to be applied either directly in the solvent layer (Skúlason et al. 2007) or in the form of a homogeneous charge background (Filhol and Neurock 2006). In both cases, the applied potential can then be calculated from the work function of the solvated catalyst surface in relation to the vacuum level, which is generally taken as the potential in the vacuum far above the thin, icelike solvent (Chan and Nørskov 2015).

Such more or less ordered layers of water with low molecular mobility not only exist on simple metal surfaces but have also been observed in other materials such as photo- and electrochemically highly relevant phosphides (Wood et al. 2013) and oxides (Sumita et al. 2010; Zhao et al. 2012). Yet, where on metals long range ordered, hexagonal structures dominate, the more complex structure of oxides gives rise to only a local ordering, with, on average, higher lateral molecular mobility (Tilocca and Selloni 2004). This, nevertheless, still suggests that a first description of solvation effects on oxides can be achieved by adding a number of icelike solvent layers to the catalyst surface (Siahrostami and Vojvodic 2015). Indeed, recent CHE-based calculations of, e.g., the water oxidation reaction on IrO_2 (Gauthier et al. 2017) showed a pronounced influence of solvation on the stabilities of the reaction intermediates OH^* and OOH^* (cf. Sect. 1.1), which are both polar and capable of donating hydrogen bonds. Of these two properties, the latter was found to be of greater importance for the relative binding energies, as they showed very little dependence on solvent layer thickness beyond one bilayer, which suggests that electrostatic effects play only a limited role here.

While the ice-layer approach has been shown to yield very good results at only insignificantly increased cost compared to vacuum simulations, there definitely are systems where such an approach is not applicable. For example, electrodes with only weak or intermediate water-binding strengths – which could become relevant in ongoing efforts to break scaling relations or increase product selectivity – showed a preference of more liquid than icelike behavior (Groß et al. 2014). In such cases, due to the disordered nature of the solvent "layers" near the surface, static total energy

calculations are not applicable, and a full sampling of solvent degrees of freedom cannot be avoided. Indeed, a great number of studies employed DFT-MD to model the solid-liquid interface also for electrode materials attracting more ordered water structures.

Sprik and co-workers, for example, employ dynamical sampling to investigate the alignment of electronic and protonic energy levels with respect to water (Cheng and Sprik 2010b; Cheng et al. 2014a) as well as acidity constants of oxides (Cheng and Sprik 2010a), based on a dynamical formulation of the computational hydrogen electrode calculated from the proton work function in a large water box (Cheng et al. 2014b). Similarly, a number of groups Chen et al. (2013), and Stecher et al. (2016) used DFT-MD sampling with explicit solvent degrees of freedom to investigate kinetic barriers in electrochemical processes on cluster models (Mattioli et al. 2013) or extended surfaces (Chen et al. 2013; Stecher et al. 2016). A detailed description of the thereby applied methods will be given in Sect. 4. Along similar lines, Otani et al. employed explicit solvation sampling in a semi-infinite simulation box, giving rise to well-defined electrode potentials (Otani and Sugino 2006), to investigate the dynamics of the hydrogen evolution reaction on platinum (Otani et al. 2008). They find non-negligible influence of both electrode potential and solvent dynamics on the reaction pathway, further underlining the importance of solvation effects. Finally, explicit solvent dynamics was applied by Hansen and Rossmeisl (Hansen et al. 2017) to investigate pH dependence of the electrode solvent interface in a grand canonical approach.

2.1.1   Ion Effects

Up to this point, the discussion of solvation effects was restricted to pure water, neglecting solvated ions present in electrolytes. In electrochemical solid-liquid interfaces, these play the role of mobile counter-charges which are attracted by the buildup of charge in the electrode due to the applied potential, forming the so-called electric double layer (EDL). Due to the presence of charge carriers of opposite signs in solid and liquid phase, respectively, the EDL is characterized by strong electric fields, which themselves can have non-negligible influences on reactions within the interface. Unfortunately, the simulation of the EDL is hampered by two finite size effects. On the solid side of the interface, experimentally realistic field strengths would lead to surface charges of only fractions of an electron per surface unit cell (Rossmeisl et al. 2008). Thus, reactions involving the transport of charges to or from the electrode would lead to large jumps in the potential. Similarly, in the liquid phase, realistic ionic strengths would necessitate very large water boxes. Due to the condition of charge neutrality of the EDL, arising from the fact that electrolyte is conductive (Cheng and Sprik 2014), both these problems are related to the applied electrode potential and can be tackled with the methods discussed in the introduction to this section.

In the literature, explicit simulations of solid-liquid interfaces containing ions have been carried out both in the icelike interface approximation (Chan and Nørskov 2015) and fully dynamical simulations (Cheng and Sprik 2014). Both approaches compare favorably with regard to the respective experiments but are

computationally more expensive than plain CHE calculations in vacuum due to added solvent and ionic degrees of freedom. Even in an icelike approximation, the position of the solvated ions needs to be sampled to yield an authentic representation of the EDL (Rossmeisl et al. 2008). The added computational cost all but precludes high-throughput screening of catalyst materials. The next section will therefore focus on more approximate, but potentially much cheaper continuum approaches to solvents and ions.

2.2   Implicit Approaches

Historically, the idea to treat the electrostatic response of a pure solvent implicitly as a polarizable continuum goes back to works of Kirkwood (1934) and Onsager (1936) and therefore precedes most modern computational methods by far. Based on the assumptions of isotropy and homogeneity of the solvent and linearity of its response to electrostatic perturbations, it is possible to determine the total electrostatic potential $\Phi(\mathbf{r})$ of a solute with charge distribution $\rho_{\text{sol}}(\mathbf{r})$ embedded in a dielectric continuum from a generalized Poisson equation (Tomasi et al. 2005):

$$\nabla \varepsilon(\mathbf{r}) \nabla \Phi(\mathbf{r}) = \rho_{\text{sol}}(\mathbf{r}). \tag{7}$$

The function $\varepsilon(\mathbf{r})$ thereby denotes the dielectric permittivity of the system, which far away from the solute takes on the solvent's macroscopic bulk value (e.g., $\varepsilon(\mathbf{r} \to \infty) \approx 80$ for water) but is equal to one in the vicinity of the solute. This reflects the fact that solvent molecules carrying the response can only approach the solute to within a finite distance, thus forming a so-called solvation cavity (cf. Fig. 2).

Though seemingly quite simple, Eq. (7) gives rise to a multitude of different models and methods. The dielectric function, for example, not only mediates
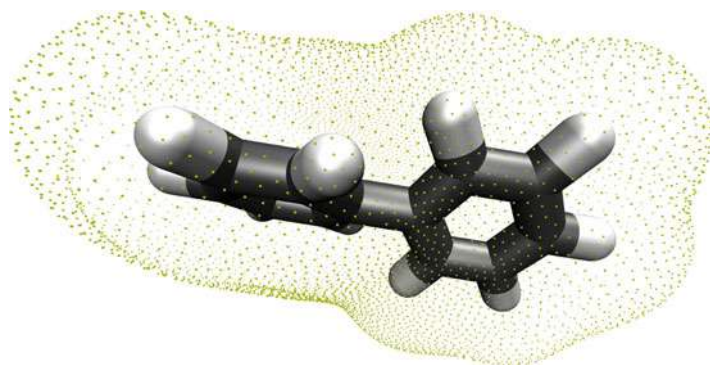
Fig. 2 Sketch of a solvation cavity based on an electron iso-density surface for biphenyl. Discrete, equidistant points on the cavity surface (green dots) were generated for the MPE model with a pseudo-dynamic simulation (Sinstein et al. 2017) in the FHI-aims code (Blum et al. 2009)

between solute and solvent but in principle also has to include entropic effects due to the loss of degrees of freedom solvent molecules undergo near the solute, e.g., forming solvation shells. As such $\varepsilon(\mathbf{r})$ is always an effective parameter of the solvation model and can be chosen in a number of ways, the two most prominent being a hard step function or a soft transition, both of which can again be defined in a number of ways, such as iso-surfaces of the density (Fig. 2) (Andreussi et al. 2012; Sinstein et al. 2017), soft spheres (Fisicaro et al. 2017), or spheres with van der Waals radii (Klamt and Schüürmann 1993). Furthermore, there are also numerous ways to define the basis used to represent the electrostatic response, ranging from two-dimensional charge distributions on the solvation cavity surface (Klamt and Schüürmann 1993), to multipole moments (Sinstein et al. 2017) and real-space grids (Mathew et al. 2014). Given this great wealth of methodological options, we direct the interested reader to the excellent review by Tomasi et al. (2005) for a more detailed discussion of solvation methods.

As an illustration of the inner workings of an implicit solvent model, we here give a short description of the recently reformulated multipole moment expansion (MPE) model (Sinstein et al. 2017). Based on the linearity of the Poisson equation and a stepwise transition in $\varepsilon(\mathbf{r})$, the problem is split into an inner part with $\varepsilon_{\text{in}} = 1$ and an outer part where the permittivity takes on the solvent's macroscopic bulk value (ε_{b}).

$$\Phi(\mathbf{r}) = \begin{cases} \Phi_{\text{int}}(\mathbf{r}) = \Phi_H(\mathbf{r}) + \Phi_R(\mathbf{r}), & \text{inside,} \\ \Phi_{\text{ext}}(\mathbf{r}) = \varepsilon_{\text{b}}^{-1}\Phi_H(\mathbf{r}) + \Phi_Q(\mathbf{r}), & \text{outside.} \end{cases} \tag{8}$$

Here, $\Phi_R(\mathbf{r})$ and $\Phi_Q(\mathbf{r})$ denote the potential inside and outside of the solvation cavity due to the polarization of the solvent. Both the internal and the external potential are regularized with $\Phi_H(\mathbf{r})$, which is the electrostatic potential of the solute in the absence of the solvent. The latter is, due to the necessity to solve the Schrödinger equation, generally known in any quantum chemistry code anyway. The solution of Eq. (7) then reduces to determining only the polarization potentials $\Phi_{R,Q}$ which are generally smoother and, most importantly, free of Coulomb singularities from the solute's nuclei. Additionally, the specific regularization given here was shown to completely eliminate errors stemming from charges outside of the solvation cavity common to other approaches (Sinstein et al. 2017). In the MPE, the three potentials $\Phi_{H,R,Q}(\mathbf{r})$ are first expanded in multicentered multipole moments up to a certain order. The unknown coefficients of $\Phi_R(\mathbf{r})$ and $\Phi_Q(\mathbf{r})$ are then determined via boundary conditions on the full potential and its derivative at a number of points on the cavity surface. This procedure yields an approximation to the electrostatic contribution to the solvation free energy with only the shape of the cavity – or in the case of a density iso-surface the iso-value – as a parameter. Additional parameters can be introduced with corrections for non-electrostatic contributions, such as van der Waals interactions or entropic contributions of the cavity (Scherlis et al. 2006; Andreussi et al. 2012).

This means that this model contains a number of free parameters (two or three at a minimum (Sinstein et al. 2017)) which need to be determined, e.g., by fitting (Andreussi et al. 2012; Ringe et al. 2016; Sinstein et al. 2017) to a sufficiently large database of experimental solvation free energies (Shivakumar et al. 2010; Marenich et al. 2012). While such a procedure works very well for molecules in solution, it leads to a problem for the use of implicit solvation methods in (photo-)electrochemistry due to the lack of appropriate reference data. Transferability of solvation parameters gained from molecular solvation free energies to extended surfaces is still a topic of ongoing research. Furthermore, it was shown for the example of the MPE model (Sinstein et al. 2017) that also for small molecules, the choice of functional (e.g., generalized gradient or hybrid functional) exerts some influence over the optimal choice of solvation parameters. This was found to be partially due to the different extents of the electron density given by different functionals but also due to the implicit solvent model canceling some DFT errors. Examining cross-correlations between parameter sets and functionals, a most general set was found for the HSE06 functional yielding mean absolute errors (MAE) smaller than 50 meV for neutral solutes. The same set of parameters still gives a reasonable MAE of $\approx$85 meV for cationic solutes but fails completely for anions ($\approx$500 meV). This behavior is in fact not restricted to the MPE model alone but common among implicit solvation models and especially pronounced in those relying on a density-based solvation cavity (Dupont et al. 2013). Recent work showed that this failure can be ameliorated by employing fixed soft spheres around the solute's atoms (Fisicaro et al. 2017). Nevertheless, also with the added charge-independent degrees of freedom awarded to the cavity generation by using soft spheres, the authors found that no single parameter set can satisfactorily describe both neutrals and cations on one hand and anions on the other.

The reason for this ultimately lies in the simplicity of both the employed electrostatic and non-electrostatic model. Virtually all continuum models show for identical cavities a symmetry of the electrostatic response with respect to positive or negative charges. Additionally, in the simple non-electrostatic model employed here, the response is completely independent of solute charge, as it only depends on the properties of the cavity. Especially for water, such a behavior does not reflect the physical realities (Latimer et al. 1993; Mobley et al. 2008) as sketched in Fig. 3 and thus has to be compensated for in implicit solvation models either through separate sets of parameters for cations, neutrals, and anions or through explicit treatment of such asymmetries (Bardhan and Knepley 2014).

We will return to the problem of parametrization again at the end of this section. First, though, it is important to note here that, while potentially yielding a good description of solvation effects in pure polar media, Eq. (7) cannot, by construction, describe electrolytes. In the form given in Eq. (7), the only charge carriers accounted for are those of the solute, while the solvent is assumed to be free of charges. An obvious fix for this is to simply augment the right-hand side with an additional charge density term $\rho_{\text{ion}}(\mathbf{r}; \Phi(\mathbf{r}))$ representing ions in solution.

$$\nabla \varepsilon(\mathbf{r}) \nabla \Phi(\mathbf{r}) = \rho_{\text{sol}}(\mathbf{r}) + \rho_{\text{ion}}(\mathbf{r}; \Phi(\mathbf{r})). \qquad (9)$$
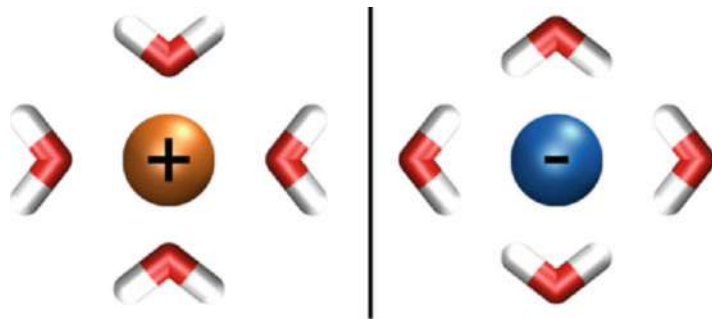
Fig. 3 Sketch of the first solvation shell of cations (left) and anions (right). Differently charged solutes lead to very different solvent responses, which are usually not captured by implicit solvation models. In the two solvation motifs shown, each water molecule has very different freedoms to move, which would be reflected in the local dielectric permittivity

Yet, as ions are mobile in solution, they will arrange themselves according to the electrostatic potential. Assuming that they are in equilibrium with both the solvent and the solute, the ion densities can be represented through their Boltzmann distribution function

$$\rho_{\text{ion}}(\mathbf{r}; \Phi(\mathbf{r})) = c_+ e^{-\beta z_+ \Phi(\mathbf{r})} - c_- e^{\beta z_- \Phi(\mathbf{r})}, \tag{10}$$

where $c_{\pm}$ and $z_{\pm}$ denote concentration and charge of cations and anions, respectively, and $\beta = 1/k_B T$ is the inverse temperature. This modification of the Poisson equation is then known as Poisson-Boltzmann theory (Chapman 1913; Gouy 1917) and has found many applications in literature (Honig and Nicholls 1995; Fang and Liu 2010; Boda et al. 2002; Bohinc et al. 2012).

Yet, it is important to note that even within the above-discussed assumptions (homogeneity, isotropy, linear response, and equilibrium), Eq. (10) constitutes an approximation to the actual equilibrium ion density, as it neglects non-electrostatic interactions between solutes and ions. For example, ions described by Eq. (10) could get arbitrarily close to the solute and each other, which is clearly unphysical. Modern implementations of Poisson-Boltzmann theory (Andreussi et al. 2012; Fang et al. 2013; Mathew et al. 2014; Ringe et al. 2016) thus introduce additional constraints on the ionic densities to correct for finite sizes of the ions (Borukhov et al. 1997) – including their own solvation shells – and the fact that due to their solvation shells ions can only approach solutes up to a finite distance forming a so-called Stern layer (Stern 1924). While these modifications restore the correct physical behavior of the ion densities, they come at the price of introducing additional free parameters, which again need to be determined beforehand.

For example, in the case of the Stern layer and size-modified Poisson-Boltzmann equation (Ringe et al. 2016), there are in total four parameters defining the solute-solvent interactions (cf. also the MPE model described above) and three ionic parameters. While these could, given a large enough training set, be determined by

fitting them all together at the same time, such an approach would bear the dangers of over-fitting, i.e., a severe lack of transferability between solvents and salts. The common approach thus is to first fit non-ionic parameters for calculations in ion-free solvents (Andreussi et al. 2012; Ringe et al. 2016; Sinstein et al. 2017) and then keep these fixed while determining ionic parameters. While for ion-free solvents there are a number of databases collecting experimental solvation free energies (Shivakumar et al. 2010; Marenich et al. 2012), reference data is decidedly more sparse for the ionic case, especially considering that there needs to be data available at different ionic concentrations as well. A solvent model should reproduce properties not just at a single ion concentration but over a certain range to be useful for electrochemical simulations. Recent approaches to determining ionic parameters have thus focused on Setchenow coefficients (Setchenow 1892), which describe the linear decrease or increase of the solvation free energy with salt concentration. This relationship tends to hold for salt concentrations up to 2–5M with corresponding coefficients being available for a wide range of molecular solutes (Long and McDevit 1952; Pérez-Tejeda et al. 1987; Keeley et al. 1988; Waxman et al. 2015). Studies based on the SMPB model showed (Ringe et al. 2017) not only that implicit models can reproduce Setchenow coefficients and thus relative solvation free energies very well but also that ionic parameters correlate with the properties of the salt. Specifically, they demonstrated that the SMPB ionic parameters can be scaled from well-characterized systems such as NaCl electrolytes with the respective experimentally accessible hydration number of the salt (Marcus 1985). The reason for this is that the hydration numbers reflect the size of the ion with its solvation shell and thus yield a measure of the sizes of the respective Stern layers (Kiriukhin and Collins 2002).

Finally, the local solvation environment around molecules, to which all these parametrizations have been fitted, does not necessarily reflect the properties of the EDL on a surface. The extent to which existing parameter sets can be transferred to (photo-)electrochemical systems is thus not entirely clear. Future work in this direction is clearly needed.

In summary, with explicit and implicit solvation methods, there are two distinct approaches to including solvent effects in (photo-)electrochemical simulations. While explicit methods have potentially greater accuracy due to their trivial inclusion of non-mean-field effects such as hydrogen bonding, they are, for sufficiently realistic system sizes and density functionals, rather expensive. Implicit approaches are, on the other hand, relatively cheap and not limited to certain (high) ion concentrations but need to be carefully parametrized to the specific system. Thus, the optimal choice here likely depends on the question to be answered.

3   Reactive Sites

In Eq. 4 and the following, reactive sites and molecules adsorbed at them were simply denoted by an asterisk (*). All structural information of the reaction is included in this one character. This notation, while undeniably convenient and

useful (Reuter et al. 2017), somewhat obfuscates the complexity of the underlying situation. It is well known, that the exact choice of model for the reactive site critically influences (Neurock 2003) the binding energies of reaction intermediates and thus any assessment of the reactivity of a catalyst material. Even pure metallic single crystal surfaces already present more than one possible binding site, including different facets. In more complex but chemically also more promising systems such as alloys, oxides, phosphides, or carbides, the situation is generally even less clear-cut. Under reaction conditions, these tend to diverge from the clean, idealized surfaces often employed in CHE studies. In part, this was already recognized in the early days of photo- and electrochemical surface simulations where, as a first step, surface oxygenation and hydrogenation was determined using the CHE (Valdes et al. 2008). For example, the most stable surface termination for a rutile TiO_2 (110) facet was shown to be potential-dependent. Each of the distinct terminations in turn shows a different reactivity in terms of free energy profiles calculated with the CHE.

While the effect of the potential on the specific surface termination can thus already be included in CHE-based simulations, they still rely on idealized periodic surface models, which do not necessarily reflect reality in a catalytic device. The most common setup for TiO_2-based catalysis, for example, is in the form of nano-particles, decorated with smaller co-catalyst particles (cf. Sect. 3.2) or molecules (Graetzel 2012). The catalyst nano-particles thereby are rife with non-periodicities, such as defects, kinks, terraces, and edges, all of which could significantly influence the catalyst reactivity (Diebold 2003). The importance of these deviations from the pristine periodic structure thereby stems from the occurrence of under-coordinated metal atoms which can exhibit exceptionally strong binding to certain adsorbates (Diebold et al. 1998; Gong et al. 2006) and can also attract localized charge carriers in the form of polarons (cf. Fig. 4) (Setvin et al. 2014). These can, if localized near a suitable reactive site, drive reduction and oxidation reactions (Deskins et al. 2010; Zawadzki et al. 2012). In the following, we discuss some of the complexities of incorporating surface non-periodicities into simulations of photo-electrochemical systems.

3.1 Defects

Computer simulations of extended systems, both quantum mechanical and classical in nature, rely heavily on the method of periodic boundary conditions (PBCs). There, a given simulation supercell, or in the case of surfaces a slab geometry, is essentially repeated infinitely. Short-range interactions are treated directly, while long-range ones, such as the Coulomb interaction, are summed up in reciprocal space. This approach, together with an appropriate sampling of the Brillouin zone in reciprocal space, has proven to yield highly accurate results in the different fields of computational materials science. PBCs have been extensively applied in DFT calculations to straightforwardly simulate the electronic band structure and density of states of metals, semiconductors, and insulators.
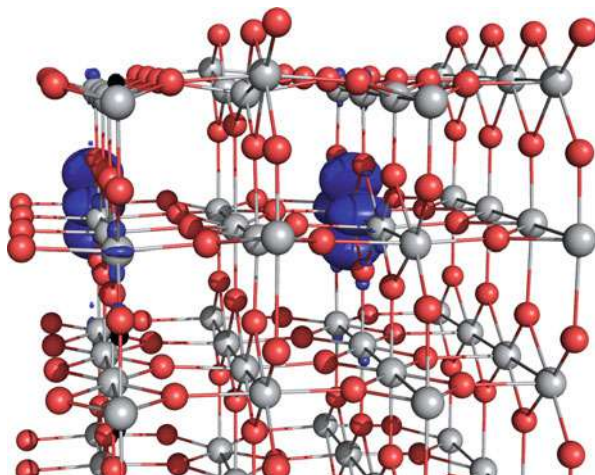
Fig. 4 Illustration of polaron localization near a neutral oxygen vacancy in TiO$_2$, with titanium atoms shown in silver and oxygen depicted in red. Removal of one oxygen from the surface leaves behind two electrons, which localize near the overall positive defect. Polarons, here represented as blue iso-surfaces (iso-value = 0.02e) of the absolute value of the spin density, are shown in their most stable configuration (Deskins et al. 2011; Shibuya et al. 2012; Kick et al. 2019)

Unfortunately, defects, terraces, kinks, and edges present a problem for the PBC approach as they are by nature non-periodic. In order to avoid unrealistic defect densities and defect interactions, one has to increase the size of the simulated supercell or slab, greatly increasing the computational cost as well. This increase in cost, as well as the problem of sampling every likely defect geometry, is at the heart of why early simulations of photo-electrochemical reactions concentrated almost exclusively on idealized model surfaces.

The problems of simulating non-periodicities are exacerbated even further in nonmetallic catalysts such as carbides (Choyke et al. 2013), phosphides (Demers and van de Walle 2012), and oxides (Janotti et al. 2010; Janotti and Van de Walle 2011; Berger et al. 2015), due to the appearance of charged defects. In order to avoid divergence of the reciprocal space sum over infinitely repeated simulation cells, only overall charge neutral systems can be considered with PBCs. As explicit counter-charges, e.g., in the form of oppositely charged defects, cannot easily be incorporated in the small systems accessible to electronic structure calculations, the standard approach is to introduce a space charge in the form of a uniform charge background. While this can yield excellent results for similar systems of equal charge, it introduces an a priori unknown offset of the potential and thus the total energy, complicating the comparison of simulation supercells.

There are essentially two distinct approaches to correcting for both defect-defect interactions and offsets of the potential. One is to carefully extrapolate to the dilute limit from supercells of increasing sizes (Lany and Zunger 2009; Komsa and

Pasquarello 2013; Freysoldt et al. 2014). In the case of neutral defects, this is a straightforward process, yet Makov and Payne (1995) showed that in the case of charged simulation supercells, such a procedure would converge infeasibly slowly – roughly with $1/\varepsilon L$, where ε is the dielectric permittivity of the medium and L the linear dimension of the supercell. In order to compensate for the long-range interaction of charges, they put forward a correction scheme based on a model point charge embedded in a dielectric continuum. This highly simplified approach, though, can lead to a significant overcompensation due to a neglect of screening through short-range reorganization (Van de Walle and Neugebauer 2004). This led to the emergence of a number of other corrections which yield significantly improved energetics for the comparison of charged supercells (Lany and Zunger 2009; Janotti et al. 2010; Komsa and Pasquarello 2013; Freysoldt et al. 2014).

The other approach to simulating charged defects in solids is to forego the use of PBCs altogether and use a non-periodic cluster model instead. In analogy to the explicit and implicit solvation approaches discussed in Sect. 2, only a small cluster is treated fully quantum mechanically, while the influence of the environment is simulated on a lower, classical, level of theory (Sherwood et al. 2003). In this quantum mechanical/molecular mechanical (QM/MM) approach, also the environment is described by a cluster, albeit vastly larger compared to the QM region. Thus, arbitrary charge states of the active QM region can be calculated without spurious offsets, yet at the cost of additional approximations regarding the interaction of QM and MM zones. In the case of ionic materials, the MM environment also needs to account for the long-range contributions to the overall potential of the simulation cell to yield the correct level alignment with respect to the fully periodic system (Berger et al. 2014), which can be achieved by fitting to periodic reference calculations (Sokol et al. 2004). On top of that, through the close contact of QM and MM atoms in a solid, spurious delocalization of charges from the QM into the MM region needs to be counteracted, e.g., by applying suitable nonlocal effective core potentials to MM atoms within range of the QM basis functions. A schematic of a solid-state QM/MM setup for TiO_2 is depicted in Fig. 5.

In addition to a more straightforward treatment of charged systems, QM/MM embedding can lead to much more favorable computational costs compared to fully periodic calculations, allowing, for example, the treatment of catalytically relevant surfaces on the level of hybrid or double hybrid functionals (Berger et al. 2014), or even with coupled cluster theory (Kubas et al. 2016), the current "gold standard" of computational chemistry. In addition to yielding accurate binding energies of various catalytically relevant adsorbates (Berger et al. 2014; Kubas et al. 2016), embedded calculations have demonstrated realistic descriptions of defect levels and formation energies in TiO_2 at different charge states (Berger et al. 2015). In the dilute limit, surface oxygen vacancies were shown to be stable across a wide range of conditions with regard to doping (i.e., Fermi level position) and oxygen chemical potential. For photo-electrochemical applications, most interesting was the finding that, analogously to the bulk (Janotti et al. 2010), positively charged defects are most stable for a diverse set of conditions.
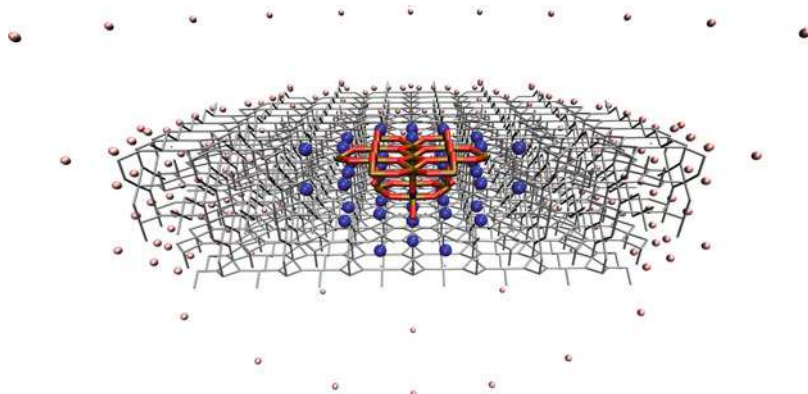
Fig. 5 Schematic of a QM/MM setup for TiO_2. The QM region contains 17 titanium (brown) and 34 oxygen atoms (red). The QM region is embedded into a roughly 3000 atom strong TiO_2 cluster (silver) represented by a polarizable force field. Seventy-seven positively charged MM Ti-centers near the QM region are replaced by effective core potentials (blue spheres). The long-range electrostatic Madelung potential of a periodic calculation is simulated with the help of fitted charges (pink spheres) far away from the QM region. For illustration purposes only the back half ($y > 0$) of the MM setup is shown here

Finally, the embedding approach is not constrained to only point-defects but can also be applied to the study of terraces, corners, and edges. By definition, these present a large number of under-coordinated sites and are thus prime candidates in the search for the reactive site. Indeed, a recent study by Bhattacharya and co-workers (Bhattacharya et al. 2017) showed that corner sites can also act as charge carrier traps. Moreover, employing ab initio thermodynamics (Reuter 2016), the authors showed that changes to the coordination of edge atoms drastically alter the relative stability of surface sites in terms of oxidation or reduction.

In summary, material non-periodicities such as point-defects, terraces, corners, and edges have been shown to exhibit properties that could significantly influence both photo- and electrocatalytic properties of materials. Through the increased availability of computational methods in recent years, defects have wandered more and more into the spotlight of theoretical research. With that they could open the door to controlled, purposeful nano-engineering of defects for more targeted and efficient catalysts.

3.2 Co-catalysts

Ever since the early days of photo-catalysis, semiconducting photocatalysts have been decorated with co-catalyst particles, which have been found to aid in the overall photoreaction (Graetzel 2012). The co-catalyst particle is thereby assumed to fulfill one of three possible functions (Yang et al. 2013; Ran et al. 2014):

- *Exciton dissociation*: Co-catalysts can aid in the separation of charge carriers, created through a photoexcitation. One of the carriers can then get trapped in the co-catalyst and partake in a reduction or oxidation reaction (cf. Sect. 3.1).
- *Lowering barriers*: Providing additional reaction sites, co-catalysts can contribute to the lowering of activation free energies or overpotentials, e.g., through different binding energies compared to the catalyst surface.
- *Catalyst stabilization*: Many semiconductors with bandgaps in the visible range –which are the best candidates for photocatalyst materials – are susceptible to oxidation through photo-generated electron holes. By trapping these holes and applying them in oxidation reactions, co-catalysts can help to protect the semiconductor surface from self-decomposing.

For all of these functions, the electronic energy levels in the co-catalyst have to be more or less closely aligned to either valence or conduction band of the semiconductor, in order to allow the exchange of charge carriers between them. A typically found sketch of the mechanisms involved is given in Fig. 6. In the last decade, a great variety of co-catalysts have been put forward (Yang et al. 2013; Ran et al. 2014), from metal clusters of varying degree of earth abundance (Ikeda et al. 2001; Tsuji et al. 2004; Wang and Hammer 2006; Zong et al. 2008; Rayalu et al. 2013; Li et al. 2014a; Ran et al. 2017), to bioinspired molecules (Hou et al. 2011; Bi et al. 2015). Thereby, co-catalysts offer a great deal of flexibility in their design, as, for example, shown by Su et al. (2014), who engineered and calculated bimetallic co-catalyst particles which facilitate the photocatalytic hydrogen evolution while suppressing the back reaction.

Computational treatment of co-catalysts on surfaces faces, to a certain degree, similar problems as other deviations from pristine surfaces. Most co-catalyst particles are sized at least a few tens of nano-meters, making them impossible to simulate in full with ab initio methods. Even small clusters in the non-scalable size regime can easily demand very large surface unit cells to avoid spurious interaction between the co-catalysts (Hou et al. 2011). Additionally, combined theoretical and experimental studies have found metal particles on oxides to aggregate near vacancies (Wahlström et al. 2003; Matthey et al. 2007), which facilitate the growth of nano-clusters. Furthermore, Gong et al. (2008) found Au clusters to predominantly aggregate near terraces, while Pt clusters form near steps of the support material. Indeed, non-periodicities in the substrate material were found to be essential for pinning clusters, thus preventing the otherwise highly mobile particles simply forming larger and possible less reactive clusters through sintering.

After determining the geometries, the next step in a computational study of the reactivity of co-catalysts is the analysis of their electronic structure, such as level alignment (Hou et al. 2011; Su et al. 2014) and local charge depletion or aggregation (Wang and Hammer 2006). This is necessary to test if charge carriers can localize in the co-catalyst without any additional barriers (cf. Fig. 6) and thus if they can contribute to the system's reactivity. Then, the effect of the co-catalyst, e.g., on the overall overpotential of a photochemical can be estimated via the CHE approach, with the same limitations discussed in Sect. 1.2. The necessary
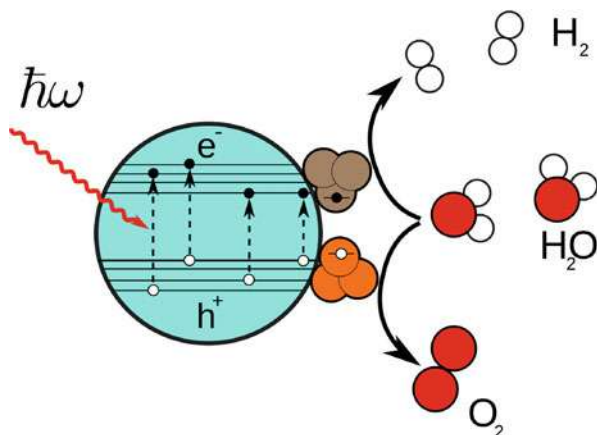
Fig. 6 Schematic of photocatalytic water splitting depicting both oxygen and hydrogen evolution co-catalysts. Holes and electrons are depicted as empty and full circles, respectively, with energy levels drawn as horizontal lines. Note the alignment of energy levels of the co-catalysts with respect to the photo-absorbing particle. Transitions from support to co-catalyst should be downhill in energy to allow charge migration, yet not too much to avoid losing energy necessary for the photocatalytic reaction

potential to drive a reaction on the co-catalyst can then be compared to the energy available in the photo-generated charge carrier localized at a respective facet of the substrate. Oberhofer and Reuter (2013) employed a modified version of the CHE which explicitly accounts for the hole localization energy, to screen gold clusters in the non-scalable size regime up to Au_{55} for their suitability as a co-catalyst with different facets. Due to the high computational costs involved in calculating large Au clusters on a hybrid level of theory, this study neglected support effects beyond that of the localized electron hole. Nevertheless, a number of promising candidates were found to yield clear improvements over pristine TiO_2 surfaces.

In summary, due to the large system sizes necessary to accurately describe supported co-catalyst particles, few studies have attempted to do so to date. Yet, with both increases in computing power and improvements to methodology, considering, e.g., solid-state embedding, theory can further contribute to the important field of co-catalyzed photocatalysis in the future.

4   Kinetic Barriers

Finally, the overall energetic cost of a catalytic reaction is not solely due to the thermodynamic cost of formation of the reaction intermediates but also may include free energy barriers between the intermediates. While these do not influence the equilibrium constant of the reaction, i.e., the ratio of reactants to products in equilibrium, they do determine the dynamical properties of the reaction such as the turnover

frequency. High turnover frequencies are of paramount importance for the viability of a catalyst in industrial applications. Therefore, for a complete assessment of the suitability of a given material as a catalyst, kinetic free energy barriers should be considered as well. Indeed, in theoretical studies of, e.g., heterogeneous catalysis, it has been a long-standing practice to calculate barriers which can then be employed in the study of even very complex reaction networks to predict turnover frequencies (Reuter 2016).

Photo-electrochemical simulations have unfortunately lagged behind here for a number of reasons, all more or less related to problems of time- and length-scales. The defining property of electrochemical reaction steps is the transfer of protons and electrons, charged species which incur significant response from the dielectric environment (cf. Sect. 2). In determining a barrier between reaction intermediates, one thus – in principle – has to sample all environment degrees of freedom for each value of the reaction coordinate. This means that, in contrast to heterogeneous catalysis, entropic contributions to the free energy barrier cannot necessarily be approximated as a simple harmonic contribution.

Nevertheless, a lot of insight can already be gained from the potential energy surfaces (PES) of elementary reaction steps. Mattioli et al. (2013), for example, employed the nudged elastic band method (NEB) to determine energy barriers between intermediates of the water oxidation reaction on small cobalt-oxide clusters to elucidate likely reaction sites and mechanisms. A similar approach was followed by Li et al. (2014b) to elucidate the reaction mechanism of $NaBH_4$ hydrolysis. Both approaches draw their intermediate configurations from equilibrium ab initio molecular dynamics (AIMD) simulations, yet the actual reactions are described on a static level with NEB to determine the minimum energy pathways, disregarding entropic contributions. They are therefore well suited to give a first mechanistic insight into the respective reactions and could serve as a starting point for more general free energy calculations, for example, using collective coordinates constructed from the minimum PES data (Li et al. 2014b).

Fully dynamical simulations of (photo-)electrochemical reactions, on the other hand, suffer from a number of problems on their own. Next, to the high computational costs associated with AIMD simulations of realistically sized systems, accounting for the electrochemical driving force on metal electrodes is far from straightforward. Real electrochemical reactions, though, occur at a constant potential with an infinite reservoir of electrons, with respect to the catalyst surface. Yet, due to the limited size of the simulation supercells, the reservoir of electrons in the simulation is limited as well. This means that all simulations are naturally performed at a constant total charge which corresponds to a change of interfacial charge density and thus potential across the reaction pathway (Skúlason et al. 2007; Wasileski and Janik 2008). Earlier studies by Rossmeisl and co-workers tried to estimate this finite size error by simply simulating the system in differently sized simulation cells and extrapolating to the infinitely extended surface (Rossmeisl et al. 2008; Skúlason et al. 2010). This approach is, however, fairly impractical, especially for more complex reactions. Recently, Chan and Nørskov (2015) therefore presented a much simpler approach, which requires only a single barrier calculation which is

later augmented by a simple correction factor due to the change in work function Φ and surface charge q between initial (I) and transition state (TS).

$$\Delta E_{corr} = \frac{1}{2}(q_{TS} - q_I)(\Phi_{TS} - \Phi_I) \qquad (11)$$

Charges $q_{I,TS}$ can thereby be determined with any sufficiently accurate estimation scheme, such as Bader or Hirshfeld charges, directly from the DFT calculation of the respective states. Note that this simple correction does not only apply to barriers but is equally valid for free energy differences between reaction intermediates. Initially demonstrated for static calculations of only energy barriers (Chan and Nørskov 2016; Shi et al. 2016), this approach can straightforwardly be applied to dynamical simulations as well, as demonstrated by Cheng et al. (2015, 2017). They employed ab initio meta-dynamics to determine reaction barriers for the CO reduction on a Cu(100) surface to shed light on possible reaction pathways at different applied potentials. Only by including the calculation of free energy barriers into the picture, they were able to fully explain experimental observations. They accurately predicted the competition between electrochemical reactions and surface sites and the selectivity toward different products at different potentials (Cheng et al. 2017).

The situation is somewhat different in photo-electrocatalysis on semiconductors, where localized polaronic charge carriers (electrons or holes) are known to drive the reaction (Cowan and Durrant 2013). There, rather than a good description of constant potentials, it is important to accurately model polarons and their interaction with the reactants. In again a mixed dynamical/statistical approach, Selloni and co-workers (Chen et al. 2013) determined the PES of the first – and most likely rate determining – step of water oxidation on TiO_2 anatase (101) driven by an explicit electron hole. Specifically, they determined the energy profiles of the proton transfer step in two different electronic configurations with the hole on the surface (initial state) and in the water (final state), respectively. As a reaction coordinate, they thereby employed a geometric criterion of the proton being attached to the surface adsorbed water or the water molecule in the first solvation shell. Their results highlight a problem with this kind of approach, as a neglect of the solvent reorganization leads to the electron hole staying at the surface adsorbed water even if the proton is moved all the way toward the designated final step.

Finally, a fully dynamic scheme was presented by Stecher et al. (2016) who utilized a solid-state embedding scheme (cf. Sect. 3.1) to allow for an explicit and finite size effect free treatment of the electron hole. They also employed a Marcus theory-inspired reaction coordinate based on the energy gap between initial and final state (Mones et al. 2009; Mones and Csányi 2012), which explicitly includes solvent reorganization effects and thus minimizes hysteresis. Using umbrella sampling, they determined the free energy barrier for the first reaction step of water oxidation TiO_2 rutile (110) to be roughly 0.2eV, comparable to estimates by Selloni and co-workers.

All of the above studies share a common theme in that they show non-negligible barriers much larger than the thermal energy k_BT at room temperature. Thus, without additional driving forces, the respective (photo-)electrochemical

reactions are predicted to show very small turnover frequencies. For the modeling scientist, this means that although first estimates based solely on thermodynamic considerations and the CHE can be very useful in identifying rate limiting steps and reaction intermediates, only a full picture including free energy barriers can accurately determine the suitability of a given catalyst.

5   Summary

Since its inception, the computational hydrogen electrode method has undeniably led to great insights in electrochemistry and photo-electrochemistry. Together with the scaling relations, which at least for metals allow extrapolation of a material's suitability as a catalyst from just a few binding energies, the CHE enabled the large-scale screening of likely catalyst materials. On the other hand, in terms of practical implementation, most studies utilizing the CHE employed a number of simplifications for reasons of computational efficiency. Yet, with the development of more complex catalysts, especially those with the ability to go beyond the performance limits dictated by the scaling relations, the need for augmentations to the classical CHE approach became more apparent. In this chapter, we discussed three possible improvements over standard CHE calculations.

In the original applications of the CHE, solvation effects were assumed to play a minor role for the energetics of neutral reaction intermediates and thus omitted. Additionally, leaving out sampling of all solvent degrees of freedom led to huge boosts in efficiency. Yet, considering, e.g., photo-electrochemical reactions on semiconductors going through charged or even just differently polar intermediates, omission of solvation effects can lead to erroneous results. In Sect. 2, we discuss methodological approaches to including solvation effects on the level of the electronic structure calculations, both with their own sets of advantages and disadvantages. Explicit solvation in the form of ab initio molecular dynamics is generally very expensive and, regarding, e.g., electrolyte concentrations, limited with respect to system size. Matters can be improved by employing the approximation of immobile, icelike water on the catalyst surface, which, though, is not applicable to all systems. The upshot here is the correct description of non-mean-field effects, such as potential-dependent solvent reorientations and hydrogen bonding. Implicit solvation models, on the other hand, offer much improved computational efficiency, with, using appropriate models, less constraints on system sizes and ion concentrations. This comes at the cost of non-mean-field interactions being described only as effective corrections. On top of that, most implicit solvent models to date are parametrized to fit test sets of small- to medium-sized molecules and therefore do not necessarily reproduce the solid-liquid interface. Finally, depending on the specific reaction of interest, the solvation environment is not always in equilibrium with the reaction center due to differing timescales of reaction and solvent reorganization. In this case, effective solvent models, which essentially assume an equilibrium response of the solvent to the solute, cannot be straightforwardly applied. Given the wealth of available methods, there is no single

optimal choice as each has its own set of drawbacks and advantages, and each specific approach needs to be tailored to the specific problem at hand.

Another fundamental question, treated in Sect. 3, is that of the exact nature of the reactive site. On simple monometallic catalysts, one can often expect reasonably pristine surfaces, where the exact choice of reactive site is less influential on, e.g., the adsorption energetics. This picture is radically altered on more complex materials such as oxides or nitrides, where terraces, steps, and – under reaction conditions – defects are abound. These can have tremendous influences on (photo-)electrochemical reactions by trapping charge carriers or pinning co-catalyst particles. Being deviations from perfect pristine periodic structures, they are hard to tackle in computer simulations, as they necessitate very large simulation cells to avoid spurious interactions between periodic images. In the case of charged defects, this situation is exacerbated through the interaction between the charge carrier and the neutralizing background charge, which has to be accounted for through extrapolation from increasing system sizes. Another approach shown to effectively tackle the problem of charged defects is the solid-state embedding or QM/MM technique. There, only a finite region of interest is treated quantum mechanically, while the environment is represented as an equally finite cluster of polarizable MM atoms. Adding a correction for the long-range Madelung potential, this approach effectively yields a description of non-periodicities in the QM region in the infinitely dilute limit. Thereby, it can also be used with higher-level methods such as (double-)hybrid DFT or even highly accurate coupled cluster methods. These advantages come at the cost of possibly increased inaccuracies due to an imperfect description of the electronic structure at the QM/MM interface, which means that convergence with cluster size and shape needs to be carefully tested.

A further fundamental aspect in the search for the reactive site, generally overlooked in most applications of the CHE method so far, is the role of co-catalysts. These can range from bioinspired molecules to metal or metal-oxide clusters and, as laid out in Sect. 3.2, can have a number of effects in a photocatalyst. Depending on the exact role, they need to be energetically aligned with conduction or valence band edges of the photo-absorbing support material. Most computational studies so far therefore concentrated on structures and electronic alignment of co-catalyst and support. Using a reformulation of the CHE methodology, a first study of Au clusters in the non-scalable size regime showed several promising co-catalysts for water oxidation on varying TiO_2 facets. Nevertheless, theoretical research into the nature and effects of co-catalysts in photocatalysis is still in its early stages and much needs to be done in the future.

The third and final kind of improvement over standard CHE-based (photo-) electrochemical studies discussed in Sect. 4 is the study of kinetic barriers. CHE calculations in the past focused on the optimization of catalysts with respect to the thermodynamic overpotential and free energy barriers along the reaction pathway were generally neglected. On metal catalysts this was mainly due to the problems associated with simulating reactions at constant potential – instead of constant charge – during proton transfer steps. Several correction schemes have been proposed, and in the last few years a number of studies estimating potential energy

barriers have been published. On semiconductors, on the other hand, photocatalytic reactions are known to be driven by surface localized charge carriers, which again leads to the aforementioned problems due to the neutralizing charge background. As in the case of surface defects, these problems can be tackled by charge correction or QM/MM embedding schemes.

The other drawback in simulating barriers at the solid-liquid interface is again the question of sampling the solvent response. Recent simulations on both metals and semiconductor catalysts showed that the reorganization of the solvent exerts a great influence on the reaction barriers. Given this fact, even the choice of reaction coordinate is not entirely straightforward, as simple geometric coordinates fail to account for solvent response and thus can lead to very large hysteresis effects. More complex reaction coordinates, based on a Marcus theory model of charge transport, demonstrably yielded a much improved description of the reaction free energy profile.

In summary, the last few years have seen tremendous methodological advances for the simulation of electro- or photo-electrochemical surface reactions. First insights into their microscopic details were made possible by the theoretical framework of the CHE, together with a number of additional approximations used in its practical implementation. Yet, like pieces of a jigsaw puzzle, more and more details have been included into the picture, such that a fully quantitative simulation of (photo-)electrochemistry from first principles is now within reach. Depending on the system and the questions to be answered, some of these simulations will be based on the framework of the CHE, one of its generalizations, or completely different, fully dynamical ways to align computations to experiments. This chapter focused on a number of methodological improvements to the way the CHE has been applied in the past. These allowed us, as a community, to graduate from model simulations on idealized surfaces in vacuum to defected catalysts in realistic environments, bringing us closer and closer to experimental realities.

References

Abild-Pedersen F, Greeley J, Studt F, Rossmeisl J, Munter T, Moses PG, Skulason E, Bligaard T, Nørskov JK (2007) Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. Phys Rev Lett 99(1):016105

Aktins PW, de Paula J (2014) Atkins' physical chemistry, 10th edn. Oxford University Press, Oxford

Andreussi O, Dabo I, Marzari N (2012) Revised self-consistent continuum solvation in electronic-structure calculations. J Chem Phys 136(6):064102

Bardhan JP, Knepley MG (2014) Communication: modeling charge-sign asymmetric solvation free energies with nonlinear boundary conditions. J Chem Phys 141(13):131103. https://doi.org/10.1063/1.4897324

Berger D, Logsdail AJ, Oberhofer H, Farrow MR, Catlow CRA, Sherwood P, Sokol AA, Blum V, Reuter K (2014) Embedded-cluster calculations in a numeric atomic orbital density-functional theory framework. J Chem Phys 141(2):024105

Berger D, Oberhofer H, Reuter K (2015) First-principles embedded-cluster calculations of the neutral and charged oxygen vacancy at the rutile TiO_2 (110) surface. Phys Rev B 92(7):075308

Bhattacharya S, Berger D, Reuter K, Ghiringhelli LM, Levchenko SV (2017) Theoretical evidence for unexpected O-rich phases at corners of MgO surfaces. Phys Rev Mat 1(7):071601

Bi W, Li X, Zhang L, Jin T, Zhang L, Zhang Q, Luo Y, Wu C, Xie Y (2015) Molecular co-catalyst accelerating hole transfer for enhanced photocatalytic H_2 evolution. Nat Commun 6:8647

Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K, Scheffler M (2009) Ab initio molecular simulations with numeric atom-centered orbitals. Comput Phys Commun 180(11):2175–2196

Boda D, Fawcett WR, Henderson D, Sokołowski S (2002) Monte Carlo, density functional theory, and Poisson–Boltzmann theory study of the structure of an electrolyte near an electrode. J Chem Phys 116(16):7170–7176

Bohinc K, Shrestha A, Brumen M, May S (2012) Poisson-Helmholtz-Boltzmann model of the electric double layer: analysis of monovalent ionic mixtures. Phys Rev E 85(3):031130

Borukhov I, Andelman D, Orland H (1997) Steric effects in electrolytes: a modified Poisson-Boltzmann equation. Phys Rev Lett 79(3):435

Chan K, Nørskov JK (2015) Electrochemical barriers made simple. J Phys Chem Lett 6(14):2663–2668

Chan K, Nørskov JK (2016) Potential dependence of electrochemical barriers from ab initio calculations. J Phys Chem Lett 7(9):1686–1690

Chapman DL (1913) LI. A contribution to the theory of electrocapillarity. Philos Mag 25(148):475–481

Chen J, Li YF, Sit P, Selloni A (2013) Chemical dynamics of the first proton-coupled electron transfer of water oxidation on TiO_2 anatase. J Am Chem Soc 135(50):18774–18777

Cheng J, Sprik M (2010a) Acidity of the aqueous rutile TiO_2 (110) surface from density functional theory based molecular dynamics. J Chem Theor Comput 6(3):880–889

Cheng J, Sprik M (2010b) Aligning electronic energy levels at the TiO_2/H_2O interface. Phys Rev B 82(8):081406

Cheng J, Sprik M (2014) The electric double layer at a rutile TiO_2 water interface modelled using density functional theory based molecular dynamics simulation. J Phys Condens Matter 26(24):244108

Cheng J, Liu X, Kattirtzi JA, VandeVondele J, Sprik M (2014a) Aligning electronic and protonic energy levels of proton-coupled electron transfer in water oxidation on aqueous TiO_2. Angew Chem Int Ed 53(45):12046–12050

Cheng J, Liu X, VandeVondele J, Sulpizi M, Sprik M (2014b) Redox potentials and acidity constants from density functional theory based molecular dynamics. Acc Chem Res 47(12):3522–3529

Cheng T, Xiao H, Goddard WA III (2015) Free-energy barriers and reaction mechanisms for the electrochemical reduction of CO on the Cu (100) surface, including multiple layers of explicit solvent at pH 0. J Phys Chem Lett 6(23):4767–4773

Cheng T, Xiao H, Goddard WA (2017) Full atomistic reaction mechanism with kinetics for CO reduction on Cu (100) from ab initio molecular dynamics free-energy calculations at 298 K. Proc Natl Acad Sci USA 114:201612106

Choyke WJ, Matsunami H, Pensl G (2013) Silicon carbide: recent major advances. Springer Science & Business Media, Berlin

Cowan AJ, Durrant JR (2013) Long-lived charge separated states in nanostructured semiconductor photoelectrodes for the production of solar fuels. Chem Soc Rev 42(6):2281–2293

Demers S, van de Walle A (2012) Intrinsic defects and dopability of zinc phosphide. Phys Rev B 85(19):195208

Deskins NA, Rousseau R, Dupuis M (2010) Defining the role of excess electrons in the surface chemistry of TiO_2. J Phys Chem C 114(13):5891–5897

Deskins NA, Rousseau R, Dupuis M (2011) Distribution of Ti^{3+} surface sites in reduced TiO_2. J Phys Chem C 115(15):7562–7572

Diebold U (2003) Structure and properties of TiO_2 surfaces: a brief review. Appl Phys A 76(5):681–687

Diebold U, Lehman J, Mahmoud T, Kuhn M, Leonardelli G, Hebenstreit W, Schmid M, Varga P (1998) Intrinsic defects on a TiO_2 (110)(1× 1) surface and their reaction with oxygen: a scanning tunneling microscopy study. Surf Sci 411(1–2):137–153

Dupont C, Andreussi O, Marzari N (2013) Self-consistent continuum solvation (SCCS): the case of charged systems. J Chem Phys 139(21):214110

Fang YH, Liu ZP (2010) Mechanism and tafel lines of electro-oxidation of water to oxygen on RuO_2 (110). J Am Chem Soc 132(51):18214–18222

Fang YH, Wei GF, Liu ZP (2013) Theoretical modeling of electrode/electrolyte interface from first-principles periodic continuum solvation method. Catal Today 202:98–104

Feibelman PJ (2002) Partial dissociation of water on Ru (0001). Science 295(5552):99–102

Filhol JS, Doublet ML (2013) An ab initio study of surface electrochemical disproportionation: the case of a water monolayer adsorbed on a Pd (111) surface. Catal Today 202:87–97

Filhol JS, Neurock M (2006) Elucidation of the electrochemical activation of water over Pd by first principles. Angew Chem Int Ed 45(3):402–406

Fisicaro G, Genovese L, Andreussi O, Mandal S, Nair NN, Marzari N, Goedecker S (2017) Soft-sphere continuum solvation in electronic-structure calculations. J Chem Theor Comput 13(8):3829–3845

Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86(1):253

Gambu TG, Petersen MA, van Steen E (2017) Probing the edge effect on the ORR activity using platinum nanorods: a DFT study. Catal Today. https://doi.org/10.1016/j.cattod.2017.12.026

Gauthier JA, Dickens CF, Chen LD, Doyle AD, Nørskov JK (2017) Solvation effects for oxygen evolution reaction catalysis on IrO_2 (110). J Phys Chem C 121(21):11455–11463. https://doi.org/10.1021/acs.jpcc.7b02383

Gong XQ, Selloni A, Batzill M, Diebold U (2006) Steps on anatase TiO_2 (101). Nat Mater 5(8):665

Gong XQ, Selloni A, Dulub O, Jacobson P, Diebold U (2008) Small Au and Pt clusters at the anatase TiO_2 (101) surface: behavior at terraces, steps, and surface oxygen vacancies. J Am Chem Soc 130(1):370–381

Göttle AJ, Koper MT (2017) Proton-coupled electron transfer in the electrocatalysis of CO_2 reduction: prediction of sequential vs. concerted pathways using DFT. Chem Sci 8(1):458–465

Gouy G (1917) Sur la fonction électrocapillaire. Ann Phys (Paris) 9:129–184

Graetzel M (2012) Energy resources through photochemistry and catalysis. Elsevier, Amsterdam

Groß A, Gossenberger F, Lin X, Naderian M, Sakong S, Roman T (2014) Water structures at metal electrodes studied by ab initio molecular dynamics simulations. J Electrochem Soc 161(8):E3015–E3020

Halter DP, Palumbo CT, Ziller JW, Gembicky M, Rheingold AL, Evans WJ, Meyer K (2018) Electrocatalytic H_2O reduction with f-elements: mechanistic insight and overpotential tuning in a series of lanthanide complexes. J Am Chem Soc 140(7):2587–2594. https://doi.org/10.1021/jacs.7b11532

Hansen MH, Rossmeisl J (2016) pH in grand canonical statistics of an electrochemical interface. J Phys Chem C 120(51):29135–29143

Hansen MH, Nilsson A, Rossmeisl J (2017) Modelling pH and potential in dynamic structures of the water/Pt (111) interface on the atomic scale. Phys Chem Chem Phys 19(34):23505–23514

Haynes WM (2014) CRC handbook of chemistry and physics. CRC press, Boca Raton

Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268(5214):1144–1149

Hou Y, Abrams BL, Vesborg PC, Björketun ME, Herbst K, Bech L, Setti AM, Damsgaard CD, Pedersen T, Hansen O et al (2011) Bioinspired molecular co-catalysts bonded to a silicon photocathode for solar hydrogen evolution. Nat Mater 10(6):434

Hu QM, Reuter K, Scheffler M (2007) Towards an exact treatment of exchange and correlation in materials: application to the "CO adsorption puzzle" and other systems. Phys Rev Lett 98(17):176103

Ikeda S, Sugiyama N, Pal B, Marcí G, Palmisano L, Noguchi H, Uosaki K, Ohtani B (2001) Photocatalytic activity of transition-metal-loaded titanium (IV) oxide powders suspended in aqueous solutions: correlation with electron–hole recombination kinetics. Phys Chem Chem Phys 3(2):267–273

Janik MJ, Taylor CD, Neurock M (2007) First principles analysis of the electrocatalytic oxidation of methanol and carbon monoxide. Top Catal 46(3–4):306–319

Janotti A, Van de Walle CG (2011) LDA+U and hybrid functional calculations for defects in ZnO, SnO$_2$, and TiO$_2$. Phys Status Solidi B 248(4):799–804

Janotti A, Varley J, Rinke P, Umezawa N, Kresse G, Van de Walle C (2010) Hybrid functional studies of the oxygen vacancy in TiO$_2$. Phys Rev B 81(8):085212

Kandoi S, Gokhale A, Grabow L, Dumesic J, Mavrikakis M (2004) Why Au and Cu are more selective than Pt for preferential oxidation of CO at low temperature. Catal Lett 93(1–2): 93–100

Keeley DF, Hoffpauir MA, Meriwether JR (1988) Solubility of aromatic hydrocarbons in water and sodium chloride solutions of different ionic strengths: benzene and toluene. J Chem Eng Data 33(2):87–89

Kick M, Reuter K, Oberhofer H (2019) Intricacies of DFT+U, Not Only in a Numeric Atom Centered Orbital Framework. J Chem Theory Comput 15(3):1705–1718

Kiriukhin MY, Collins KD (2002) Dynamic hydration numbers for biologically important ions. Biophys Chem 99(2):155–168

Kirkwood JG (1934) Theory of solutions of molecules containing widely separated charges with special application to zwitterions. J Chem Phys 2(7):351–361

Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2(5):799–805

Komsa HP, Pasquarello A (2013) Finite-size supercell correction for charged defects at surfaces and interfaces. Phys Rev Lett 110(9):095505

Koper MT (2013a) Theory of multiple proton–electron transfer reactions and its implications for electrocatalysis. Chem Sci 4(7):2710–2723

Koper MT (2013b) Theory of the transition from sequential to concerted electrochemical proton–electron transfer. Phys Chem Chem Phys 15(5):1399–1407

Kubas A, Berger D, Oberhofer H, Maganas D, Reuter K, Neese F (2016) Surface adsorption energetics studied with "gold standard" wave-function-based ab initio methods: small-molecule binding to TiO$_2$ (110). J Phys Chem Lett 7(20):4207–4212

Lany S, Zunger A (2009) Accurate prediction of defect properties in density functional supercell calculations. Model Simul Mater Sci Eng 17(8):084002

Latimer WM, Pitzer KS, Slansky CM (1993) The free energy of hydration of gaseous ions, and the absolute potential of the normal calomel electrode. In: Pitzer KS (ed) Molecular structure and statistical thermodynamics: selected papers of Kenneth S Pitzer. World Scientific, Singapore, pp 485–489

Li J, Cushing SK, Zheng P, Senty T, Meng F, Bristow AD, Manivannan A, Wu N (2014a) Solar hydrogen generation by a CdS-Au-TiO$_2$ sandwich nanorod array enhanced with Au nanoparticle as electron relay and plasmonic photosensitizer. J Am Chem Soc 136(23):8438–8449

Li P, Henkelman G, Keith JA, Johnson JK (2014b) Elucidation of aqueous solvent-mediated hydrogen-transfer reactions by ab initio molecular dynamics and nudged elastic-band studies of NaBH$_4$ hydrolysis. J Phys Chem C 118(37):21385–21399

Long F, McDevit W (1952) Activity coefficients of nonelectrolyte solutes in aqueous salt solutions. Chem Rev 51(1):119–169

Maayan G, Gluz N, Christou G (2018) A bioinspired soluble manganese cluster as a water oxidation electrocatalyst with low overpotential. Nat Catal 1(1):48

Makov G, Payne M (1995) Periodic boundary conditions in ab initio calculations. Phys Rev B 51(7):4014

Marcus Y (1985) Ions in solution and their solvation. Wiley, New Jersey

Marenich A, Kelly C, Thompson J, Hawkins G, Chambers C, Giesen D, Winget P, Cramer C, Truhlar D (2012) Minnesota solvation database–version 2012. University of Minnesota, Minneapolis

Mathew K, Sundararaman R, Letchworth-Weaver K, Arias T, Hennig RG (2014) Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways. J Chem Phys 140(8):084106

Matthey D, Wang J, Wendt S, Matthiesen J, Schaub R, Lægsgaard E, Hammer B, Besenbacher F (2007) Enhanced bonding of gold nanoparticles on oxidized TiO_2 (110). Science 315(5819):1692–1696

Mattioli G, Giannozzi P, Amore Bonapasta A, Guidoni L (2013) Reaction pathways for oxygen evolution promoted by cobalt catalyst. J Am Chem Soc 135(41):15353–15363

Meng S, Wang E, Gao S (2004) Water adsorption on metal surfaces: a general picture from density functional theory studies. Phys Rev B 69(19):195404

Michaelides A, Alavi A, King DA (2004) Insight into H_2O-ice adsorption and dissociation on metal surfaces from first-principles simulations. Phys Rev B 69(11):113404

Mobley DL, Barber AE, Fennell CJ, Dill KA (2008) Charge asymmetries in hydration of polar solutes. J Phys Chem B 112(8):2405–2414

Mones L, Csányi G (2012) Topologically invariant reaction coordinates for simulating multistate chemical reactions. J Phys Chem B 116(51):14876–14885

Mones L, Kulhánek P, Simon I, Laio A, Fuxreiter M (2009) The energy gap as a universal reaction coordinate for the simulation of chemical reactions. J Phys Chem B 113(22):7867–7873

Neurock M (2003) Perspectives on the first principles elucidation and the design of active sites. J Catal 216(1–2):73–88

Nielsen M, Björketun ME, Hansen MH, Rossmeisl J (2015) Towards first principles modeling of electrochemical electrode–electrolyte interfaces. Surf Sci 631:2–7

Nørskov JK, Rossmeisl J, Logadottir A, Lindqvist L, Kitchin JR, Bligaard T, Jonsson H (2004) Origin of the overpotential for oxygen reduction at a fuel-cell cathode. J Phys Chem B 108(46):17886–17892

Oberhofer H, Reuter K (2013) First-principles thermodynamic screening approach to photocatalytic water splitting with co-catalysts. J Chem Phys 139(4):044710

Ogasawara H, Brena B, Nordlund D, Nyberg M, Pelmenschikov A, Pettersson L, Nilsson A (2002) Structure and bonding of water on Pt (111). Phys Rev Lett 89(27):276102

Onsager L (1936) Electric moments of molecules in liquids. J Am Chem Soc 58(8):1486–1493

Otani M, Sugino O (2006) First-principles calculations of charged surfaces and interfaces: A plane-wave nonrepeated slab approach. Phys Rev B 73(11):115407

Otani M, Hamada I, Sugino O, Morikawa Y, Okamoto Y, Ikeshoji T (2008) Electrode dynamics from first principles. J Phys Soc Jpn 77(2):024802–024802

Pérez-Tejeda P, Maestre A, Balón M, Hidalgo J, Muñoz MA, Sánchez M (1987) Setschenow coefficients for caffeine, theophylline and theobromine in aqueous electrolyte solutions. J Chem Soc Farad Trans 1 83(4):1029–1039

Peterson AA, Abild-Pedersen F, Studt F, Rossmeisl J, Nørskov JK (2010) How copper catalyzes the electroreduction of carbon dioxide into hydrocarbon fuels. Energy Environ Sci 3(9):1311–1315

Qian M, Cui S, Jiang D, Zhang L, Du P (2017) Highly efficient and stable water-oxidation electrocatalysis with a very low overpotential using FeNiP substitutional-solid-solution nanoplate arrays. Adv Mater 29(46):1704075

Ran J, Zhang J, Yu J, Jaroniec M, Qiao SZ (2014) Earth-abundant cocatalysts for semiconductor-based photocatalytic water splitting. Chem Soc Rev 43(22):7787–7812

Ran J, Gao G, Li FT, Ma TY, Du A, Qiao SZ (2017) Ti_3C_2 MXene co-catalyst on metal sulfide photo-absorbers for enhanced visible-light photocatalytic hydrogen production. Nat Commun 8:13907

Rao RR, Kolb MJ, Halck NB, Pedersen AF, Mehta A, You H, Stoerzinger KA, Feng Z, Hansen HA, Zhou H et al (2017) Towards identifying the active sites on RuO_2 (110) in catalyzing oxygen evolution. Energy Environ Sci 10(12):2626–2637

Rayalu SS, Jose D, Joshi MV, Mangrulkar PA, Shrestha K, Klabunde K (2013) Photocatalytic water splitting on Au/TiO_2 nanocomposites synthesized through various routes: enhancement in photocatalytic activity due to SPR effect. Appl Catal B 142:684–693

Reda M, Hansen HA, Vegge T (2018) DFT study of stabilization effects on N-doped graphene for ORR catalysis. Catal Today 312:118–125

Reuter K (2016) Ab initio thermodynamics and first-principles microkinetics for surface catalysis. Catal Lett 146(3):541–563

Reuter K, Scheffler M (2001) Composition, structure, and stability of RuO_2 (110) as a function of oxygen pressure. Phys Rev B 65(3):035406

Reuter K, Plaisance CP, Oberhofer H, Andersen M (2017) Perspective: on the active site model in computational catalyst screening. J Chem Phys 146(4):040901

Ringe S, Oberhofer H, Hille C, Matera S, Reuter K (2016) Function-space-based solution scheme for the size-modified Poisson–Boltzmann equation in full-potential DFT. J Chem Theor Comput 12(8):4052–4066

Ringe S, Oberhofer H, Reuter K (2017) Transferable ionic parameters for first-principles Poisson-Boltzmann solvation calculations: neutral solutes in aqueous monovalent salt solutions. J Chem Phys 146(13):134103

Rossmeisl J, Skúlason E, Björketun ME, Tripkovic V, Nørskov JK (2008) Modeling the electrified solid–liquid interface. Chem Phys Lett 466(1):68–71

Rossmeisl J, Chan K, Ahmed R, Tripković V, Björketun ME (2013) pH in atomic scale simulations of electrochemical interfaces. Phys Chem Chem Phys 15(25):10321–10325

Scherlis DA, Fattebert JL, Gygi F, Cococcioni M, Marzari N (2006) A unified electrostatic and cavitation model for first-principles molecular dynamics in solution. J Chem Phys 124(7):074103

Schnur S, Groß A (2009) Properties of metal–water interfaces studied from first principles. New J Phys 11(12):125003

Setchenow M (1892) Action de l'acide carbonique sur les solutions des sels á acides forts. Ann Chim Phys 25:226

Seh ZW, Kibsgaard J, Dickens CF, Chorkendorff I, Nørskov JK, Jaramillo TF (2017) Combining theory and experiment in electrocatalysis: insights into materials design. Science 355(6321):eaad4998

Seitz LC, Dickens CF, Nishio K, Hikita Y, Montoya J, Doyle A, Kirk C, Vojvodic A, Hwang HY, Norskov JK et al (2016) A highly active and stable IrO_x/$SrIrO_3$ catalyst for the oxygen evolution reaction. Science 353(6303):1011–1014

Setvin M, Hao X, Daniel B, Pavelec J, Novotny Z, Parkinson GS, Schmid M, Kresse G, Franchini C, Diebold U (2014) Charge trapping at the step edges of TiO_2 anatase (101). Angew Chem Int Ed 53(18):4714–4716

Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ et al (2003) QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. J Mol Struct 632(1–3):1–28

Shi C, Chan K, Yoo JS, Nørskov JK (2016) Barriers of electrochemical CO_2 reduction on transition metals. Org Process Res Dev 20(8):1424–1430

Shibuya T, Yasuoka K, Mirbt S, Sanyal B (2012) A systematic study of polarons due to oxygen vacancy formation at the rutile TiO_2 (110) surface by GGA+U and HSE06 methods. J Phys Condens Matter 24(43):435504

Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theor Comput 6(5):1509–1519

Siahrostami S, Vojvodic A (2015) Influence of adsorbed water on the oxygen evolution reaction on oxides. J Phys Chem C 119(2):1032–1037. https://doi.org/10.1021/jp508932x

Sinstein M, Scheurer C, Matera S, Blum V, Reuter K, Oberhofer H (2017) Efficient implicit solvation method for full potential DFT. J Chem Theor Comput 13(11):5582–5603

Skúlason E, Karlberg GS, Rossmeisl J, Bligaard T, Greeley J, Jónsson H, Nørskov JK (2007) Density functional theory calculations for the hydrogen evolution reaction in an electrochemical double layer on the Pt (111) electrode. Phys Chem Chem Phys 9(25):3241–3250

Skúlason E, Tripkovic V, Björketun ME, Gudmundsdottir S, Karlberg G, Rossmeisl J, Bligaard T, Jónsson H, Nørskov JK (2010) Modeling the electrochemical hydrogen oxidation and evolution reactions on the basis of density functional theory calculations. J Phys Chem C 114(42):18182–18197

Sokol AA, Bromley ST, French SA, Catlow CRA, Sherwood P (2004) Hybrid QM/MM embedding approach for the treatment of localized surface states in ionic materials. Int J Quantum Chem 99(5):695–712

Stampfl C, Ganduglia-Pirovano MV, Reuter K, Scheffler M (2002) Catalysis and corrosion: the theoretical surface-science context. Surf Sci 500(1–3):368–394

Stecher T, Reuter K, Oberhofer H (2016) First-principles free-energy barriers for photoelectrochemical surface reactions: proton abstraction at TiO_2 (110). Phys Rev Lett 117(27):276001

Stern O (1924) The theory of the electrolytic double-layer. Z Elektrochem 30(508):1014–1020

Su R, Tiruvalam R, Logsdail AJ, He Q, Downing CA, Jensen MT, Dimitratos N, Kesavan L, Wells PP, Bechstein R et al (2014) Designer titania-supported Au–Pd nanoparticles for efficient photocatalytic hydrogen production. ACS Nano 8(4):3490–3497

Sumita M, Hu C, Tateyama Y (2010) Interface water on TiO_2 anatase (101) and (001) surfaces: first-principles study with TiO_2 slabs dipped in bulk water. J Phys Chem C 114(43):18529–18537

Thiyagarajan N, Janmanchi D, Tsai YF, Wanna WH, Ramu R, Chan SI, Zen JM, Yu SSF (2018) A carbon electrode functionalized by a tricopper cluster complex: overcoming overpotential and production of hydrogen peroxide in the oxygen reduction reaction. Angew Chem Int Ed 57:1

Tilocca A, Selloni A (2004) Structure and reactivity of water layers on defect-free and defective anatase TiO_2 (101) surfaces. J Phys Chem B 108(15):4743–4751

Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105(8):2999–3094

Tsuji I, Kato H, Kobayashi H, Kudo A (2004) Photocatalytic H_2 evolution reaction from aqueous solutions over band structure-controlled $(AgIn)_x(Zn_2)_{(1-x)}S_2$ solid solution photocatalysts with visible-light response and their surface nanostructures. J Am Chem Soc 126(41):13406–13413

Valdes A, Qu ZW, Kroes GJ, Rossmeisl J, Nørskov JK (2008) Oxidation and photo-oxidation of water on TiO_2 surface. J Phys Chem C 112:9872

Van de Walle CG, Neugebauer J (2004) First-principles calculations for defects and impurities: applications to III-nitrides. J Appl Phys 95(8):3851–3879

Wahlström E, Lopez N, Schaub R, Thostrup P, Rønnau A, Africh C, Lægsgaard E, Nørskov J, Besenbacher F (2003) Bonding of gold nanoclusters to oxygen vacancies on rutile TiO_2 (110). Phys Rev Lett 90(2):026101

Wang J, Hammer B (2006) Role of Au^+ in supporting and activating Au_7 on TiO_2 (110). Phys Rev Lett 97(13):136107

Wasileski SA, Janik MJ (2008) A first-principles study of molecular oxygen dissociation at an electrode surface: a comparison of potential variation and coadsorption effects. Phys Chem Chem Phys 10(25):3613–3627

Waxman EM, Elm J, Kurtén T, Mikkelsen KV, Ziemann PJ, Volkamer R (2015) Glyoxal and methylglyoxal setschenow salting constants in sulfate, nitrate, and chloride solutions: measurements and Gibbs energies. Environ Sci Technol 49(19):11500–11508

Wood BC, Schwegler E, Choi WI, Ogitsu T (2013) Hydrogen-bond dynamics of water at the interface with InP/GaP (001) and the implications for photoelectrochemistry. J Am Chem Soc 135(42):15774–15783

Yang J, Wang D, Han H, Li C (2013) Roles of cocatalysts in photocatalysis and photoelectrocatalysis. Acc Chem Res 46(8):1900–1909

Zawadzki P, Laursen AB, Jacobsen KW, Dahl S, Rossmeisl J (2012) Oxidative trends of TiO_2—hole trapping at anatase and rutile surfaces. Energy Environ Sci 5(12):9866–9869

Zhang B, Zheng X, Voznyy O, Comin R, Bajdich M, García-Melchor M, Han L, Xu J, Liu M, Zheng L, García de Arquer FP, Dinh CT, Fan F, Yuan M, Yassitepe E, Chen N, Regier T, Liu P, Li Y, De Luna P, Janmohamed A, Xin HL, Yang H, Vojvodic A, Sargent EH (2016) Homogeneously dispersed multimetal oxygen-evolving catalysts. Science 352(6283):333–337

Zhao Z, Li Z, Zou Z (2012) Structure and properties of water on the anatase TiO_2 (101) surface: from single-molecule adsorption to interface formation. J Phys Chem C 116(20):11054–11061

Zong X, Yan H, Wu G, Ma G, Wen F, Wang L, Li C (2008) Enhancement of photocatalytic H_2 evolution on CdS by loading MoS_2 as cocatalyst under visible light irradiation. J Am Chem Soc 130(23):7176–7177

Part IX
Hierarchical Materials Modeling: Mechanical Performance

Multiscale Modeling of Structural Materials: Chemistry and Mechanical Performance

62

Francisco J. Martin-Martinez and Markus J. Buehler

Contents

Abstract

Through an overview of selected chapters in this section we navigate the state of the art of different materials with crucial structural features at different length scales. It includes a discussion on hierarchical biomaterials like silk, new aerogels with enhanced thermal and mechanical properties, and infrastructure materials like wood, concrete, and asphalt, whose complex composition leads to specific modeling challenges. The wide range of materials discussed cover different length scales from nano- to macro-, and different macroscopic properties, e.g., from soft to stiff. All of them together represent a selection of case studies that exemplifies the importance of multiscale modeling and simulation in combination with experimental techniques to design and develop more sustainable materials for structural applications in the future.

F. J. Martin-Martinez · M. J. Buehler (✉)
Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: fmartinm@mit.edu; mbuehler@mit.edu

1   Introduction

Developing more resilient, lighter and stronger structural materials to support our current technologies and infrastructure, while coping with sustainable principles and supporting a circular economy requires a deep fundamental understanding of materials chemistry and hierarchical structure that can be only be achieved by a synergistic combination of theory and experiment. The approaches to model and simulate the key structural features of any material strongly depend on the scale at which the key underlying mechanisms take place, as well as on the desired performance needed for the intended application. Thus, a proper description of materials structure and properties through the implementation of this multiscale modeling paradigm provides a predictive framework to design and develop new materials under sustainable principles. It also allows a deeper understanding of the fundamental mechanisms that enable the performance of the materials, to provide the foundations for developing novel materials, recycling them, or identifying new applications. In this section, as we navigate from silk, to aerogels, asphalt, wood, and concrete case studies, we describe key structural features of these materials along different length scales, and how their complex composition (e.g., asphalt, concrete), their hierarchical structure (silk, lignocellulosic biomass, bamboo), or their specific nanoscale molecular arrangement (e.g., aerogels) is important to generate those unique properties that define the applications of these materials.

2   Methods

Different modeling and simulation methods are applied to describe the properties of the materials at different time and length scales. Starting from quantum mechanics methods based on density functional theory (DFT) to molecular dynamics (MD) simulations, coarse grained (CG) modeling, and finite elements methods (FEM), we review how to connect the study of chemical reactivity and mechanical properties from the nanoscale to the mesoscale and up to the continuum scale. DFT is applied to study small molecular systems like lignin fragments, and periodic systems varying from polymers to cement crystals, and how these can be tackled by these fundamental calculations. In some cases, conceptual DFT descriptors are also employed to explain qualitatively and quantitatively the reactivity of any molecular system. Polarizability, dipolar momentum, and also molecular electrostatic potential (MEP) are discussed specially in chapters related to wood. At a larger scale, MD simulations with reactive force fields, as well as nonreactive, are discussed in most of the systems in the section, i.e., silk, aerogels, concrete, asphalt, wood. For example, in the particular case of silica aerogels, MD modeling and simulations are shown to play a crucial role in advancing the fundamental understanding of structure-property-function relationships. Both fully atomistic and CG MD modeling methods are discussed to characterize aerogels' structural, mechanical, and thermal properties. In the case of silk, dissipative particle dynamics (DPD) are

presented as a remarkable approach to develop a mesoscale framework to study the self-assembly of silk proteins into polymer networks and to model the effect of shear flow and terminals on the processed silk fiber. In bulk materials, observations from MD, discrete element method (DEM), and FEM modeling at multiple length scales are used to further understand the role of cohesive-frictional interactions in the strength development of cementitious materials. This is also closely connected to the methods used for studying bamboo, and asphalt, as presented in the related chapters of the section. All these methods together provide a sample of the diversity, applicability, and impact of multiscale modeling when describing the properties of any structural material.

3   Case Studies

Starting from materials whose properties and structural applications are closer to the nanoscale (e.g., silk) to those bulk-like materials whose applications are more closely related to their use in a massive quantity (e.g., asphalt, wood, or concrete), a sequence of case studies helps us to extract common methodologies related to design, modeling, and simulation, in connection with experimental production and characterization.

Silk is a unique natural material whose mechanical properties have intrigued scientist and engineers for decades (Tarakanova and Buehler 2012). Its unique hierarchical structure underlies an outstanding performance in many different applications related to medicine, materials science, and engineering. In Nature, silkworms and spiders have developed glands to produce different types of silk for prey capture or as protective shells. Nowadays, there are available recombinant synthesis techniques and microscale processing approaches to produce silk with tailored properties. As discussed in ▸ Chap. 63, "Silk-Based Hierarchical Materials for High Mechanical Performance at the Interface of Modeling, Synthesis, and Characterization," by Tarakanova et al. a wide range of applications has been achieved by combining modeling, synthesis, and characterization.

Reaching larger scale applications, porous, lightweight, robust, and cost-effective silica aerogels (Sangha et al. 2011) have been developed from supercritical drying of wet gels. As described in ▸ Chap. 64, "Silica Aerogels: A Review of Molecular Dynamics Modelling and Characterization of the Structural, Thermal, and Mechanical Properties," by Yeo et al. this is made by introducing extreme porosity while, at the same time, maximizing their strength and toughness. Experimental and simulation studies on silica aerogels have led to commercialization in numerous engineering applications, like aerospace and civil engineering. In this direction, multiscale modeling techniques have helped the application of aerogels in Cherenkov radiators, integrated circuits, solar energy collection devices, and thermal insulation, as discussed in the chapter dedicated to aerogels.

At an intermediate scale between the applications of silk biomaterials and the large-scale applications of asphalt or concrete, wood is an intriguing material that displays also a hierarchical structure starting from the nanoscale to

achieve an outstanding mechanical performance at the macroscale. As a case study, the increasing interest in the use of bamboo is remarkable, as described in ▶ Chap. 65, "Toughening and Strengthening Mechanisms in Bamboo from Atoms to Fibers," by Askarinejad et al. Bamboo is a fast-growing, naturally occurring bio-composite material, whose use as an eco-friendly structural material has increased dramatically during the last decade. To get further understanding of bamboo structure and performance, numerical methods such as FEM have been applied. This has supported many experimental studies on bamboo, including measuring strength, Young's modulus of matrix and fiber, and analysis of microstructures and fiber distribution.

Despite the intrinsic mechanical performance of bamboo, wood's composition of cellulose, hemicellulose, and lignin makes it useful not only from the structural point of view but also as a source of new materials to be extracted from wood waste. In fact, lignocellulosic biomass is a key feedstock for the development of future sustainable technologies to produce carbon-based materials fuels and chemical, also known as biorefineries. Martin-Martinez et al. (▶ Chap. 66, "Multiscale Modeling of Lignocellulosic Biomass") provide an insight to the current state of the art in multiscale modeling of lignocellulosic biomass (especially lignin) (Sangha et al. 2011), as well as an overview of molecular simulations that have helped understanding lignocellulosic biomass structure.

The development of biorefineries does not diminish the importance to innovate for increasing the sustainability of some petroleum-based technologies. Asphalt is of special growing interest. Being the leftover of petroleum distillation, asphalt is a complex mixture of organic molecules that are difficult to characterize and hard to model. These challenges, together with the fact that asphalt have worked decently well for its intended application with a considerably low cost of production, have hampered the development of technologies devoted to nanoscale control of asphalt properties, as well as modeling and simulations platforms for asphalt. This has changed the last decade, and extensive modeling and characterization are being devoted to understand asphalt composition and structure-property relationships. As discussed by Mullins et al. (▶ Chap. 67, "Simple Asphaltene Thermodynamics, Oilfield Reservoir Evaluation, and Reservoir Fluid Geodynamics"), this research expands to oilfield reservoir evaluation, and reservoir fluid geodynamics. Reservoir crude oils are a complex mixture consisting of dissolved gases, liquids, and dissolved (or colloidally suspended) solids, the asphaltenes. It is essential to know the properties of the crude oil in order to know whether to produce the reservoir fluid. Several oilfield case studies are presented by Mullins et al. (▶ Chap. 67, "Simple Asphaltene Thermodynamics, Oilfield Reservoir Evaluation, and Reservoir Fluid Geodynamics") to elucidate major production concerns, showing that "*when new technology is combined with new science, great accomplishments come into view.*"

Closely connected to infrastructure applications, Palkovic and Büyüköztürk (▶ Chap. 68, "Multiscale Modeling of Cohesive-Frictional Strength Properties in Cementitious Materials") show how computational modeling has been a valuable tool for studying the role of molecular structure and chemical composition in

cementitious materials. Their chapter described the linking of molecular details to macroscale properties that define the durability of the material. This is done through a holistic multiscale approach that translates the essential physics across atomistic, mesoscale and microstructural length scales. Observations from MD, discrete element method (DEM), and FEM modeling at multiple length scales are combined to further understand the role of cohesive-frictional interactions in the strength development of cementitious materials (Palkovic et al. 2016).

All these case studies exemplify the potential of multiscale modeling to elucidate fundamental mechanisms and to enable the rational design and manufacture of structural materials, whose properties can be tailored along different scale for an uncountable number of applications.

4   Conclusions and Perspectives

Moving forward, many challenges are still ahead for the development of more sustainable structural materials. One of the challenges is related to computational modeling, both in terms of new algorithms and more advanced computers. The implementation of Machine Learning (ML) and Artificial Intelligence (AI) into the already established computational methodologies has the potential to contribute to significant speed up of simulations. In a similar manner, the use of cloud computing and the entrance of big corporations into cloud computing applied to materials science will dramatically increase the computational resources available for modeling and simulations. The other big challenge is related to the production of the materials. New advances in additive manufacturing, microfluidics, and synthetic biology will enable an unprecedented level of control over the structure of any materials.

In the case of silk-based materials, multiscale modeling methods are expected to overcome still-existing bottlenecks to accurately describe the mesoscale. This limitation also applies to most of the materials discussed in this section for which the gap between nanoscale mechanisms and macroscale behavior imposes a significant challenge. In the case of silica aerogels, the next big leap is expected in the manipulation of their molecular structures in a bottom-up production approach, leading to aerogels that are stronger, tougher, and more optically transparent, while retaining their low weight and thermal conductivity. In a different field, understanding lignin structure and composition, as well as its integration with cellulose and hemicellulose into the hierarchical structure of lignocellulosic biomass, is crucial for deciphering biosynthetic pathways, producing plants with more easily degradable cell walls, and developing better technologies for biomass depolymerization. At the molecular level, the structure of bamboo microfibrils composed by crystalline and amorphous cellulose in the matrix of lignin-carbohydrate complex (LCC) required also mesoscale modeling to fully decipher the connection between LCC structure and strength and stiffness. In the case of cementitious materials, improved parameterization of cohesive-frictional force fields with molecular models of C-S-H that implement different chemical compositions provides new possibilities

for future exploration. This is especially important for understanding the role of additives. Also, microstructure models to consider mortar and concrete are still a challenge. The development of more advanced computational methods will also likely contribute to a much broader understanding of many oil reservoir concerns over space and time.

References

Palkovic SD, Brommer DB, Kupwade-Patil K et al (2016) Roadmap across the mesoscale for durable and sustainable cement paste - a bioinspired approach. Constr Build Mater 115:13
Sangha AK, Petridis L, Smith JC et al (2011) Molecular simulation as a tool for studying lignin. Environ Prog Sustain Energy 31:47
Tarakanova A, Buehler MJ (2012) A materiomics approach to spider silk: protein molecules to webs. JOM 64:214

Silk-Based Hierarchical Materials for High Mechanical Performance at the Interface of Modeling, Synthesis, and Characterization

63

Anna Tarakanova, Davoud Ebrahimi, Zhao Qin, and Markus J. Buehler

Contents

Abstract

Silk is an exceptional emerging material for diverse applications in medicine, materials science, and engineering. In the natural world, many organisms, including silkworms and spiders, have developed glands to produce different types of silk for applications ranging from prey capture to protective shells. In the lab, synthetic silks have emerged with the development of recombinant synthesis techniques and microscale processing approaches. In recent years, the combination of experimental work and computational investigation has made

A. Tarakanova · D. Ebrahimi · Z. Qin
Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: annat@mit.edu; davebra@mit.edu; qinzhao@mit.edu

M. J. Buehler (✉)
Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: mbuehler@mit.edu

great progress in characterizing and exploiting silk as a functional material for a myriad of applications. Silk can be characterized first as a sustainable material, as it is easily and controllably degradable. Moreover, it is biocompatible and biodegradable, finding many applications in medicine. Silk's properties stem from the unique hierarchy in its structure and the specific cooperativity of these different layers of hierarchy. In this chapter, we will give an overview of silk's unique hierarchical structure and resultant properties and provide case studies where a combination of modeling and experimental synthesis and characterization approaches have been used to study various features of silk-like materials. In particular, we will discuss three case studies: (1) responsive chimera silk-elastin-like materials, (2) silk scaffolds for silk-mineralization, and (3) functional applications of regenerated silk for inkjet printing, and filtration membrane synthesis.

1 Overview

Silk is an exceptional emerging material for diverse applications in medicine, materials science, and engineering. In the natural world, many organisms, including silkworms and spiders, have developed glands to produce different types of silk for applications ranging from prey capture to protective shells. In the lab, synthetic silks have emerged with the development of recombinant synthesis techniques and microscale processing approaches. In recent years, the combination of experimental work and computational investigation has made great progress in characterizing and exploiting silk as a functional material for a myriad of applications. Silk can be characterized first as a sustainable material, as it is easily and controllably degradable. Moreover, it is biocompatible and biodegradable, finding many applications in medicine. Silk's properties stem from the unique hierarchy in its structure and the specific cooperativity of these different layers of hierarchy. In this chapter, we will give an overview of silk's unique hierarchical structure and resultant properties and provide case studies where a combination of modeling and experimental synthesis and characterization approaches have been used to study various features of silk-like materials. In particular, we will discuss three case studies: (1) responsive chimera silk-elastin-like materials, (2) silk scaffolds for silk-mineralization, and (3) functional applications of regenerated silk for inkjet printing, and filtration membrane synthesis.

2 Silk Properties in Nature

In nature, silks are produced by a diverse family of organisms, including silkworms, spiders, honeybees, beetles, ants, and even mussels and certain crustaceans (Sutherland et al. 2010). The properties of silks produced by these different organisms vary based on primary protein sequence and the specific processing environment within the anatomy of the insect. There is tremendous variety even within a single family. Spiders, for example, may spin up to seven mechanically different silk

fibers, and this diversity stems from the unique biology of the spider's body that contains multiple glands that modulate chemical and mechanical processing of the silk (Guerette et al. 1996). The range of properties associated with silk results in a broad spectrum of functions.

Silk is chiefly known for its superb mechanical properties. It surpasses high energy-absorbing materials such as Kevlar and carbon fiber and provides an extremely lightweight alternative (Gosline et al. 1999; Agnarsson et al. 2010; Swanson et al. 2009; Vollrath 2010). A combination of high tensile strength on par with high-grade alloy steel (at 1–2 GPa (Swanson et al. 2009)) and extensibility (up to 60% strain at failure (Swanson et al. 2009)) results in superior toughness and exceeding performance when normalized by its density.

Tensile properties are dependent on humidity conditions as water plays the role of a plasticizer in silk (Plaza et al. 2006). In fact, when humidity is increased and water is absorbed through the silk fiber, the fiber undergoes significant shrinkage and reversibly converts the materials into a rubber (Liu et al. 2005; Shao and Vollrath 1999; Gosline et al. 1984). This process is termed supercontraction (Liu et al. 2005) and renders silk a great candidate material for lightweight actuators in micromachines, sensors, biomimetic muscles, and green energy applications.

Similar to other materials with a high content of amorphous structure, silk undergoes a glass transition, from a glassy to a rubbery state upon increasing temperature (Plaza et al. 2006). Tensile properties of silk fibers transition from stiff to compliant as temperature are increased. Unlike synthetic fibers, silk has higher toughness at lower temperature (Yang et al. 2005). In fact, its retention of exceptional material properties across a wide temperature range ($-66\,°C$ to $100\,°C$) suggests applications as a super-fiber in harsh environments (Yang et al. 2005). Silk's mechanical properties paired with its unique responsive features provide a unique platform for developing biomimetic fibers that are strong, lightweight, and elastic and undergo reversible phase change or supercontraction.

3 Structure of Spider Silk

Spider silk is a highly hierarchical system (Table 1, Fig. 1a) with evolutionarily "optimized" structure at many length scales (Tarakanova and Buehler 2012). We survey these optimization schemes to give a flavor of the many levels of control that are theoretically available to design multiscale morphology for nanoengineering applications.

Silk protein is composed of beta-sheet nanocrystals interspersed within less orderly amorphous domains, where the underlying molecular structure is dominated by weak hydrogen bonding. Beta sheet crystals contain a dense network of hydrogen bonds, while dispersed hydrogen bonds join together amorphous domains, composed of unordered structure, extendable helices, and beta turns (Tarakanova and Buehler 2012).

In fact, nanocrystal size, moderated by the reeling speed of the silk, is the primary factor controlling the strength and toughness of silk (Keten et al. 2010). Smaller crystals exhibit superior properties up to a critical size of 2–4 nm, as a result of the

Table 1 Summary of structural hierarchy of spider silk spanning length scales from angstroms to centimeters, along with key mechanisms that define its mechanical function. (Table modified from Tarakanova and Buehler 2012)

Structure	Hierarchical level	Description	Key mechanism and experimental validation
Amino acid sequence	Primary protein structure (Å)	The primary protein structure is composed of a sequence of amino acid residues which define the folding of the protein. In silk, poly-alanine regions are responsible for stiff beta-sheet rich domains while glycine rich regions define structures within amorphous domains	Basic building block that dictates secondary protein structure. Experimental techniques such as NMR and X-ray diffraction are limited in their atomic resolution for certain complex materials, so molecular dynamics permits a discrete qualification of primary structure through folding mechanisms. Recent work has been successful in sequencing silk genes and using this information to generate recombinant engineered versions of the natural protein (Ayoub et al. 2007). Such changes of amino acid sequence can be experimentally realized through genetic engineering or peptide synthesis (Rising et al. 2010)
Beta-sheet nanocrystals	Secondary protein structure (nm)	The secondary protein structure defines local conformations, such as alpha helices and beta turns. Beta sheet crystals in silk have a critical size of 2–4 nm, exhibiting exceptional strength due to cooperative hydrogen bonding between the sheets. This invokes a highly dissipative stick-slip mechanism that provides the basis for larger-scale toughness and strength	Nanoconfinement of crystal optimizes cooperative hydrogen bond activity, resulting in shear deformation shown to result in most regulated and least catastrophic failure, through the self-healing capacity of hydrogen bond formation. Critical crystal sizes were validated against experiment, based on AFM images (Du et al. 2006) and using diffraction techniques (Du et al. 2006)
Protein composite (beta-sheet crystals, beta turns, 3_1 helices)	Tertiary protein structure ($\sim$10–20 nm)	Tertiary protein structure is the three-dimensional arrangement of interacting secondary structures. The interplay between beta sheet crystals and amorphous domains ensures strength and extensibility of the composite structure	The cooperative interplay and load-balance of protein composite domains achieves a unique hyperelastic material behavior in silk. Molecular dynamics allows the tracking of nanoscale protein-unfolding regimes which are difficult to discern experimentally

<div align="right">(continued)</div>

Table 1 (continued)

Structure	Hierarchical level	Description	Key mechanism and experimental validation
Fibrils	(∼100 nm)	Fibrils exhibit dimensions of 50 ± 30 nm in diameter, where cohesive interaction of protein domains composing the fibril is responsible for homogeneous load distribution upon loading, ensuring optimal strength	Nanoconfinement of fibril size results in homogeneous deformation of all protein domains within the fibril, thereby achieving greatest strength and toughness through contributions of many protein components acting together. Optimal fibril size is validated, e.g., through SEM images (Du et al. 2006)
Fibers	(∼1–2 μm)	Bundles of fibrils into fibers enhance mechanical properties of silk threads	Bundling of several geometrically confined fibrils into a fiber enhances material properties through reaching homogeneous deformation state (flaw tolerance). Structure and geometry of fibers are visible through SEM imaging (Du et al. 2006)
Spider web	(>cm)	Spider web geometry and material composition determine the deformation mechanisms of the web, responsible for optimized prey capture mechanisms	Localization of failure due to a combined yielding-stiffening behavior at deformation. In situ experiments conducted on garden webs show a very similar result (Cranford et al. 2011; Swanson et al. 2007)

difference in deformation mechanism on the molecular scale (Keten et al. 2010). In small crystals, a stick-slip motion is observed as hydrogen bonds brake and reform. The self-healing ability of smaller crystals owing to the continuous reformation of hydrogen bonds protects the crystals from catastrophic failure as hydrogen bonds are shielded from exposure to water, which facilitates rupture. In contrast to soft and brittle larger crystals, which fail catastrophically by crack-flaw formation and consequent bending, smaller crystals exhibited shear deformation. The effective capacity of hydrogen bonds to resist failure by self-healing in conjunction with nanoconfinement at the crystal scale results in strong, stiff crystals of characteristic size, as verified experimentally (Du et al. 2006).

At the next scale, beta sheet crystals are embedded in the amorphous meshwork of less ordered domains. These molecular nanocomposites – crystals embedded in an amorphous matrix – create a network of fibrils that bundle into fibers, forming mature threads. At the fibril scale as well there exists a critical length at which the molecular properties of the protein network are most effectively scaled up to the micrometer scale (Giesa et al. 2011). The underlying mechanism responsible
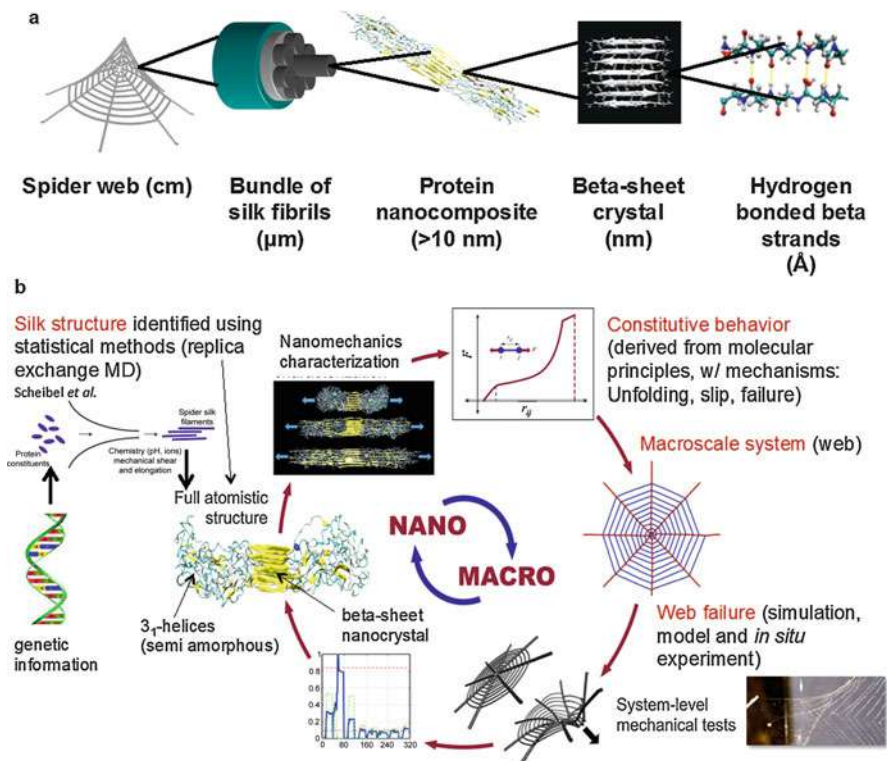
Fig. 1 (**a**) Hierarchical structure of spider silk, from amino acids to an orb web architecture. (**b**) A summary of the multiscale approach used to link the molecular to the macroscale. Silk protein structure, based on the information about its genetic makeup, is identified and characterized to define a constitutive behavior. The molecularly derived behavior is incorporated into a macroscale web system. Web scale deformation experiments link macroscale deformation to molecular mechanisms (Rising et al. 2010). (Figure adapted with permission of Springer)

for this remarkable behavior is a homogeneity of stress distributions as fibril sizes approach the critical length scale. There is a delocalization of stress at the crack tip as fibrils reach the critical fibril size, resulting in completely homogeneous deformation where the entire sample reaches uniform stress concentrations before failure, improving performance (Giesa et al. 2011).

At the macroscale, for instance, in the orb web of a spider, mechanical robustness stems not from spider silk's high ultimate stress and strain, but rather from the nonlinear response of silk threads and their geometrical arrangement within the web (Cranford et al. 2012). When a silk fiber deforms, there is a softening at a yield point and substantial stiffening at large strain until failure. This is critical for localizing load-induced deformation in comparison to standard engineering materials having linear elastic or elastic-plastic deformation behavior. Under distributed loads, such as wind, the stiff behavior of silk threads under small deformation is essential in

maintaining the web's structural integrity (Cranford et al. 2012). This nano-to-macroparadigm, depicted in Fig. 1b, underlines the interplay between hierarchical structure and function through mechanical response.

The structure of silk suggests new design paradigms in applications ranging from high performance fibers and multifunctional materials produced from artificially spun silk or derived materials (Omenetto and Kaplan 2010; Rammensee et al. 2008) to novel material platforms for tissue engineering or even structural applications in the defense or aerospace industry (Buehler and Yung 2009; Vepari and Kaplan 2007).

4 Silk Processing

In addition to silk's unique hierarchical structure, the processing of silk within the duct of the insect is particularly important for the properties of the resultant material. Using dissipative particle dynamics (DPD) coarse-grained models, a mesoscale framework was developed to study the self-assembly of silk proteins into polymer networks and to model the effect of shear flow and terminals on the processed silk fiber (Fig. 2). Models captured the presence of several different "blocks": "A" (poly(alanine) containing, hydrophobic domains), "B" (GGX (X = R, L, Y, or Q) rich, hydrophilic domain), and "H" (hexahistidine fusion tag, introduced for facile purification). It was shown that intermediate ratios of hydrophobic to hydrophilic blocks (domains) of the amino acids in recombinant spider silk result in well-connected networks (Fig. 2a). Moreover, shear flow increased the connectivity of the network of polymers, which translated to outstanding silk fiber formation. Stronger polymer networks were formed when longer polymer chains were present (Fig. 2b). The addition of a charged N-terminal domain (NtSp) improved alignment of the polymer network with the shear flow (Fig. 2c). Moreover, the H(AB)12-NtSp showed greater tendency to aggregate compared with the H(AB)12 (Fig. 2d). Molecular dynamics studies also identified that the shear stress on aqueous silk in the spinning duct of an insect results in helix to beta sheet transitions, key to the strength and function of mature, spun silk fibers (Giesa et al. 2016).

5 Chimera Silk-Elastin-Like Materials

Silk is a good material candidate for developing responsive, tunable, nature-inspired biomaterials. In particular, silk-elastin-like protein polymers (SELPs) have gained attention as bio-inspired composites for their biocompatibility, degradability, and stimuli-responsive tunability (Huang et al. 2015, 2016; Wang et al. 2014; Xia et al. 2011, 2014). In the past two decades, genetically engineered SELPs were shown to be quite versatile. SELPs can be processed in a variety of ways: as nanoparticles, films, nanofibers, thin coatings, hydrogels, and scaffolds, providing a diverse set of structures for material applications (Xia et al. 2011, 2014; Hwang et al. 2009a; Ner et al. 2009; Qiu et al. 2009, 2010; Zhu et al. 2016). These applications include
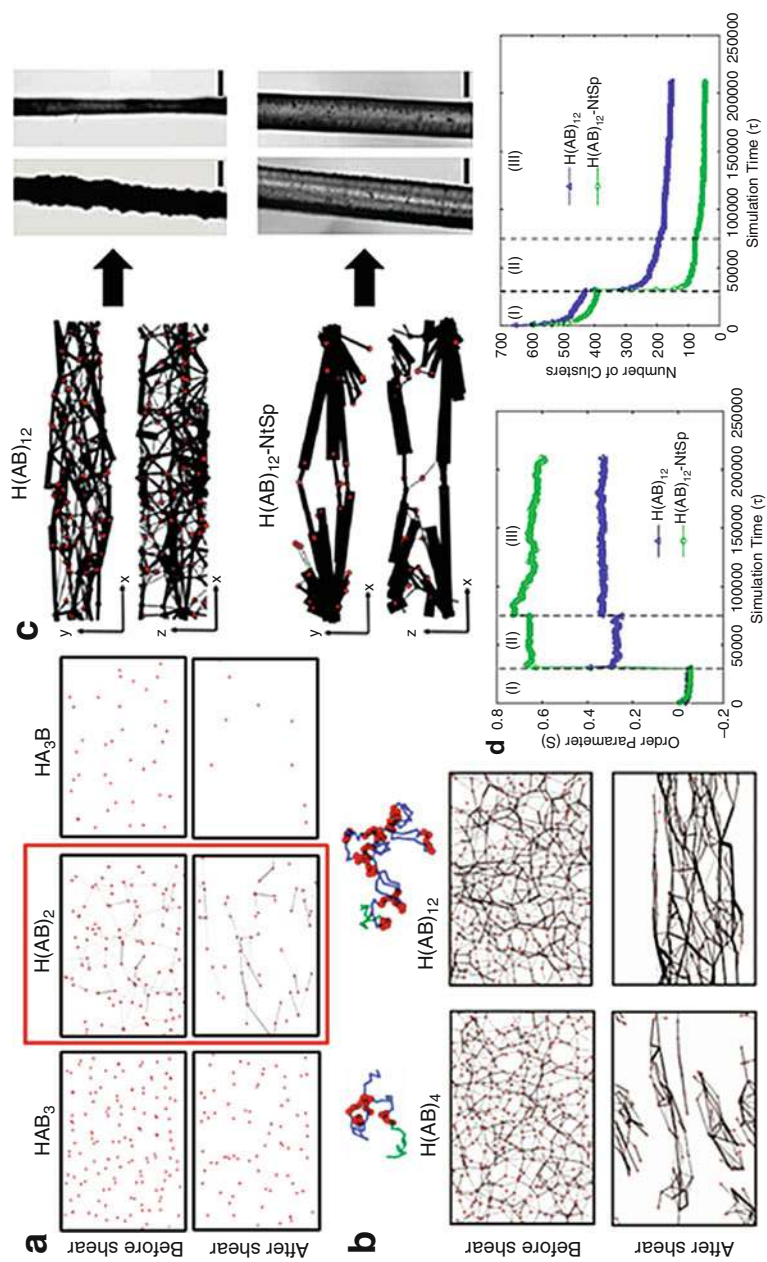
Fig. 2 (continued)

biosensors, tissue engineering, targeted drug delivery release systems, gene therapy, and nanocarriers (Gustafson et al. 2009; Haider et al. 2005; Hwang et al. 2009b; Megeed et al. 2004; Anumolu et al. 2011), among others.

SELPs are composed of alternating silk-like and elastin-like domains, combining the properties of the component parts. Silk-like domains (GAGAGS) mimic the *Bombyx mori* silkworm silk sequence. They assemble into tightly packed structures and provide stability and mechanical resilience. Elastin-like pentapeptide domains (GXGVP) are representative of the elastin protein sequence and exhibit an inverse temperature transition, modulated by changing the second X residue of the pentapeptide. SELPs combine the mechanical strength, resilience, and self-assembling properties inherent to silk together with tunable mechanics derived from the elastin domains, which in physiological conditions exhibit reversible sensitivity to stimuli, including temperature, pH, ionic strength, electric fields, and enzymes (Huang et al. 2016). By combining silk-like and elastin-like domains, SELPs achieve useful mechanical properties and discrete tunability.

Molecular modeling is instructive for studying these materials to obtain a molecular perspective of their unique features. Together, molecular modeling and experiments based on recombinantly synthesized SELP sequences can be combined to probe the molecular scale temperature transition effects and single-molecule mechanical responses to thermal stimulation. In particular, a recent SELP model has identified thermally stimulated structural transitions at the single molecule scale and extracted temperature effects on molecular unfolding pathways and mechanical signatures.

Silk-elastin-like protein polymers, based on the amino acid sequence $[(GVGVP)_4(GYGVP)(GVGVP)_3(GAGAGS)]_{14}$, a 14-mer built from units composed of eight alternating silk-like and elastin-like blocks (Fig. 3a), can be synthesized using recombinant DNA technology. The turbidity profiles for 1 mg/ml of the SELP aqueous solution structures display a temperature transition range between 28 °C and 45 °C, henceforth referred to as T_t (Fig. 3b). To distinguish structures below and above transition range, two systems are considered: at 7 °C and 57 °C, well below and above T_t. Below the transition, at 7 °C, the SELP

Fig. 2 Modeling, production, and processing of silk block co-polymer systems. (**a**) Modeling at the mesoscale coarse grain shows that alternating hydrophobic/hydrophilic domains at a 1:1 ratio promote network formation over other block distributions. (**b**) As length is scaled, mature networks form through shear flow, which can be correlated to potential targets for physical processing and spinning. (**c**) Additional structures can be modeled using this framework, including the addition of the N-terminal domain (NtSp), which is predicted to promote better molecular alignment. Accompanying production and processing, model inputs lead to effective predictors for spinning, which was demonstrated from the core H(AB)12 system to the N-terminal variant. (**d**) Larger order parameter means better alignment of the polymer networks with the shear flow. In all cases shear results in greater alignment in the direction of shear, with H(AB)12-NtSp aligning the most. Moreover, the N-terminus caused clusters to consolidate, leading to unique behavior of continuous aggregation, even after shear. (Figures adapted from Lin et al. (2015), Rim et al. (2017), Jacobsen et al. (2017) with permission)
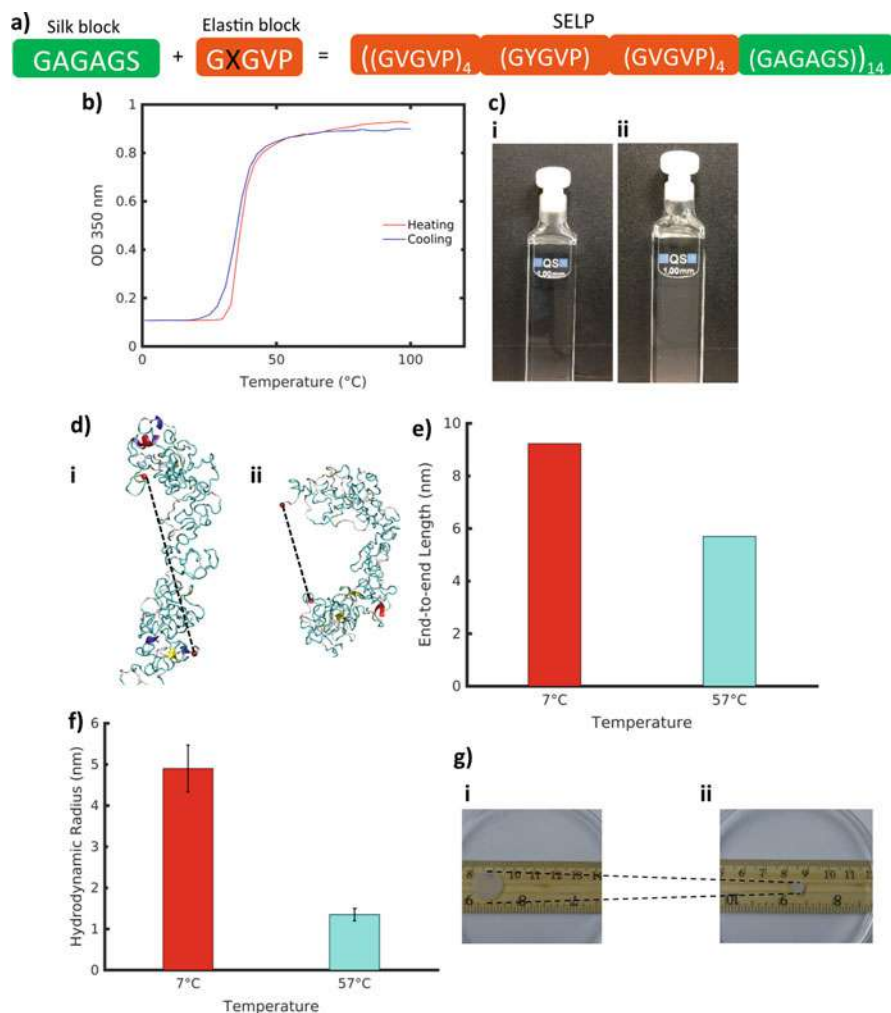
Fig. 3 (**a**) SELP sequence considered in experiment and simulation. SELP is composed of alternating silk-like (GAGAGS) and elastin-like (GXGVP) blocks, where X represents the interchangeable residue responsible for modulating the transition temperature. (**b**) UV spectrophotometry heating and cooling curves show a reversible transition range between 28 and 45 °C. (**c**) SELP samples at (i) low (7 °C) and (ii) high (57 °C) temperature. (**d**) Representative SELP structures from simulation, at (i) 7 °C and (ii) 57 °C. Dotted lines represent end-to-end molecular distance. (**e**) End-to-end distance of representative SELP structures pictures in d) at 7 °C and 57 °C. (**f**) Hydrodynamic radius from dynamic light scattering of SELP at 7 °C and 57 °C. (**g**) SELP hydrogel samples at (i) 7 °C and (ii) 57 °C. (Reprinted with permission from Tarakanova et al. (2017a). Copyright 2017 American Chemical Society)

solution appears transparent. As the temperature is raised above the transition for the polymer, to 57 °C, the SELP solution becomes cloudy, indicative of the phase transition taking place (Fig. 3c).

To capture this behavior at the nanoscale, SELP molecular structures can be predicted using Replica Exchange Molecular Dynamics simulations (Fig. 3d). At the single-molecule scale, the transition occurs as a gradual structural change, manifested by a decreasing radius of gyration with temperature. Both the molecular models and the synthesized polypeptides exhibit up to 90% of unordered secondary structure with minimal beta sheet and helical content, consistent with the high content of elastin-like sequence within the polymer. A comparison between representative structures at 7–57 °C shows a structural collapse at high temperature as the structure bends across the principal axis and assumes a more compact conformation (Fig. 3d). Measurement of the end-to-end distance confirms a reduced molecular size (Fig. 3e).

A similar trend to a compacted structure with increasing temperature is observed in the synthesized polymer by measuring the hydrodynamic radius (R_h) of the SELP free chain. R_h at temperatures below and above T_t is determined by dynamic light scattering (DLS). DLS measurement at 7 °C and 57 °C displays a reduction in the R_h of the free chain from 3.8 ± 0.6 nm to 1.4 ± 0.2 nm (Fig. 3f). Such drastic reduction in molecular size suggests that the free chains of the SELP folds at high temperature above T_t, leading to a decrease of the overall size of the SELP free chains.

At the macroscale, a corresponding behavior is observed in silk-elastin hydrogels that are fabricated and tested at different temperatures. A shrinkage of 57% in the hydrogel radius is measured between 7 °C and 57 °C (Fig. 3g). The structural collapse observed at the single molecule scale is propagated up to the macroscale. It is proposed that the packing geometry of SELP molecules in the hydrogel ultimately propels the shrinking phenomenon, causing structural reorganization at the molecular scale.

Modeling at the molecular scale permits characterization of temperature-dependent mechanics at a fundamental scale. Unfolding pathways for representative SELP structures below and above T_t, at 7 °C and 57 °C, respectively, can be studied using steered molecular dynamics simulations. Drastically different deformation behavior can be observed at low and high temperatures (Fig. 4a–c). At 7 °C, the mechanical response includes a steep linear regime followed by a plateau, compared with a gentle linear slope at 57 °C. The unraveling of the high temperature structure reveals a smooth unfolding mechanism (Fig. 4d). The molecule unraveled from a single cluster, uniformly like a ball of yarn, corresponding to a linear deformation regime. By contrast, as the low temperature structure unfolds, a plateau in the force-displacement appears at a displacement of about 150 nm (Fig. 4e). A series of smaller clusters detaches from the main densely folded region as the molecule is pulled. These clusters act to dissipate the force, resisting the pulling, and produce the plateau that differentiates the low and high temperature deformation curves.

To understand the internal molecular landscape through the unfolding process, the intramolecular hydrogen bond evolution at high and low temperature is considered (7–57 °C, respectively). Notably, the number of hydrogen bonds that exists
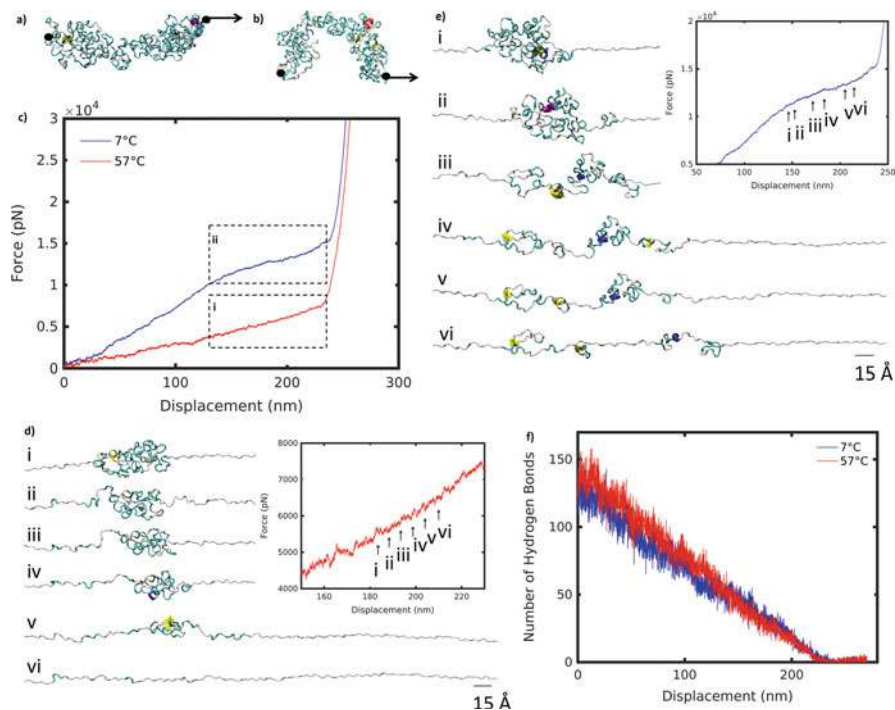
Fig. 4 Steered molecular dynamics set-up for SELP at (**a**) 7 °C and (**b**) 57 °C. Molecule is fixed at the C-terminal end and pulled at the N-terminal end. (**c**) Force-displacement curves at 7 °C and 57 °C for pulling speed 50 m/s. Unfolding snapshots for region (i) (**d**) and (ii) (**e**), in plot of (**c**), respectively. Structures correspond to displacements indicated by numbered arrows on the force-displacement plot. Scale bar: 15 Å. (**f**) Hydrogen bond evolution at 7 °C and 57 °C during SMD pulling. (Reprinted with permission from Tarakanova et al. (2017a). Copyright 2017 American Chemical Society)

within the SELP molecule is higher by 15% at high temperature, and this difference persists through two thirds of the unfolding stages (Fig. 4f). This observation is counterintuitive, as higher temperature is expected to more easily disrupt weak intramolecular hydrogen bonds, and is attributed to the presence of elastin-like segments, which are known to assume a folded, hydrogen-bond-rich conformation above T_t. The enhanced hydrogen bonding at high temperature may help to maintain a large, uniform region in the unfolding pathway. Fewer hydrogen bonds at low temperature create a less compact structure that permits the separation of small, independent clusters as the molecule experiences a pulling force. In fact, it is predicted theoretically that the energy barrier to unfold the structure at 57 °C is greater than the energy barrier at 7 °C due to the presence of a dense network of hydrogen bonds at high temperature (Tarakanova et al. 2017b).

The inverse temperature transition observed in SELPs is a reversible, controllable process that motivates applications in drug delivery, shape changing materials,

and biomimetic devices. Using a combination of experimental, modeling, and theoretical methods, the structure and nanomechanics of SELPs can be identified at different temperatures. These findings offer a foundational perspective on the molecular scale mechanisms of temperature-induced phase transition in SELPs and suggest a novel approach to combine simulation and experiment to study materials for multifunctional biomimetic applications.

6   Silk as a Scaffold for Silk-Mineralization

Biomineralization is the process of producing minerals (inorganic material) by proteins (organic material) in living organisms. Through complex hierarchical assembly of composite organic-inorganic material, mineralized tissue becomes stiffer and harder (Launey et al. 2010). Understanding the organic-inorganic interface to fine-tune mineralization is the key to control biomaterial function in repair and regeneration of tissues, preventing the process in dynamic tissues as in the case of calcific aortic valve disease where calcium deposition prevents its function, or promoting mineralization in disease like osteoporosis where one loses bone density.

Recently, silk protein-based biomineralized composites have been studied as tissue scaffolds. An efficient scaffold material for tissue regeneration should be biocompatible with suitable stiffness, strength, and degradability (Chan and Leong 2008). Silk is a biocompatible protein approved by the US Food and Drug Administration to be used in some biomedical applications (Altman et al. 2003). Strength, stiffness, and degradability of the silk protein could be tuned by controlling its crystalline beta sheet content (Nova et al. 2010; Dinjaski et al. 2018). A suitable scaffold should mimic the structure and properties of natural extra cellular matrix (ECM) (Polo-Corrales et al. 2014). Moreover, ECM-cell bidirectional signaling interactions should be understood and controlled. It is known that ECM properties such as biochemistry, structure, and mechanics could change the interplay between genotype and phenotype (Benoit et al. 2008; Teixeira et al. 2004; Engler et al. 2006). Silk-hydroxyapatite (HA) (Bhumiratana et al. 2011) and silk-silica (Mieszawska et al. 2010a) scaffolds have been studied for bone tissue repair with human mesenchymal stem cells (hMSCs) to regenerate bone tissue. HA or silica promoting peptides are tethered onto silk protein and provide biochemical and signaling properties to enhance hMSCs differentiation in addition to adsorption of ions to form minerals (Marolt et al. 2006; Rockwood et al. 2011; Meinel et al. 2004).

In modular protein engineered biomaterials, different peptide building blocks could be combined to generate different scaffolds with diverse properties (Sengupta and Heilshorn 2010). For instance, Fig. 5 shows an integrated modeling and experimental approach for design of silk-based biomaterials and control of silicification (Dinjaski et al. 2016). The study provides valuable insight into key factors to control mineralization such as protein folding, and exposure and alignment of charged units. Genetically engineered constructs were designed with three basic building blocks (Fig. 5a): Histidine tag (green box) is added to facilitate purification, the artificial
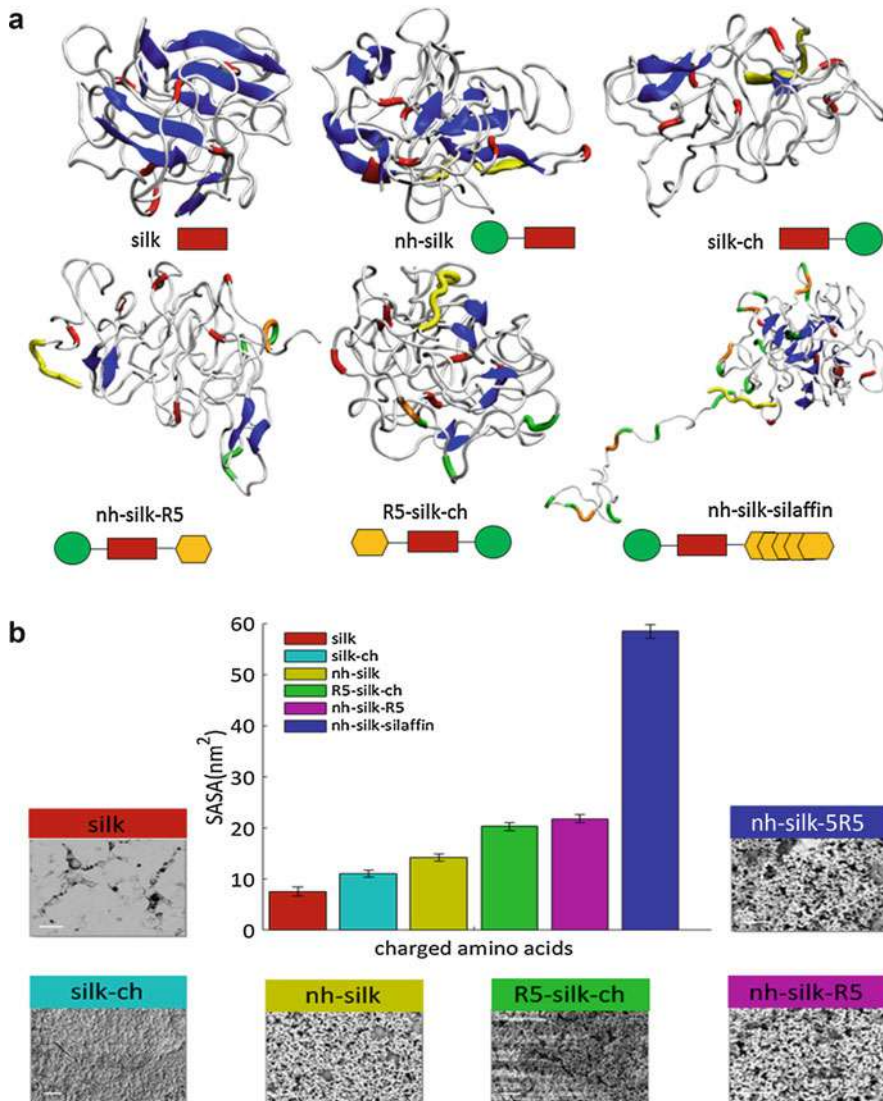
Fig. 5 (**a**) Folded protein structures: His tag (green box) is added to facilitate purification, the artificial silk polymer (red box) is added to form β-sheet structures, and R5 (orange box) is added to facilitate silicification. (**b**) Solvent accessible surface area for positively charged amino acids correlates with the amount of silicification in experiment. (Figures adapted from Dinjaski et al. (2016) with permission)

silk polymer (red box), derived from the consensus repeat of *N. clavipes* dragline silk protein, is added to form β-sheet structures, and R5 (yellow box) domain derived from *C. fusiformis silaffin* gene is added to facilitate silicification. "nh" and "ch" stand for N- and C-terminal fused histidine tags. Figure 5a shows sample snapshots of different protein structures in explicit water. β-sheets and random

coils are colored in blue and white, respectively. Histidine, arginine (in silk block), arginine (in R5 bonding peptide), and lysine are charged residues of the proteins which are color coded in yellow, red, orange, and green, respectively. Results (Fig. 5b) show that the presence of positively charged amino acids is essential for catalyzing biomineralization; however, the process is not amino acid specific. Moreover, the location of the charged biomineralization domain in the fusion protein could affect protein folding and consequently surface exposure of charged amino acids. In summary, solvent accessible surface area for positively charged amino acids correlates with the amount of silicification in experiment. Silk and silk-ch have the least accessibility of charged amino acids, resulting in no silicification and low silicification in experiments, respectively. On the other hand, nh-silk-R5 and nh-silk-silaffin (5R5) show the highest amount of mineralization. The study furthers the understanding of intracellular pathways involved in bone regeneration triggered by recombinant silk–silica chimeras (Martín-Moldes et al. 2018).

The same three building blocks were used to construct silk-silica interfaces. Then the intracellular pathways involved in the process of mineralization were assessed, when stem cells were grown on those silica substrates. Integrins are a family of cell adhesion receptors consisting of heterodimeric transmembrane proteins called α and β subunits. They interact with ECM proteins, triggering activation of intracellular pathways, such as mitogen-activated protein kinase (MAPK). Figure 6a–c shows modeling of the headpiece from integrin αVβ3 in solution, in contact with the silica surface and the silk-chimera protein surface. Similar models were prepared for α5β1 integrin. Figure 6d shows gene expression analyses of hMSCs seeded on 15mer-ch and R5-15mer-ch films with or without silicification and on tissue culture plastic (TCP) for one day. When compared to TCP, cells seeded on nonsilicified films showed no difference in expression of integrin subunits. Meanwhile, the cells seeded on the silicified films showed increased gene expression of α5, αV, and β3 integrin subunits, but not the β1 integrin subunit. The detected expressions suggest αVβ3 and α5β1 as two candidates for triggering the hMSCs. Based on the results of the MD simulations (Fig. 6e), only αVβ3 integrin is activated in contact with the silicified surfaces, since the head piece opens up with respect to the integrin in solution, while there is a decrease in head piece distance in contact with silk-chimera protein surface (nonsilicified surface) and for α5β1 integrin in contact with silica surface. The use of a blocking antibody against integrin αVβ3 in experiment validated results of the simulation, and the induction of the osteoblast differentiation genes was abolished. In summary, in the absence of the silica surface (nonsilicified) or in the presence of a neutralizing antibody against αVβ3 integrin, there was no activation of the MAPK cascades, thus no induction of bone extracellular matrix genes and therefore no differentiation.

The main components of bones are apatite crystals, the inorganic mineral phase, collagen fibrils, and the organic protein phase, which control nucleation and growth of the minerals (Nudelman et al. 2010). By combing silk with different amount of minerals such as hydroxyapatite or silica, compressibility of the composite scaffold could be improved which is required to match properties of natural bone tissue at the site of implementation (Bhumiratana et al. 2011). Two types of silk proteins could be used to generate tissue scaffold: natural silkworm silk (Mieszawska
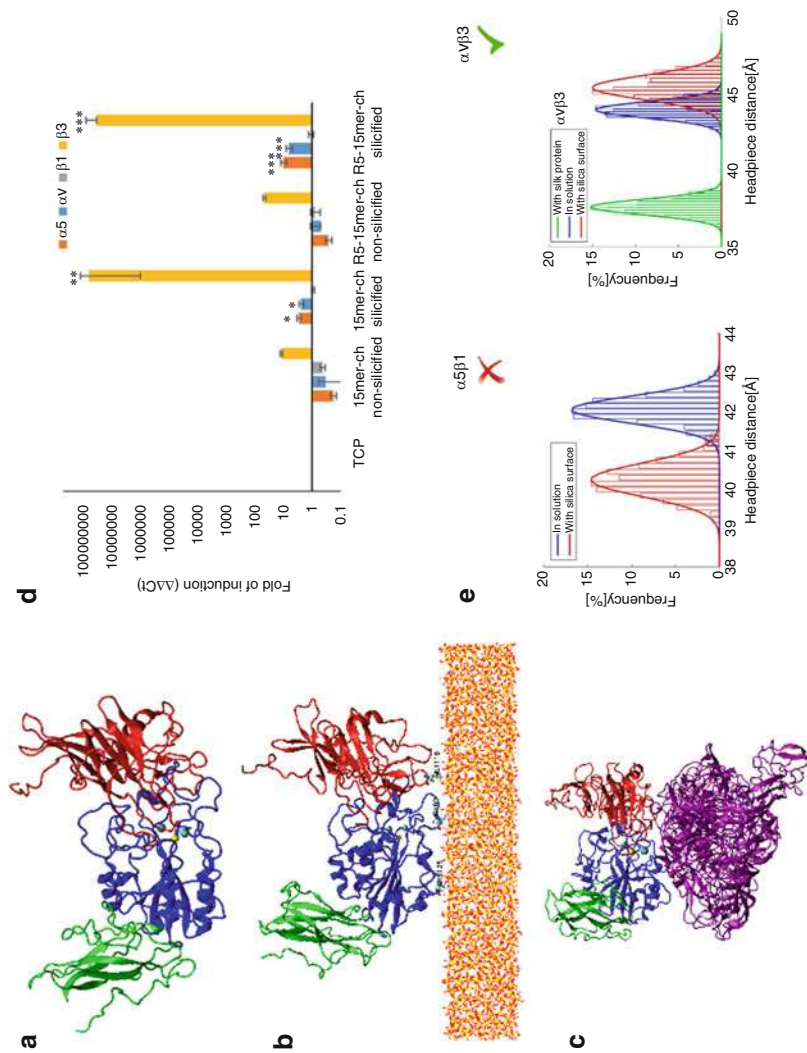
Fig. 6 (continued)

et al. 2010b; Sofia et al. 2001; Liu et al. 2008) and recombinant spider silk (Moisenovich et al. 2012). The integrated modeling/experimental approach could be used to develop modeling guided platform for rational design of recombinant silk proteins for biomaterials and material design needs. This strategy could forecast the influence of the addition of biomineralization domains to silk sequence on the fusion protein structure and folding, and identify the optimal processing parameters by understanding the mechanisms and identifying the key factors involved in the process.

7   Design and Synthesis of Silk-Based Filtration Membranes

Membranes for water treatment have progressively been developed to address global challenges of water pollution in recent years, because they are energy- and waste-efficient to remove molecular-level contaminants (Shannon et al. 2008; Vandezande et al. 2008; Werber et al. 2016). A variety of new materials and novel fabrication methods (e.g., block copolymer self-assembly (Wang et al. 2011; Krieg et al. 2011; Deng et al. 2013), template synthesis (Liang et al. 2010; Peng et al. 2007), track-etching technique (Eguizabal et al. 2014; Striemer et al. 2007), chemical vapor deposition (Mourhatch et al. 2010; Karan et al. 2012), and layer-by-layer assembly (Joseph et al. 2014) have been developed to improve the purification efficiency of these membranes. However, it remains challenging to prepare low-cost water purification membranes while retaining the mechanical strength and high purification performance.

Recently, a de novo strategy was used for making multilayer nanoporous membranes via integration of computational simulation and experiments. The core methodology is to use soft and hard nano-building blocks for constructing multilayer nanoporous structures (Ling et al. 2017). Here, the soft layer with a smaller pore size (several nanometers in diameter) serves as a size-selectivity layer to effectively filter out molecular level contaminants. The hard layer with a large pore size and high porosity allows water to pass through quickly and provides structural supports to the soft layer. In addition, the alternating arrangement of these two layers can further improve permeability of these membranes. Many natural systems, such as nacre, enamel, and bone, use proteins and calcium-based minerals to build their sophisticated multilayer architectures with enhanced physical

---

Fig. 6 (**a**, **b**, and **c**) Simulated head piece domain of the integrin in molecular dynamics (MD) modeling, in solution (water molecules are not shown), in contact with silica surface, and in contact with silk-chimera protein surface. (**d**) Gene expression of α5 (orange bars), αV (blue bars), β1 (gray bars), and β3 (yellow bars) subunits of human mesenchymal stem cells (hMSCs) seeded on tissue culture plastic (TPC). There are two candidates for triggering the differentiation of hMSCs to osteoblasts in contact with silicified surfaces: αVβ3 and α5β1. (**e**) Activation of integrins measured as opening of their head pieces in MD simulations shows αVβ3 integrin and not α5β1 integrin is activated in the presence of silica surface. (Figures adapted from Martín-Moldes et al. (2018) with permission)

properties (Chen et al. 2012). Moreover, both proteins and calcium-based minerals have showed promising application in water purification; thus, we selected these two components as starting building blocks.

In order to disclose the condition in which these two components can form the porous multilayer structures, simple coarse-grained models of the protein nanofibers (NFs) and calcium-based minerals nanoplates (NPs) were developed. Each NF and NP was modelled by an elastic network with carefully chosen elastic constants to reflect the materials' elastic properties and mass concentrated at the junction beads to reflect the material densities. Using the collection of these models, molecular dynamics (MD) simulations are performed to simulate their assembly during the deposition process in the water flow.

As examples, Fig. 7 shows the comparison of the assembly process of NFs and NPs. It is found that by tuning the material stiffness, density, and interfacial interactions between the subunits, the NFs and NPs can either assemble to form a multilayer structure with a clear boundary between two material phases or a
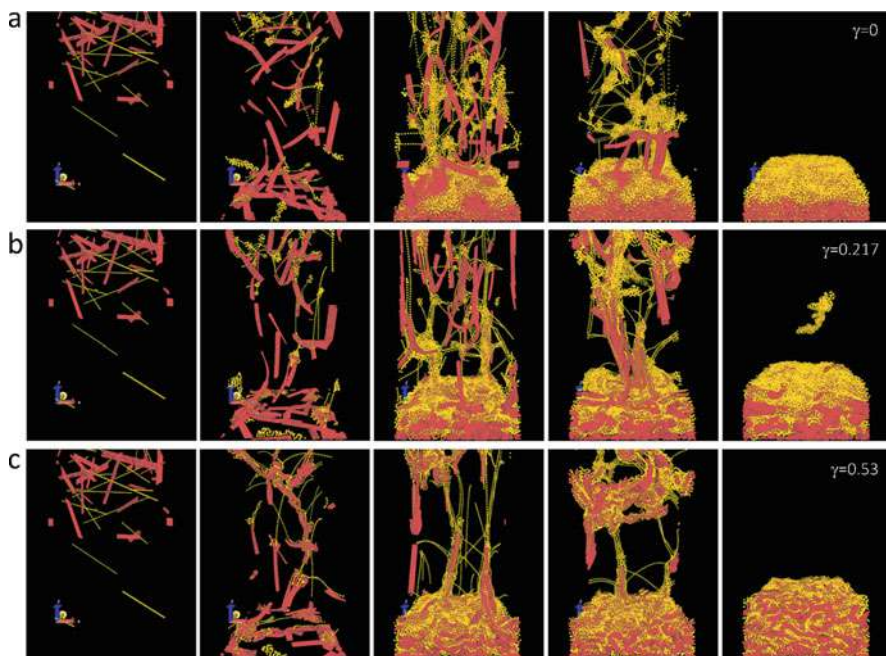
Fig. 7 Coarse-grained computational MD simulations for SNF/HAP assembly and deposition. Elastic network models are built on mechanics and geometry feature of the unit building blocks. HAP-SNF surface energy γ (J/m^2) is not certain as it depends on the molecular structure and the hierarchy structure of the interface in contact. This is important for forming the layered structure. For the SNF/HAP model, HAP and SNF particles were randomly distributed initially and then subjected to a gravity field that accounts for the drag force from the water flow. a, b, and c are the snapshots of the HAP/SNF assembly process during deposition with γ set as 0 (**a**), 0.217 (**b**), and 0.53 J/m^2 (**c**), respectively. (Figures adapted from Ling et al. (2017) with permission)

homogenous mixture of the two materials. It was discovered that for a wide range of mineral density (as the material density of protein is almost fixed) and stiffness of NFs and NPs, the interfacial energy (γ) between NF and NP plays the dominating role to govern the way of their assembly. It is revealed by computational simulations that for weak NF/NP interaction, NFs are assembled into a continuous network during the deposition process and their deposition rates are smaller than that of NP flakes that lack such self-assembly process. Turning to strong NF/NP interactions, NF and NP flakes are well mixed by self-assembly before being deposited on the substrate. Accordingly, a NF/NP interaction weaker than both NF/NF interaction and NP/NP interaction ($\gamma < 0.217$ J/m^2, which is the average surface energy measured for protein filament assembly) is necessary to form the layered structure (Fig. 7). This discovery is critical for selecting the suitable combination of materials to form the multilayer structure, because many protein/mineral combinations have strong interfacial interactions and are not suitable to be used as raw materials for forming the multilayer structure.

On the basis of simulation findings, different material combinations were scanned and silk nanofibril (SNF) and HAP were chosen to conduct experiments. It is noted that silks in nature do not have the propensity to form mineralized bulk materials as collagen and experimental evidence shows that SNF can easily exfoliate from the HAP surface to validate its weak interaction with HAP. Moreover, different from a liquid exfoliation (top-down) route, here, a new and more energetically efficient approach was developed to produce SNFs, that is, to control the self-assembly of silk protein. First, aqueous silk solution (0.1 wt%) was assembled into elongated SNFs with a height of ∼3 nm and a contour length up to 5 µm. Then, the SNFs were used as templates to grow HAP nanosheets through in situ biomineralization. These HAP nanocrystals, with a length of 100–300 nm and a width of 20–30 nm, showed a uniform needle-like morphology and were stabilized by the connected SNF networks. Thus, such HAP nanocrystals were finely dispersed and highly stable in the SNF solution, in contrast to the flower-like HAP aggregates in silk solution from control experiments. Finally, the SNF/HAP dispersion was assembled into water insoluble macroscopic membranes via vacuum filtration. Indeed, these SNF/HAP membranes show highly ordered multilayer structures with ∼200 nm thick alternated HAP and SNF layers (a 4-µm-thick membrane as an example, Fig. 8). The number of layers is tunable by changing the consumption of dispersion (Fig. 8). Typically, 1 mL of ∼0.2 wt% dispersion generates a double-layer membrane, with HAP settling in the bottom layer and the SNFs remaining on the top (Fig. 8a, b). Each layer has a similar thickness of approximately 500 nm. Both layers have uniform dispersions of SNFs and HAPs, respectively, but they show distinct pore sizes (Fig. 8c, d). The SNF layer has a narrow pore size distribution with a diameter of 8 ± 2 nm (could serve as size-selectivity layer). In contrast, the HAP layer has a larger pore size distribution with a diameter of 28 ± 5 nm (can enhance the water permeation). By progressively adding the volume of the SNF/HAP dispersion, the number of layers is increased and reaches tens of layers (Fig. 8f–g). The alternating arrangement of SNFs and HAPs layers can be observed in all cases.
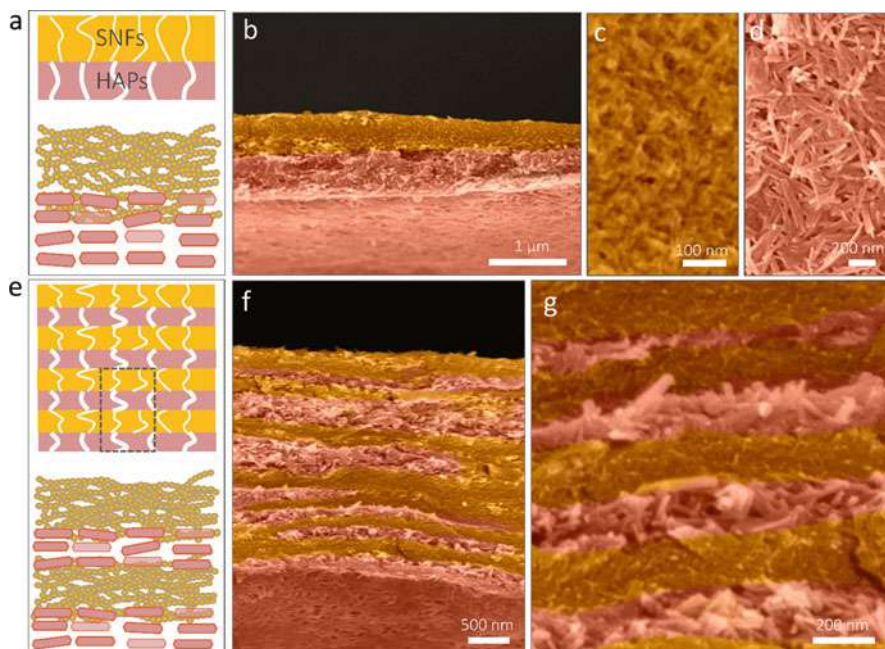
Fig. 8 The well-organized multilayer structures of SNF/HAP membranes. (**a–d**) The double-layer structure of the SNF/HAP membranes was formed through vacuum filtration of 1 mL of dispersion. The picture (**a**) gives the schematic of double layer structures. The top layer is the SNF-rich layer with small pore sizes. The bottom layer is the HAP layer with larger pore sizes. Picture (**b**) is the cross-sectional SEM image of a double layer membrane. Pictures (**c**) and (**d**) are the top view SEM images of SNF-rich and HAP-rich layer, respectively. (**e–g**) The multilayer structure of the SNF/HAP membrane was generated from 3 mL SNF/HAP dispersion using a 3.5 cm diameter mold. Picture (**e**) gives the schematic of multilayer structures. Picture (**f**) is the cross-sectional SEM image of a multilayer membrane. Picture (**g**) is the high resolution cross-sectional SEM image of a multilayer membrane. False color is used in SEM images. (Figures adapted from Ling et al. (2017) with permission)

The core methodology used soft silk nanofibrils and hard mineral-building blocks for constructing multilayer nanoporous structures. Here, the soft layer (protein nanofibrils) with a smaller pore size (several nanometers in diameter) gives size-selectivity to effectively filter out molecular level contaminants. The hard layer (mineral nanoplates) with a large pore size and high porosity allows water to pass through quickly and provides structural support to the soft layer. In addition, the alternating arrangement of these two layers can further improve permeability of these membranes. Comparative computational simulations based on coarse-grained models of protein nanofibrils and mineral plates reveal that the multilayer structure can only form with weak interactions between nanofibrils and mineral plates. On the basis of this finding, a series of protein/mineral systems were scanned and silk nanofibrils (SNFs), and a hydroxyapatite (HAP) system was used to conduct the

experimental fabrication. In the manufacture process, a new biomimetic route was developed by using protein self-assembly and in situ biomineralization. The resultant SNF/HAP membranes indeed present the highly ordered nano-sized multilayer structures. Profiting from these well-organized hierarchical multilayer nanoporous structures, the resultant membranes not only show ultrafast water penetration (over five times greater than other reported ultrathin membranes with similar thickness), but also exhibit universal and high efficiency to remove and even reuse (in some cases) heavy metal ions, dyes, proteins, and nanoparticles. A multilayer membrane, on the basis of the simulation findings, may be formed by using different materials. Natural protein-based (e.g., amyloid, collagen fibril) and engineered peptides and protein-based (e.g., silk-, silk-elastin-, silaffin-, apatite-binding) fibrils and their corresponding biomineral plates (e.g., calcification, silicification) are also suitable for building multilayer nanoporous structures. Therefore, a biomimetic process for membrane synthesis could lead to large-scale production of multilayer membrane materials with a broad-spectrum (and possibility to design for a specific spectrum) and efficiency for water purification, useful in waste-water treatment, biomedicine, food industry, and life science applications.

8 Inkjet Printing of Regenerated Silk

Recently, pure silk fibroin (SF) derived from *Bombyx mori* cocoons has been used as ink for direct printing (Ling et al. 2016). Figure 9 illustrates the general process of the inkjet 3D-printing of the silk membrane as an optical strain gauge. The florescent silk solution was prepared and contained the materials in a syringe right before printing. The concentrations of silk and ions were tested to make sure the viscoelastic property of the material is appropriate so it can be printed without clogging at the front tip of the syringe needle and also keeps a relatively stable form on the substrate surface after printing. Specifically, the degummed silk fiber was dissolved in $CaCl_2$/formic acid (FA) solvent (Zhang et al. 2015) with the concentration of SF ink reaching 30 wt% without any aggregation or gelling (Fig. 9a). The syringe is mounted to an in-home-build 3D printer that can move in all the directions and works by controlling the squeezing speed of the syringe during printing (Fig. 9b). All movements of the printer are digitally controlled by stepper motors and a computer numerical control board. Using such printing facility, different florescent patterns can be printed that are made of silk materials, such as a rectangular network with different florescent colors in two different orthogonal directions and parallel bands (Fig. 9c). Patterns were printed on different kinds of deformable substrates including polypropylene film, rubber, and PDMS because of the high viscoelasticity of silk solution. After rapid curing owing to quick evaporation of FA (about 10 min), the printed silk pattern can get deformed together with the substrate, even for a common rubber glove after wearing (Fig. 9d) and getting stressed in full tension (Fig. 9e).
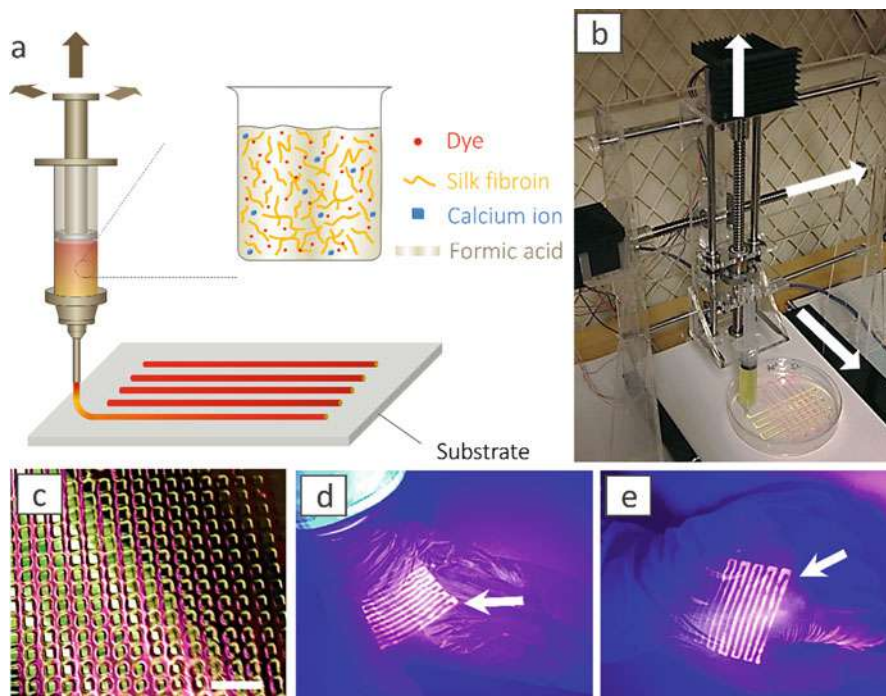
Fig. 9 3D printing of silk. (**a**) Schematic illustration of the 3D printing process. A fluorescent silk/Ca2+/FA ink is printed onto a substrate. (**b**) Photograph of home-made 3D printing device. (**c**) Photograph of printed silk pattern on polypropylene film. The scale bar is 1 cm. (**d**, **e**) The photograph of printed silk pattern on gloves with palm open (**d**) and clenched fist (**e**). The florescent silk pattern can get deformed with the deformation of gloves. (Figures adapted from Ling et al. (2016) with permission)

By using this printing technique, a new method was developed and a set of membrane materials based on silk could be used as accurate optical strain gauges. It is noted that the pure silk fibroin (SF) derived from *Bombyx mori* cocoons is biocompatible and biodegradable due to the absence of toxicity and immunogenicity (Omenetto and Kaplan 2010), making them ideal candidates for in situ and implantable strain measurements (Qin et al. 2015). By using inkjet 3D-printing, silk membrane materials could be accurately printed with a structure according to a specific design. This study shows that the materials are extremely flexible, allowing them to conform to the irregular curved surfaces and bond well to them. Figure 10a presents the schematics and microscopic structure of a prototype of the printed silk strain gauge. The silk membrane and the printed fluorescent silk pattern are bonded very well with each other as both fluorescence microscope and scanning electron microscope (SEM) images confirm a sharp boundary without any significant gap between the two materials in the scale spun from millimeter to micrometer. The
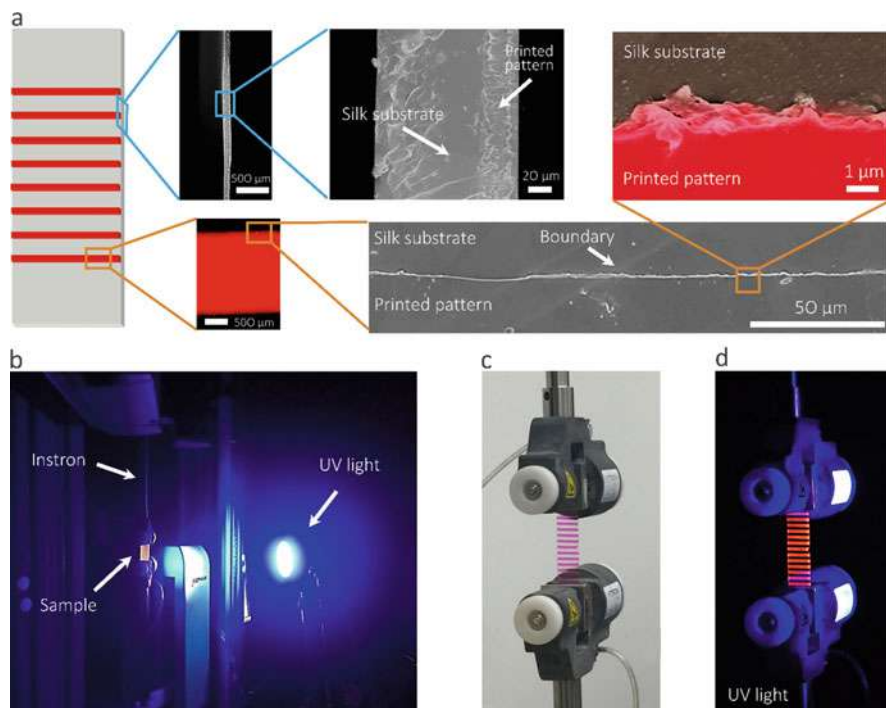
Fig. 10 Morphologic characterization and strain measurement setup of 3D-printing silk membrane. (**a**) The observation of the fluorescence microscope image (the second on bottom left) and SEM images (others except schematic diagram on the left) of 3D-printing silk membrane from cross-section and surface direction. The false color was used in top right SEM image to distinguish printed pattern and substrate. (**b**) The strain measurement setup. (**c**, **d**) Pictures of the 3D-printing silk membrane under visual (**c**) and UV light (**d**) observation. (Figures adapted from Ling et al. (2016) with permission)

printed silk pattern has a uniform thickness of 30 μm. A camera was used to record the deformation of the silk strain gauges under a UV-light irradiation environment (Fig. 10b). The printed silk lines emit strong red light due to the UV excitation of rhodamine B dyes; however, the silk substrate stays transparent because of weak UV excitation of silk (Fig. 10c, d). Such a difference in the florescent features between the substrate and the printed silk pattern causes the great contrast between the two materials which significantly improves the image resolution for strain analysis as shown in Fig. 10d. Furthermore, two-dimensional discrete Fast Fourier Transform (2D-FFT) can be used to easily obtain the strain of the membrane in deformation. It is demonstrated that the material is more stretchable than most biological tissues and can accommodate strains of at least 250% in deformations that is more than two orders of magnitude larger than almost any commercially available strain gauges (Hammock et al. 2013).

9 Conclusions and Prospective

In this chapter, we have summarized the various unique properties that silk-based and silk-inspired materials embody and addressed the various structural features, dependent on hierarchical material organization and specific processing conditions, required to achieve these properties. Several case studies describe the wide range of applications of a combined modeling, synthesis, and characterization approach at different material scales. Starting from the single-molecule scale, silks can form a basis for developing chimera tunable materials (e.g., silk-elastin-like protein polymers) with phase transition capabilities for multifunctional biomimetic applications. Silks can be used in silk-silica composites to create biodegradable and biocompatible scaffolds as a modular protein engineered biomaterial for tissue repair and regeneration applications. Energy- and waste-efficient silk filtration membranes can be engineered to remove molecular-level contaminants. As an example, we describe multilayer tunable designs made of silk nanofibrils with hydroxyapatite domains. The final example provided here is a 3D-printed silk membrane for use as an optical strain gage. Through these examples, we hope to convey the broad applicability of silk-like systems and the relevance of combining modeling approaches together with experimental synthesis and characterization to design new systems and applications drawing from and building on fundamental properties of this unique material.

Moving forward, interesting challenges abound. The development of tunable silk-based materials has been realized for molecular systems and macroscale systems, as the case studies reviewed here attest. Recent work has made significant progress incorporating modelling and experiment for predictive material design and characterization. In the future, it will be important to consider advanced and hybrid multiscale modeling methods to overcome still-existing bottlenecks to accurate predictive modeling of silk-based systems at the mesoscale. A predictive computational model for dense, high-molecular weight tunable silk-based materials is still needed that can integrate molecularly driven responsive features into a larger system (e.g., a hydrogel). Further work on understanding the specific mechanisms of interfacial behavior is needed to design mineral-protein composites, expanding on the silk-silica materials described in this chapter. From the experimental point-of-view, recombinant protein design and synthesis is expensive and inefficient. Advancements in recombinant synthesis, and alternative biogenetics methods, including genetic engineering of silk worms to produce spider silk as one example, can greatly accelerate the rate of material production. 3D printing also has been a useful alternative biomanufacturing method; however, low resolution and optimization of processing methodology (e.g., solution viscosity, evaporation rate) can still be improved. Because 3D printing is a time-consuming process, development of novel self-assembly methods would help with manufacturing ease. Finally, we believe that novel optimization algorithms applied to silk-based systems can greatly expand the material design space, leading to new applications of silk-based systems.

References

Agnarsson I, Kunter M, Blackledge TA (2010) Bioprospecting finds the toughest biological material: extraordinary silk from a giant riverine orb spider. PLoS One 5(9):e11234

Altman GH et al (2003) Silk-based biomaterials. Biomaterials 24(3):401–416

Anumolu R et al (2011) Fabrication of highly uniform nanoparticles from recombinant silk-elastin-like protein polymers for therapeutic agent delivery. ACS Nano 5(7):5374–5382

Ayoub NA et al (2007) Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. PLoS One 2(6):e514

Benoit DS et al (2008) Small functional groups for controlled differentiation of hydrogel-encapsulated human mesenchymal stem cells. Nat Mater 7(10):816–823

Bhumiratana S et al (2011) Nucleation and growth of mineralized bone matrix on silk-hydroxyapatite composite scaffolds. Biomaterials 32(11):2812–2820

Buehler MJ, Yung YC (2009) Deformation and failure of protein materials in physiologically extreme conditions and disease. Nat Mater 8(3):175–188

Chan BP, Leong KW (2008) Scaffolding in tissue engineering: general approaches and tissue-specific considerations. Eur Spine J 17:S467–S479

Chen PY, McKittrick J, Meyers MA (2012) Biological materials: functional adaptations and bioinspired designs. Prog Mater Sci 57(8):1492–1704

Cranford SW, Tarakanova A, Pugno N, Buehler MJ (2011) Nonlinear material behaviour of spider silk yields robust webs. Nature 482:72–76

Cranford S et al (2012) Nonlinear material behaviour of spider silk yields robust webs. Nature 482:72–76

Deng C et al (2013) Ultrathin self-assembled anionic polymer membranes for superfast size-selective separation. Nanoscale 5(22):11028–11034

Dinjaski N, Ebrahimi D, Ling S, Shah S, Buehler MJ, Kaplan DL (2016) Integrated modeling and experimental approaches to control silica modification of design silk-based biomaterials. ACS Biomater Sci Eng 3:2877–2888

Dinjaski N et al (2018) Predicting rates of in vivo degradation of recombinant spider silk proteins. J Tissue Eng Regen Med 12(1):e97–e105

Du N et al (2006) Design of superior spider silk: from nanostructure to mechanical properties. Biophys J 91(12):4528–4535

Eguizabal A et al (2014) Nanoporous PBI membranes by track etching for high temperature PEMs. J Membr Sci 454:243–252

Engler AJ et al (2006) Matrix elasticity directs stem cell lineage specification. Cell 126(4):677–689

Giesa T et al (2011) Nanoconfinement of spider silk fibrils begets superior strength, extensibility, and toughness. Nano Lett 11(11):5038–5046

Giesa T, Perry CC, Buehler MJ (2016) Secondary structure transition and critical stress for a model of spider silk assembly. Biomacromolecules 17(2):427–436

Gosline JM, Denny MW, Demont ME (1984) Spider Silk as Rubber. Nature 309(5968):551–552

Gosline JM et al (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function. J Exp Biol 202:3295–3303

Guerette PA et al (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family. Science 272(5258):112–115

Gustafson J et al (2009) Silk-elastinlike recombinant polymers for gene therapy of head and neck cancer: from molecular definition to controlled gene expression. J Control Release 140(3):256–261

Haider M et al (2005) Molecular engineering of silk-elastinlike polymers for matrix-mediated gene delivery: biosynthesis and characterization. Mol Pharm 2(2):139–150

Hammock ML et al (2013) 25th anniversary article: the evolution of electronic skin (E-Skin): a brief history, design considerations, and recent progress. Adv Mater 25(42):5997–6037

Huang W, Rollett A, Kaplan DL (2015) Silk-elastin-like protein biomaterials for the controlled delivery of therapeutics. Expert Opin Drug Deliv 12(5):779–791

Huang W et al (2016) Design of Multi-Stimuli Responsive Hydrogels using integrated modeling and genetically engineered silk-elastin-like-proteins. Adv Funct Mater. https://doi.org/10.1002/adfm.201600236

Hwang D et al (2009b) Influence of polymer structure and biodegradation on DNA release from silk-elastinlike protein polymer hydrogels. Int J Pharm 368(1–2):215–219

Hwang W et al (2009a) Surface induced nanofiber growth by self-assembly of a silk-elastin-like protein polymer. Langmuir 25(21):12682–12686

Jacobsen MM et al (2017) Effect of terminal modification on the molecular assembly and mechanical properties of protein-based block copolymers. Macromol Biosci 17(9)

Joseph N et al (2014) Layer-by-layer preparation of polyelectrolyte multilayer membranes for separation. Polym Chem 5(6):1817–1831

Karan S et al (2012) Ultrafast viscous permeation of organic solvents through diamond-like carbon Nanosheets. Science 335(6067):444–447

Keten S et al (2010) Nanoconfinement controls stiffness, strength and mechanical toughness of beta sheet crystals in silk. Nat Mater 9(4):359–367

Krieg E et al (2011) A recyclable supramolecular membrane for size-selective separation of nanoparticles. Nat Nanotechnol 6(3):141–146

Launey ME, Buehler MJ, Ritchie RO (2010) On the mechanistic origins of toughness in bone. Annu Rev Mater Res 40:25–53

Liang HW et al (2010) Carbonaceous nanofiber membranes for selective filtration and separation of nanoparticles. Adv Mater 22(42):4691–4695

Lin SC et al (2015) Predictive modelling-based design and experiments for synthesis and spinning of bioinspired silk fibres. Nat Commun 6:6892. https://doi.org/10.1038/ncomms7892

Ling S et al (2016) Printing of stretchable silk membranes for strain measurements. Lab Chip 16(13):2459–2466

Ling S et al (2017) Design and function of biomimetic multilayer water purification membranes. Sci Adv 3(4), e1601939, https://doi.org/10.1126/sciadv.1601939

Liu L et al (2008) Preparation and characterization of nano-hydroxyapatite/silk fibroin porous scaffolds. J Biomat Sci Poly Edi 19(3):325–338

Liu Y, Shao ZZ, Vollrath F (2005) Relationships between supercontraction and mechanical properties of spider silk. Nat Mater 4(12):901–905

Marolt D et al (2006) Bone and cartilage tissue constructs grown using human bone marrow stromal cells, silk scaffolds and rotating bioreactors. Biomaterials 27(36):6138–6149

Martín-Moldes Z et al (2018) Intracellular pathways involved in bone regeneration triggered by recombinant silk–silica chimeras. Adv Funct Mater 28:1702570

Megeed Z et al (2004) In vitro and in vivo evaluation of recombinant silk-elastinlike hydrogels for cancer gene therapy. J Control Release 94(2–3):433–445

Meinel L et al (2004) Engineering bone-like tissue in vitro using human bone marrow stem cells and silk scaffolds. J Biomed Mater Res A 71a(1):25–34

Mieszawska AJ et al (2010a) Osteoinductive silk-silica composite biomaterials for bone regeneration. Biomaterials 31(34):8902–8910

Mieszawska AJ et al (2010b) Nanoscale control of silica particle formation via silk-silica fusion proteins for bone regeneration. Chem Mater 22(20):5780–5785

Moisenovich MM et al (2012) Tissue regeneration in vivo within recombinant spidroin 1 scaffolds. Biomaterials 33(15):3887–3898

Mourhatch R, Tsotsis TT, Sahimi M (2010) Network model for the evolution of the pore structure of silicon-carbide membranes during their fabrication. J Membr Sci 356(1–2):138–146

Ner Y et al (2009) Electrospinning nanoribbons of a bioengineered silk-elastin-like protein (SELP) from water. Polymer 50(24):5828–5836

Nova A et al (2010) Molecular and Nanostructural mechanisms of deformation, strength and toughness of spider silk fibrils. Nano Lett 10(7):2626–2634

Nudelman F et al (2010) The role of collagen in bone apatite formation in the presence of hydroxyapatite nucleation inhibitors. Nat Mater 9(12):1004–1009

Omenetto FG, Kaplan DL (2010) New opportunities for an ancient material. Science 329(5991):528–531

Peng XS, Jin J, Ichinose I (2007) Mesoporous separation membranes of polymer-coated copper hydroxide nanostrands. Adv Funct Mater 17(11):1849–1855

Plaza GR et al (2006) Thermo-hygro-mechanical behavior of spider dragline silk: glassy and rubbery states. J Poly Sci Part B Poly Phy 44(6):994–999

Polo-Corrales L, Latorre-Esteves M, Ramirez-Vick JE (2014) Scaffold design for bone regeneration. J Nanosci Nanotechnol 14(1):15–56

Qin Z et al (2015) Structural optimization of 3D-printed synthetic spider webs for high strength. Nat Commun 6:7038. https://doi.org/10.1038/ncomms8038

Qiu WG et al (2009) Wet-spinning of recombinant silk-elastin-like protein polymer fibers with high tensile strength and high deformability. Biomacromolecules 10(3):602–608

Qiu WG et al (2010) Complete recombinant silk-Elastinlike protein-based tissue scaffold. Biomacromolecules 11(12):3219–3227

Rammensee S et al (2008) Assembly mechanism of recombinant spider silk proteins. Proc Natl Acad Sci 105(18):6590–6595

Rim NG et al (2017) Predicting silk Fiber mechanical properties through multiscale simulation and protein design. ACS Biomater Sci Eng 3(8):1542–1556

Rising A et al (2010) Spider silk proteins: recent advances in recombinant production, structure–function relationships and biomedical applications. Cell Mol Life Sci 68(2):169–184

Rockwood DN et al (2011) Ingrowth of human mesenchymal stem cells into porous silk particle reinforced silk composite scaffolds: an in vitro study. Acta Biomater 7(1):144–151

Sengupta D, Heilshorn SC (2010) Protein-engineered biomaterials: highly tunable tissue engineering scaffolds. Tissue Engineering Part B-Reviews 16(3):285–293

Shannon MA et al (2008) Science and technology for water purification in the coming decades. Nature 452(7185):301–310

Shao ZZ, Vollrath F (1999) The effect of solvents on the contraction and mechanical properties of spider silk. Polymer 40(7):1799–1806

Sofia S et al (2001) Functionalized silk-based biomaterials for bone formation. J Biomed Mater Res 54(1):139–148

Striemer CC et al (2007) Charge- and size-based separation of macromolecules using ultrathin silicon membranes. Nature 445(7129):749–753

Sutherland TD et al (2010) Insect silk: one name, many materials. Annu Rev Entomol 55:171–188

Swanson BO, Blackledge TA, Hayashi CY (2007) Spider capture silk: performance implications of variation in an exceptional biomaterial. J Exp Zool A Ecol Genet Physiol 307A(11):654–666

Swanson BO et al (2009) The evolution of complex biomaterial performance: the case of spider silk. Integr Comp Biol 49(1):21–31

Tarakanova A, Buehler MJ (2012) A materiomics approach to spider silk–protein molecules to webs. JOM 64(2):214–225

Tarakanova A, Haung W, Qin Z, Kaplan DL, Wong JY, Buehler MJ (2017a) Modeling and experiment reveal structure and nanomechanics across the inverse temperature transition in silk-elastin-like protein polymers. ACS Biomater Sci Eng 3:2889–2899

Tarakanova A, Haung W, Qin Z, Kaplan DL, Buehler MJ (2017b) Modeling and experiment reveal structure and nanomechanics across the inverse temperature transition in *B. mori* silk-elastin-like protein polymers. ACS Biomater Sci Eng 3:2889–2899

Teixeira AI, Nealey PF, Murphy CJ (2004) Responses of human keratocytes to micro- and nanostructured substrates. J Biomed Mater Res A 71(3):369–376

Vandezande P, Gevers LEM, Vankelecom IFJ (2008) Solvent resistant nanofiltration: separating on a molecular level. Chem Soc Rev 37(2):365–405

Vepari C, Kaplan DL (2007) Silk as a biomaterial. Prog Polym Sci 32(8–9):991–1007

Vollrath F (2010) Spider silk: evolution and 400 million years of spinning, waiting, snagging, and mating. Nature 466(7304):319–319

Wang QF, Samitsu S, Ichinose I (2011) Ultrafiltration membranes composed of highly cross-linked cationic polymer gel: the network structure and superior separation performance. Adv Mater 23(17):2004–2008

Wang Q et al (2014) High throughput screening of dynamic silk-elastin-like protein biomaterials. Adv Funct Mater 24(27):4303–4310

Werber JR, Osuji CO, Elimelech M (2016) Materials for next-generation desalination and water purification membranes. Nat Rev Mater 1(5):16018

Xia XX et al (2011) Tunable self-assembly of genetically engineered silk–elastin-like protein polymers. Biomacromolecules 12(11):3844–3850

Xia XX et al (2014) Hydrophobic drug-triggered self-assembly of nanoparticles from silk-elastin-like protein polymers for drug delivery. Biomacromolecules 15(3):908–914

Yang Y et al (2005) Toughness of spider silk at high and low temperatures. Adv Mater 17(1):84–88

Zhang F et al (2015) Regeneration of high-quality silk fibroin fiber by wet spinning from CaCl2-formic acid solvent. Acta Biomater 12:139–145

Zhu JX et al (2016) Aqueous-based coaxial electrospinning of genetically engineered silk elastin core-shell nanofibers. Materials 9(4):221. https://doi.org/10.3390/ma9040221

Silica Aerogels: A Review of Molecular Dynamics Modelling and Characterization of the Structural, Thermal, and Mechanical Properties

64

Jingjie Yeo, Zishun Liu, and Teng Yong Ng

Contents

Abstract

Samuel S. Kistler showed that extremely porous silica aerogels can be made from supercritical drying of wet gels. This work, along with subsequent exper-

J. Yeo (✉)
Department of Biomedical Engineering, Tufts University, Medford, MA, USA

Laboratory for Atomistic and Molecular Mechanics (LAMM)/Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
e-mail: yeojj@mit.edu

Z. Liu
International Center for Applied Mechanics, State Key Laboratory for Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an, China
e-mail: zishunliu@mail.xjtu.edu.cn

T. Y. Ng
School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore, Singapore
e-mail: MTYNg@ntu.edu.sg

imental and simulation studies, eventually led to the commercialization of silica aerogels in numerous engineering applications, especially for thermal insulation in aerospace and civil engineering applications. Rapid progress in the synthesis of silica aerogels provided great impetus for characterizing and optimizing their molecular structures. This created significant numerical challenges in understanding their structure-property-function relationship at several hierarchies of length scales. Both fully atomistic and coarse-grained molecular dynamics modeling and simulations have been extensively developed to tackle these challenges. We reviewed the development of new empirical molecular dynamics forcefields, novel methods of generating aerogels' percolated backbones, and compelling algorithms for characterizing their structural, mechanical, and thermal properties that have resulted in unprecedented insights into silica aerogels. These developments will drive the eventual creation of a comprehensive set of multiscale modeling platforms, which can minimize the trials and errors during experimental synthesis and even bring silica aerogels into the fold of the materials-by-design paradigm.

1   Introduction

Over the past few decades, the unending pursuit of lightweight, robust, and cost-effective engineering materials have driven the development of diverse classes of ultralight materials. Aerogels are one such class of materials that is developed by introducing extreme porosity to drastically minimize their overall weight while, at the same time, maximizing their strength and toughness. S. S. Kistler and his colleagues successfully synthesized the initial form of such gels through supercritical drying of wet gels of silica, alumina, nickel tartarate, stannic oxide, tungstic oxide, gelatin, agar, nitrocellulose, cellulose, and egg albumin (Kistler 1931, 1932; Aegerter et al. 2011), thus demonstrating the incredible flexibility in the choice of the materials that can be made into aerogels. Furthermore, supercritical drying at high temperatures and pressures prevents the development of strong interfacial tensions between the phase boundary of liquid and gas. High surface tensions, generated by processes such as liquid evaporation in air, would cause the gel to collapse and shrink, thus forming much denser gels known as xerogels which have only 50% porosities (Fricke and Emmerling 1992).

For silica aerogels in particular, multistep processes involving polymerization, gelation, hydrolysis, and condensation are used to form the tortuous porous backbone of the aerogels. To synthesize silica aerogels, hydrolysis and condensation of water or alcohols drive the polymerization of linear oligomers of particulate colloidal silica through crosslinking, with subsequent continuous sol-gel transition. Initially, chemical precursors of the silica gels were sodium metasilicate (Na_2SiO_3) (Kistler 1932), but these are now largely replaced by silicon alkoxides ($Si(OR)_4$, where R and OR denote alkyl and alkoxide groups, respectively. Some examples are tetramethyl orthosilicate (TMOS) and tetraethyl orthosilicate (TEOS), where the R groups are CH_3 and C_2H_5, respectively. Polycondensation of these monomers

results in the formation of coherent porous gels which are then dried supercritically to obtain the final aerogel structures. Numerous alkoxide-related precursors are used nowadays, including 3-(2-aminoethylamino) propyltrimethoxysilane (EDAS), polyethoxysilane (PEDS), and perfluoroalkysilane (PFAS) (Aegerter et al. 2011). Such chemical diversity is required to produce different desired structural, mechanical, chemical, electrical, and optical properties in the resulting gel, in combination with variations in solvents, catalysts, and ratios of water to precursors (Aegerter et al. 2011). Aerogels with an even wider range of properties have been developed by using materials other than silica, such as carbon, sodium silicate, zirconium dioxide, as well as many other organic and inorganic compounds. An exhaustive overview of all these variations in aerogels is beyond the scope of this review, but further details can be obtained from the Aerogel Handbook (Aegerter et al. 2011).

Silica aerogels are highly suitable for an exceptionally wide range of engineering applications due to their unique properties. One of the most notable applications of aerogels is in Cherenkov radiators as Cherenkov counters (Cantin et al. 1974; Kharzheev 2008) for identification of particles. They are also used in dielectrics for integrated circuits (ICs), passive solar energy collection devices (Hrubesh 1998), and for the trapping, storing, and releasing of an extensive range of nanoparticles and molecules (Burchell et al. 2006; Tabata et al. 2014). Another prominent area of application is in thermal insulation. Commercial production of modern porous aerogel on a large, cost effective manner is driving the proliferation in the usage of silica aerogels as thermal insulators (Herrmann et al. 1995; Jelle et al. 2011; Cotana et al. 2014). For greater flexibility in tailoring the optical and thermal properties for the required application when used as thermal fillers, silica aerogels are produced as granular beads (Ackerman et al. 2001; Neugebauer et al. 2014). Some of these applications include passive solar windows (Wittwer 1989), cryogenic thermal insulation systems on space launch vehicles (Fesmire 2006; Jones 2006), and thermal insulation of windows (Rubin and Lampert 1983; Duer and Svendsen 1998). The theoretical reduction in thermal conductance of a window using just 5 mm thick aerogel as insulation is found to be more than half (Rubin and Lampert 1983), which is significantly more insulating than even double-glass windows. Experimentally, aerogel glazing units outperformed conventional double glazing as they reduce heat loss by 58% (Gao et al. 2014). Furthermore, other than capturing space dust in NASA's Stardust mission, silica aerogels have also kept two recent Mars Rovers, Spirit and Opportunity, at relatively steady temperatures even though the temperature variation on the surface of Mars can be above 100 °C (Jones 2006).

2   Properties of Silica Aerogels

Due to the extremely porous nature of silica aerogels, their densities can approach values as low as 0.003 g/cm^3 (Tillotson and Hrubesh 1992). This remarkably low density is only three times higher than the density of air which is approximately 0.0012 g/cm^3. The nanoscale architecture of silica aerogels is usually characterized

as having fractal geometries due to self-similar patterns that appear geometrically similar even when viewed at different hierarchical length scales. The degree of fractality of a structure is usually represented by its fractal dimension which is a ratio that relates the change in detail to a given change in scale. Small-angle neutron scattering (SANS) and small-angle x-ray scattering (SAXS) spectroscopy and adsorption of molecules with different cross-sectional areas are some experimental methods for determining the fractal dimension of silica aerogels (Woignier et al. 1990). Silica aerogel's fractal dimensions vary with processing conditions, where it is approximately 1.8 under basic processing conditions (Vacher et al. 1988b) and about 2.2–2.4 in both acidic and neutral conditions (Vacher et al. 1988a).

Pure silica aerogels tend to be mechanically weak due to their low density and high porosity, but there is still a broad range of mechanical properties that can be obtained by varying their processing conditions and thereby controlling crucial factors like the porosity, density, defects, and network connectivity. The mechanical properties of silica aerogels are measured through a variety of methods. By measuring the velocity of sound through the gel, a density power-law scaling exponent of 3.17 ± 0.11 and 3.25 ± 0.15 is found for the elastic and shear modulus, respectively, in aerogels with densities between 0.05 and 2 g/cm^3 (Cross et al. 1989). A higher value of 3.7 is determined by Gronauer et al. (1986) using the same experimental method. The difference in the Young's moduli obtained using either the propagation of sound or uniaxial compression tests is also compared (Gross et al. 1988), and there is good agreement only at a medium density of 0.158 g/cm^3. Further uniaxial compression tests at ambient conditions found that the elastic moduli scaled with density differently depending on the processing conditions, with scaling constants of 3.37 ± 0.13 for acid catalyzed silica aerogels, 3.77 ± 0.21 for prehydrolyzed aerogels, and 2.85 ± 0.17 for base catalyzed aerogels (Lemay et al. 1990). This is for a density range between 0.1 and 0.6 g/cm^3. Three-point beam-bending and hydrostatic compression tests are also compared and it is found that reasonable agreement can be obtained with these two methods in the elastic regime (Scherer et al. 1995). Beam-bending tests are also performed to determine the variation of elastic moduli with density in the range of 0.1–0.4 g/cm^3 and a scaling exponent of 3.8 is determined (Woignier et al. 1987). Further bending tests revealed a scaling exponent of 3.7 ± 0.2 for the density range of 0.05–0.4 g/cm^3 (Woignier and Phalippou 1989). Flexural and uniaxial compressions tests are also compared and the results differed strongly as different precautions have to be taken to ensure each method's accuracy (Alaoui et al. 2008). Both the elastic and brittle mechanical behaviors of silica aerogels are attributable to the same kinds of gel characteristics of pore volume, silanol content, and pore size (Woignier et al. 2015).

The highly porous nature of pure silica aerogels is deleterious for their mechanical properties, but this porosity is remarkably beneficial for conferring unusually low thermal conductivity, especially for purposes of thermal insulation. Figure 1 shows the incredible insulating properties of a slab of silica aerogel preventing intense heat from affecting a delicate flower. Silica aerogels' thermal transport mechanisms have been characterized experimentally (Lu et al. 1995; Heinemann et al. 1996; Zeng et al. 1996; Wei et al. 2011), and it is found that separate

Fig. 1 To demonstrate the extremely low thermal conductivity of silica aerogel, a burner is lit under the translucent blue aerogel and the flower shows no detrimental signs of excessive heat, by J. Cormac 2009, https://commons.wikimedia.org/wiki/File:Aerogelflower_filtered.jpg

components of gaseous, solid, and radiative thermal conductivity constitute the total thermal conductivity of monolithic silica aerogels. Numerous factors decrease each component of the thermal conductivity, including negligible coupling between the different modes of heat transfer, absence of gaseous convection (and therefore diminished heat transfer) within the pores, and effective restriction of local excitations within the aerogel backbone due to the high number of pores (Ebert 2011; Koebel et al. 2011). In particular, a power-law variation of the solid thermal conductivity with density is found, similar to the variation of the elastic modulus with density, such that

$$\lambda_S = C\rho^{\alpha} \tag{1}$$

where the factor C varies with the particle connectivity within the aerogel, and α is approximately 1.6, for densities, ρ, in the range of 0.3–1 g/cm^3 (Fricke 1988), and 1.5, for ρ between 0.07 and 0.3 g/cm^3 (Fricke et al. 1992).

3   Characterization of Silica Aerogels with Molecular Dynamics

Due to the intricate link between silica aerogel's molecular structure and its properties, molecular dynamics (MD) modelling and simulations have played crucial roles in advancing the fundamental understanding of silica aerogel's structure-property-function relationship. This is reliant on significant improvements in two

basic factors: the accuracy of the MD forcefields and the algorithms for constructing representative aerogel structures. These advances have driven the ability of MD models to comprehensively capture aerogel's structural, mechanical, and thermal properties. Extensive details on the theory, methods, and implementations of MD models and simulations can be found in many textbooks (Allen and Tildesley 1989; Haile et al. 1993; Frenkel and Smit 2002; Rapaport 2004); hence, these aspects will not be covered here.

3.1   Molecular Dynamics Forcefields

MD simulations are underpinned by the quality of the mathematical formulations (the equations of motion, otherwise known as forcefields) that can meticulously reproduce a subset of properties needed for examining a certain phenomenon. Several forcefields have been proposed for modeling silica, however, only a small number of these have been implemented and validated for modeling various aspects of silica aerogels. One of the primary interaction potential used to model both amorphous bulk silica (Horbach and Kob 1999; Horbach et al. 1999; Jund and Jullien 1999; Shell et al. 2002; McGaughey and Kaviany 2004; Mahajan et al. 2007; Coquil et al. 2011) and silica aerogels (Ng et al. 2012, 2016; Yeo et al. 2013; Yeo 2014) is introduced by van Beest, Kramer, and van Santen (BKS) (van Beest et al. 1990; Kramer et al. 1991). The BKS potential's mathematical formulation is

$$U\left(r_{ij}\right) = A_{ij} \exp\left(-B_{ij} r_{ij}\right) - \frac{c_{ij}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \tag{2}$$

The first and second terms in Eq. 2 are the short-range attraction and repulsion of atoms, respectively; the final term is the long-range Coulombic interaction. r_{ij} is the separation between atoms of indices i and j; q_i and q_j their atomic charges; and A_{ij}, B_{ij}, and C_{ij} the different constants for different pairs of atoms i and j. These parameters are fitted simultaneously to both experimental macroscopic data and results from microscopic ab initio, including data for α-quartz (van Beest et al. 1990). Surprisingly, the comparatively simpler mathematical formulation of the BKS potential is capable of predicting the properties of various silica polymorphs more accurately than other pair potentials such as the Tsuneyuki, Tsukada, Aoki, and Matsui (TTAM) potential (Tsuneyuki et al. 1988). The BKS potential is able to accurately depict the structural and mechanical properties of various silica polymorphs, such as the unit-cell dimensions and elastic moduli, respectively. A downside of the BKS potential for simulating aerogels is that at high quenching temperatures, atoms can approach each other too closely and lead to dynamics that are unmanageable. However, this can be overcome by steeply increasing the repulsive capabilities of the BKS potential at small atomic distances through the addition of a "24–6" Lennard-Jones potential (Guissani and Guillot 1996) to obtain

$$U\left(r_{ij}\right) = A_{ij}\exp\left(-B_{ij}r_{ij}\right) - \frac{c_{ij}}{r_{ij}^{6}} + \frac{q_{i}q_{j}}{r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{24} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (3)$$

Several studies have also implemented the Vashishta potential (Vashishta et al. 1990) to model silica aerogels (Nakano et al. 1993, 1994; Murillo et al. 2010; Patil et al. 2017). This potential introduces more sophistication by having terms that account for both two- and three-body interactions

$$U = U_{2}\left(r_{ij}\right) + U_{3}\left(r_{ij}, r_{jk}, r_{ik}\right) \quad (4)$$

where the two-body interaction term is

$$U_{2}\left(r_{ij}\right) = \frac{H_{ij}}{r_{ij}^{\eta_{ij}}} + \frac{Z_{i}Z_{j}}{r_{ij}} - \frac{\frac{1}{2}\left(\alpha_{i}Z_{j}^{2} + \alpha_{j}Z_{i}^{2}\right)}{r_{ij}^{4}}e^{-\frac{r_{ij}}{r_{4s}}} \quad (5)$$

and H_{ij} and η_{ij} are the strengths and exponents of the steric repulsion; Z_{i} and α_{i} the effective charge and electronic polarizability; r_{4s} is the decay length of the dipole interactions. The three-body interaction term is expressed as

$$U_{3}\left(r_{ij}, r_{jk}, r_{ik}\right) = B_{ijk}f\left(r_{ij}, r_{ik}\right)p\left(\theta_{jik}, \overline{\theta}_{jik}\right) \quad (6)$$

Here, B_{ijk} is the strength of the interactions. The effects of bond-stretching are represented by the function

$$f\left(r_{ij}, r_{ik}\right) = \begin{cases} \exp\left[\frac{l}{r_{ij}-r_0} + \frac{l}{r_{ik}-r_0}\right] & for\ r_{ij}, r_{ik} < r_0 \\ 0 & for\ r_{ij}, r_{ik} > r_0 \end{cases} \quad (7)$$

where r_0 is the interaction cutoff distance. The effects of bond-bending are given by the function

$$p\left(\theta_{jik}, \overline{\theta}_{jik}\right) = \left(\cos\theta_{jik} - \cos\overline{\theta}_{jik}\right)^{2} \quad (8)$$

where θ_{jik} is the angle subtended by atoms i, j, and k with atom i as the vertex.

Increasing complexity of inter-atomic interactions is needed for more rigorous simulations of the chemical environment that each atom encounters, thus requiring terms that account for multibody interactions. While pair approaches are relatively inexpensive computationally, they have a limited ability to model structurally and chemically complex materials, especially where elements like silicon and carbon have multiple possible polymorphs. These issues are addressed by varying the strength of each bond such that it is dependent on the conditions within the local environment through a bond-order function. The Tersoff potential (Tersoff 1988) is an example of a MD potential developed for this purpose and it has also been parameterized for silica (Munetoh et al. 2007), showing excellent utility in

predicting the thermal and mechanical properties of silica aerogels (Lei et al. 2013; Yeo et al. 2013; Ng et al. 2016). This potential has the general mathematical form of

$$U = f^c\left(r_{ij}\right)\left[V^R\left(r_{ij}\right) + b_{ij}\,V^A\left(r_{ij}\right)\right] \tag{9}$$

where V^R and V^A refer to the repulsive and attractive terms; f^c is a function that smoothly limits short-range interactions; b_{ij} is the bond-order that is monotonically decreasing according to the coordination of atoms i and j. Separately, the repulsive and attractive terms are

$$V^R\left(r_{ij}\right) = A_{ij}\exp\left(-\gamma_{ij}r_{ij}\right) \tag{10}$$

$$V^A\left(r_{ij}\right) = B_{ij}\exp\left(-\mu_{ij}r_{ij}\right) \tag{11}$$

$$A_{ij} = \left(A_i + A_j\right)^{\frac{1}{2}},\; B_{ij} = \left(B_i + B_j\right)^{\frac{1}{2}},\; \gamma_{ij} = \frac{\gamma_i + \gamma_j}{2},\; \mu_{ij} = \frac{\mu_i + \mu_j}{2} \tag{12}$$

The cutoff function is given as

$$f^c\left(r_{ij}\right) = \begin{cases} 1, & r_{ij} < R_{ij} \\ \frac{1}{2} + \frac{1}{2}\cos\left(\pi\,\frac{r_{ij}-R_{ij}}{S_{ij}-R_{ij}}\right) & R_{ij} < r_{ij} < S_{ij} \\ 0, & r_{ij} > S_{ij} \end{cases} \tag{13}$$

$$R_{ij} = \left(R_i R_j\right)^{\frac{1}{2}},\; S_{ij} = \left(S_i S_j\right)^{\frac{1}{2}} \tag{14}$$

where R and S denote the interaction limits. Finally, the bond-order term is

$$b_{ij} = \chi_{ij}\left(1 + \beta_i^{n_i}\,\zeta_{ij}^{n_i}\right)^{-\frac{1}{2n_i}} \tag{15}$$

$$\zeta_{ij} = \sum_{k \neq i,j} f^c\left(r_{ik}\right)\omega_{ik}g\left(\theta_{ijk}\right) \tag{16}$$

$$g\left(\theta_{ijk}\right) = 1 + \frac{c_i^2}{d_i^2} - \frac{c_i^2}{d_i^2 + \left(h_i - \cos\theta_{ijk}\right)^2} \tag{17}$$

Here, χ_{ij} dictates the strengthening or weakening of the heteropolar bonds and the remaining terms are parameters that can be fitted to experimental or ab initio data corresponding to the system under investigation.

3.2   Structural Properties

Using the MD interaction potentials in the previous section, the structural properties of silica aerogels can be represented using multiple schemes for very varied purposes such as investigating the atomic coordination, mechanical and thermal properties, chemical and gelation kinetics, and structural evolution and relaxation of aerogels. Generally, there are two broad categories of methods to obtain porous silica. The first category is "mimetic" simulations where experimental preparation of the aerogel is mimicked through dynamic representations. The second category is the construction of statistically accurate models to represent as much detail from available experimental data as computationally possible. For both of these categories, the aerogel models are usually characterized by structural and geometrical properties such as the pore size distributions, fractal dimensions, surface areas, and bond length and bond angle distributions (Gelb 2011).

Kieffer and Angell developed one of the earliest and most elementary atomistic methods to model the percolated structures of silica aerogels through MD (Kieffer and Angell 1988). In the scheme termed "negative pressure rupturing," an equilibrated sample of amorphous silica is instantaneously expanded in a stepwise manner until the percolation threshold is reached, as illustrated in Fig. 2a. All existing bonds are stretched simultaneously at each expansion step to lengths that are 10–35% beyond their equilibrium, and subsequently the structures are relaxed to remove excessive accumulated stresses. A similar method is used to simulate much larger systems of more than 40,000 atoms (Nakano et al. 1993, 1994) together with the Vashishta potential (see Sect. 3.1), as shown in Fig. 2b, c. This combination of factors generated porous structures with fractal dimensions that varied with density and good agreement with experimental data is obtained for the neutron-scattering static structure factor, bond statistics, and pore size distribution. Another study (Ng et al. 2012) also employed negative pressure rupturing but with the BKS potential instead (see Sect. 3.1), and this scheme generated silica aerogels whose fractal dimensions scaled with density in a manner similar to previous MD studies (Kieffer and Angell 1988).

By correlating with experimental data, three separate and distinct methods of modelling silica aerogels are compared for in-depth understanding of aerogel synthesis and the underlying pore structure (Pohl et al. 1995). The first method involves packing spherical silica particles of varying diameters into a cell to determine their surface areas. The second method proceeded in similar manner to negative pressure rupturing. Cluster-cluster aggregation of silica clusters is the final method and it produces the structure of aerogels upon gelation. The first method agrees reasonably well with theoretical curves as the porosity increases. For the second method, it is found that differing pore size distributions is obtained if the expansion rates are varied. In terms of the composition and the characteristic size, the final method is the most representative of experimental aerogels.

A recent MD scheme that modifies the approach of negative pressure rupturing is used to reconstruct the silica aerogel's percolated backbone together with the
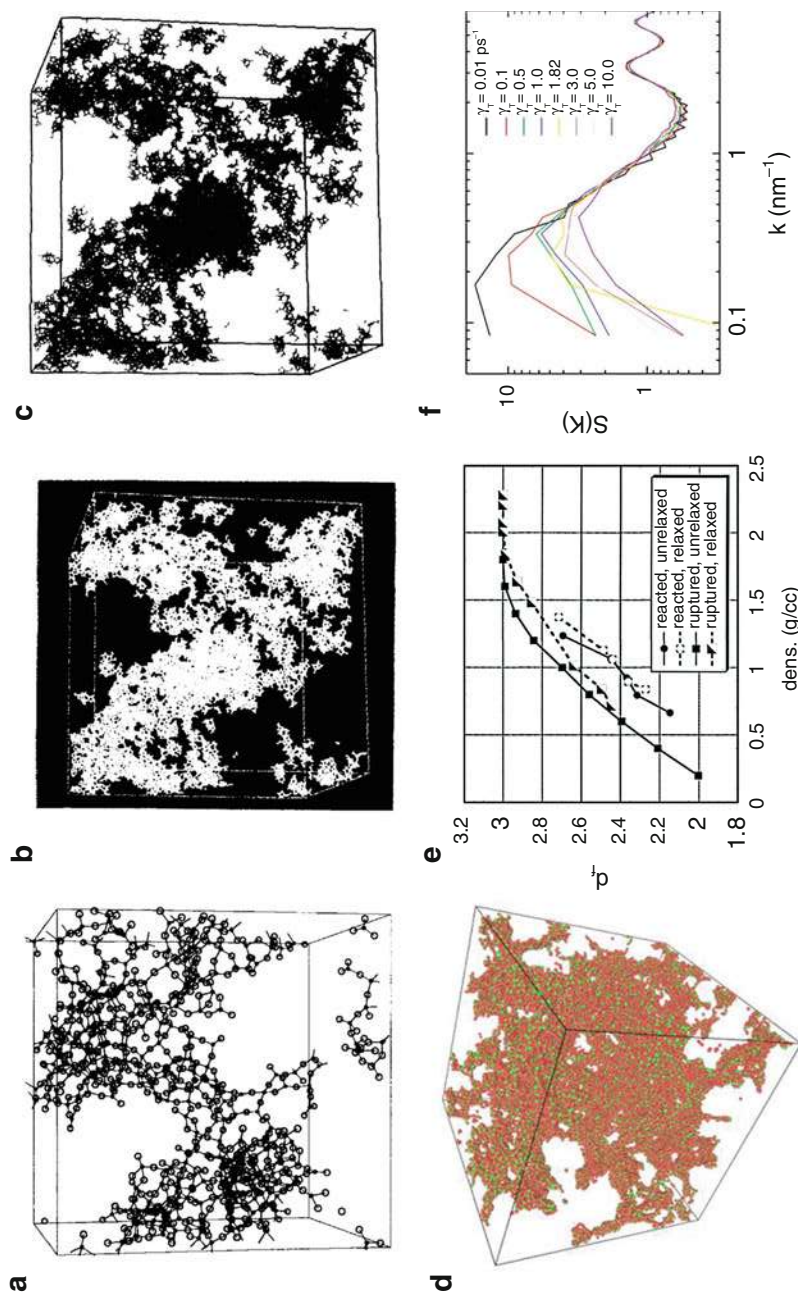
Fig. 2 Various porous backbone of silica aerogels generated through (**a–c**) negative pressure rupturing. (Reproduced from Kieffer and Angell 1988; Nakano et al. 1993, 1994 with permission from Elsevier and the American Physical Society) and (**d**) the expanding, heating, and quenching scheme. (Reproduced from Yeo et al. 2013 with permission from IOP Publishing Ltd). Numerous structural information of percolated silica aerogels has also been investigated, such as (**e**) the variation of fractal dimensions with density. (Reproduced from Bhattacharya and Kieffer 2005 with permission from AIP Publishing) and (**f**) the structure factors. (Reproduced from Gelb 2007 with permission from the American Chemical Society)

Vashishta potential (Murillo et al. 2010). In this scheme, which we dubbed as the expanding, heating, and quenching scheme (Yeo et al. 2013; Yeo 2014), a crystalline lattice of β-cristobalite with 24,000 atoms is expanded to the density required. Thereafter, the entire system is allowed to diffuse and aggregate by increasing the temperature up to 3000 K with subsequent relaxation by cooling down to temperatures that are between 2500 K and 0 K. Good agreement with experimental data is found for the *structural* features, such as the fractal dimensions, characterized using pair distribution functions and simulated scattering experiments. The expanding, heating, and quenching scheme is also used in combination with the Tersoff potential (see Sect. 3.1) (Lei et al. 2013; Yeo et al. 2013), as depicted in Fig. 2d, and the scaling of the fractal dimensions with density agreed very well with previous MD studies (Nakano et al. 1994; Murillo et al. 2010).

Through reactive MD simulations, the process of oligomerization and gelation during the production of silica aerogels can also be modelled in greater detail. A series of studies (Garofalini and Melman 1986; Feuston and Garofalini 1990; Garofalini and Martin 1994) simulated solutions of water, silicic acid monomers, and silicic acid dimers with the MD potentials derived from the Born-Mayer-Huggins (Garofalini 1983) and Rahman-Stillinger-Lemberg (Stillinger and Rahman 1978) potentials. The initial study validated the forcefields used through comparisons with data from molecular orbital calculations and experimental studies of bulk silicates. Pentacoordinated silicon is found during the reaction of two silicic acid monomers in vacuo (Garofalini and Melman 1986). Further revision of the models (Feuston and Garofalini 1990) led to the discovery of more intermediates during polymerization, including ionized monomers and pentacoordinated silicon. The simulated structures also compared favorably with X-ray and neutron diffraction data as well as quantum chemical calculations. The next study then simulated significantly larger systems with 216 silicic acid monomers at densities between 1.4 and 1.6 g/cm^3 (Garofalini and Martin 1994). The simulation data are consistent with experimental sol-gel systems in terms of chain formation, the activation energy for branching, and the relative time evolution.

Further advances in observing the gelation processes are made by incorporating the transfer of charges in a more realistic all-atom reactive approach (Bhattacharya and Kieffer 2005, Kieffer and Bhattacharya 2008). Comparison of the fractal dimensions, d_f, to those obtained from negative pressure rupturing revealed that the d_f is process-dependent rather than being associated with the density and porosity of the gel structure (Bhattacharya and Kieffer 2005), as shown in Fig. 2e. It is also determined that this fractal nature is not significantly affected by the supercritical drying process. Furthermore, three different growth regimes are observed in a subsequent study of silica aerogels' polycondensation process in both dry and aqueous environments (Kieffer and Bhattacharya 2008). The regimes of compact particles, percolated network structures, or branched clusters depend on the ratio between silicon and water as well as the density of the system. The growth kinetics and mechanical stability of the individual regimes are also studied.

There are inherent limitations in spatial sizes of MD simulations due to the computational intensity of fully atomistic simulations that extends to more than

millions of atoms, although there have been attempts to achieve this recently (Gonçalves et al. 2016). Several studies aim to circumvent this limitation by developing coarse-grained (CG) models of silica aerogels. These models aggregate numerous atoms into a single particle to reduce the number of computations needed while creating the ability to probe phenomena that exist at higher spatiotemporal scales that are out of the reach of fully atomistic MD simulations. This ability comes at the expense of losing atomistic detail; therefore, a good balance must be achieved to ensure that the observed phenomenon is not a computational artifact. One such CG model is developed using data from simulations of silica sol particles that are 2 nm in size to examine the gelation and relaxation mechanisms at much larger spatiotemporal scales (Gelb 2007), such as the structure factors in Fig. 2f. This is known as the flexible reactive aerogel model. A range of gelation behavior is observed by tuning the model's effective forward rate constant for bond creation. For highly relaxed clusters in low-reactivity conditions, reactions correspondingly became infrequent. The properties of the resulting gel are invariant even if reactivity is decreased further. In contrast, higher reactivity will result in increased amounts of structures that are far from equilibrium, leading to significant shrinkage at low pressures due to the presence of high stresses. We anticipate that future developments in CG models of silica aerogels will eventually lead to accurate multiscale models that link molecular structures to macroscale phenomena such as heat transfer and fracture of silica aerogel beads.

3.3   Mechanical Properties

Studies that utilized MD simulations to characterize the nanomechanical properties of silica aerogels have been a fairly recent development compared to the characterization of silica aerogel's structural properties. After the extensive validation of various MD methods of reproducing the fractal nature of silica aerogels, the Vashishta potential (see Sect. 3.1) and an algorithm to mimic diffusion-limited aggregation (DLCA) are used to generate porous silica with densities between 1.67 and 2.2 g/cm^3 (Campbell et al. 1999). This density range resembles xerogels more closely rather than extremely percolated aerogels. The variation of the bulk, shear, and Young's moduli with increasing density is examined. The Young's moduli are found to scale with density via a power-law with an exponent of 3.5 ± 0.2. Another manner of DLCA is adopted in another study (Murillo et al. 2010), which is the expanding, heating, and quenching scheme detailed in Sect. 3.2, together with the Vashishta potential. Percolated structures with densities between 0.23 and 2.2 g/cm^3 are studied, where the density range corresponds to that of aerogels and xerogels. Power-law variations are found for the Young's moduli and strength, where the exponents are 3.11 ± 0.21 and 2.53 ± 0.15, respectively, in good agreement with experimental data (Woignier et al. 1998). This same scheme is applied in another study (Patil et al. 2017) but on a much larger scale, having more than 192,000 atoms and system sizes that are larger than 17^3 nm^3. A density range of 0.28–2.2 g/cm^3 is investigated. The power-law exponent of the Young's moduli is estimated to be
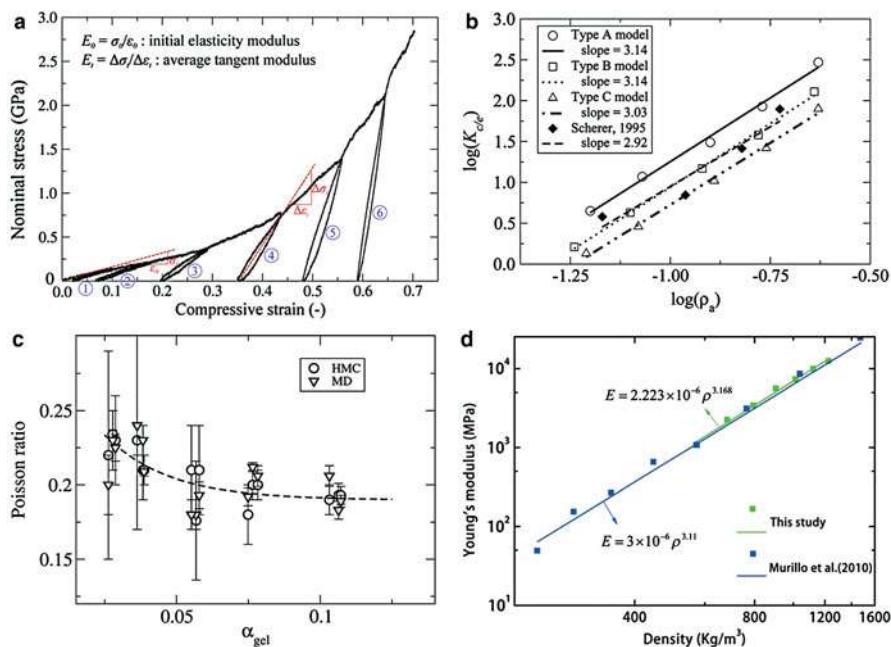
Fig. 3 Mechanical properties of silica aerogels that can be obtained from numerical simulations, such as (**a**) the relationship between stress and strain during cyclic loading. (Reproduced from Patil et al. 2017 with permission from the American Chemical Society), (**b**) the bulk moduli. (Reproduced from Ferreiro-Rangel and Gelb 2013 with permission from the American Chemical Society), (**c**) the Poisson ratio. (Reproduced from Ferreiro-Rangel and Gelb 2015 with permission from American Chemical Society), and (**d**) the elastic moduli. (Reproduced from Lei et al. 2013 with permission from The Royal Society of Chemistry)

3.25 ± 0.1 and brittle fracture is observed in tensile tests while early densification occurred for compression tests. Poisson ratios are between 0.18 and 0.21. Cyclic compressive loading at three densities of 0.28, 0.42, and 0.62 g/cm^3 caused permanent deformations under large strains due to inelastic phenomena related to residual deformation, hysteresis, and the Mullins effect (Fig. 3a). Mathematical relations are also determined between the compressive strain and the residual strain, the dissipated energy, and the average tangent modulus.

CG modelling (see Sect. 3.2) of silica aerogels has also been used to characterize the mechanical properties of silica aerogels at higher hierarchical length scales. In a series of studies using the flexible reactive aerogel model, it is found initially that the bulk modulus is significantly lower (1.0–2.3 bar) than experimental data (~7.0 bar) at similar densities (Gelb 2007). A subsequent study (Ferreiro-Rangel and Gelb 2013) looked at this issue in more detail by considering variations in particle sizes, densities, and size dispersity, along with two separate methods for obtaining the bulk modulus: fluctuation analysis or direct compression and expansion simulations. The bulk moduli derived from direct compression and expansion simulations appear
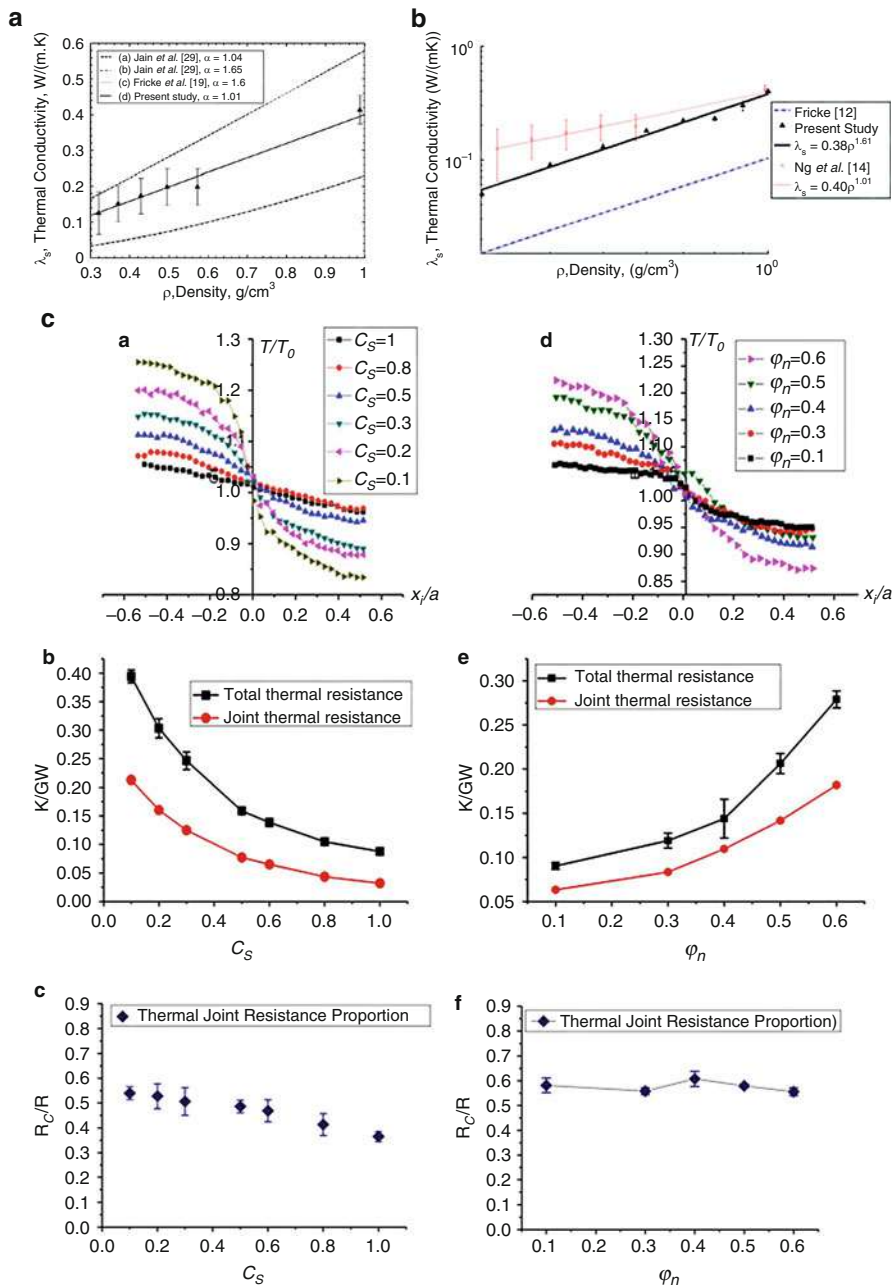
Fig. 4 (continued)

to be more reliable than those from fluctuation analysis which requires careful parameterization of the barostat mass. The power-law scaling of the bulk moduli with density has an exponent of approximately 3.1 as shown in Fig. 3b, but polydisperse particle sizes cause the bulk modulus to decrease due to the reduced density of bonds. In the most recent study (Ferreiro-Rangel and Gelb 2015), MD and hybrid Monte Carlo methods are used to investigate the uniaxial deformation and mechanical properties of silica aerogels as depicted in Fig. 3c. The power-law dependence of the Young's moduli has an exponent of approximately 3.0. Poisson ratios lie between 0.17 and 0.23. The Young's modulus and Poisson ratio are found to be largely isotropic, but anisotropy became more prominent as density decreased. Large deformation uniaxial tensile and compressive tests are also performed, where high density aerogels exhibited brittle failure, whereas low density aerogels are generally more elastic.

Extrapolation of MD simulations to bulk elastic properties has also been investigated. For instance, analytical-numerical models can be constructed with molecular details obtained from MD simulations (Lei et al. 2013). MD simulations with the Tersoff potential (see Sect. 3.1) and the expanding, heating, and quenching scheme are used to derive the Young's modulus at the molecular scale, as illustrated in Fig. 3d. These data are used to parameterize an analytical model of a much larger nanoparticle, leading to a fast and convenient analytical method of depicting the silica aerogel's Young's modulus.

3.4   Thermal Properties

MD simulations have also been used for the extensive characterization of nanoscale thermal transport in the solid backbone of silica aerogels. The vibrational density of states (vDOS) are determined for CG percolated silica structures generated by placing individual fully atomistic samples of vitreous silica in an infinite 3D percolated cluster (Rahmani et al. 1994). Through the development of a linear-frequency moments method, the vDOS of systems that are 50 nm × 35 nm × 35 nm in size and have densities between 0.10 and 0.26 g/cm^3 can be rapidly determined, especially in the low-frequency region, and the data are in good agreement with experimental results. A follow-up study (Benoit et al. 2001) generated a CG model that can reproduce small-angle neutron scattering (SANS) experimental data for the

Fig. 4 The thermal properties of silica aerogels investigated with various numerical models. (**a**), (**b**) The thermal conductivities are compared with experimental studies, where the power-law variation with density, determined through the Tersoff potential with the expanding, heating, and quenching scheme, showed excellent agreement. (Reproduced from Ng et al. 2012 and Yeo et al. 2013 with permission from Elsevier and IOP Publishing Ltd). (**c**) Other properties that are characterized include the variations in thermal resistance. (Reproduced from Liu et al. 2014 with the permission of AIP Publishing)

vDOS and inelastic scattering intensity of silica aerogels. This model is derived from a DLCA algorithm that incorporates primary particles of fully atomistic samples of vitreous silica modelled using the BKS potential (see Sect. 3.1).

The thermal conductivity of silica aerogels has also been studied in a series of MD simulations that are primarily performed using a class of MD methods known as nonequilibrium MD (NEMD) (Rapaport 2004). In each study, the MD forcefields used are validated extensively due to the paucity of simulation and experimental data available for thermal transport on such a small length scale. In the initial study (Ng et al. 2012), the thermal conductivity of dense vitreous silica is validated using the BKS potential. Vitreous silica is generated by quenching with the BKS potential while the thermal conductivity is determined through the method of reverse NEMD (Muller-Plathe and Bordat 2004). In a system with 648 atoms, the thermal conductivity is 1.20 ± 0.07 W/(m·K), in good agreement with MD results published previously (Jund and Jullien 1999; Coquil et al. 2011), while a thermal conductivity of a larger system with 3000 atoms is determined to be 1.39 W/(m·K) at 300 K, close to the experimental value of 1.40 W/(m·K). This variation in thermal conductivity is a phenomenon that is well known in MD simulations due to finite size effects (Yeo et al. 2013). Negative pressure rupturing is then applied on the larger system, producing percolated aerogels with densities between 0.32 and 0.99 g/cm^3. The thermal conductivity is determined subsequently using RNEMD and the power-law dependence with density has an exponent of 1.01, in contrast with experimental values of 1.6 (Fricke 1988). Moreover, the thermal conductivity at the density of 0.3 g/cm^3 is higher than experimental values by almost an order of magnitude, as shown in Fig. 4a. This is partly explained by the influence of processing conditions as well as the lack of defects, impurities, and sufficiently broad pore size distributions. It is verified that the largest pore diameter at a density of 0.3 g/cm^3 is significantly smaller than the pores found in experimental aerogels.

In order to minimize the effects of finite sizes and to investigate the role of the MD forcefields in determining the thermal conductivity, a subsequent study developed an enhanced model (Yeo et al. 2013) that is a combination of the expanding, heating, and quenching scheme (see Sect. 3.2) with larger system sizes of more than 52,000 atoms and the re-parameterized Tersoff potential (see Sect. 3.1). For vitreous silica, it is shown that the Tersoff potential predicts bulk thermal conductivity values and vDOS that are significantly closer to experimental data. This scheme dramatically improved the power-law scaling of thermal conductivity with density in the aerogel models, where the exponent is now 1.61, while having a much lower thermal conductivity of 0.05 ± 0.003 W/(m·K) at the lowest density of 0.3 g/cm^3 (Fig. 4b). This model is extended further to investigate the thermal resistance between secondary spherical particles of silica aerogels (Liu et al. 2014), with the exception that the percolated primary structure in each secondary particle is generated by deleting atoms prior to heating and quenching. The thermal resistance between two particles that are in close contact is in the range of 0.1–0.4 K/GW, depending on the complex interplay between contact length and defect concentration, as illustrated in Fig. 4c. Analysis of the vDOS shows that contact length fractions that are lower than 50% significantly suppress the longitudinal

acoustic (LA) modes of vibration. A similar phenomenon is found for defect concentrations higher than 50%.

4 Conclusions

The intention of this review is to provide a sufficiently broad yet in-depth understanding of historical developments in the fully atomistic and coarse-grained molecular dynamics modelling and simulations to characterize silica aerogels. The development of new empirical forcefields, including the modified BKS potential, the Vashishta potential, and the re-parameterized Tersoff potential, accurately captured numerous properties of amorphous silica. By coupling these forcefields with novel methods of generating percolated architectures, researchers can now perform multiscale optimization and characterization of silica aerogels' structure-property-function relationship. Both fully atomistic and coarse-grained molecular dynamics modelling methods have been discussed, including negative pressure rupturing, the expanding, heating, and quenching scheme and DLCA algorithms. These compelling algorithms have characterized silica aerogels' structural, mechanical, and thermal properties, delivering unprecedented insights into silica aerogels. We anticipate that the next big leap of development in silica aerogels is the manipulation of their molecular structures in a bottom-up production approach, leading to aerogels that are stronger, tougher, and more optically transparent, while retaining their incredibly low weight and thermal conductivity. This may be achieved through advances in 3D printing of ceramics in combination with rigorous characterization and optimization through a comprehensive set of multiscale modeling platforms, which ultimately minimize the trials and errors of traditional experimental synthesis and can even bring silica aerogels into the fold of the materials-by-design paradigm.

References

Ackerman WC, Vlachos M, Rouanet S, Fruendt J (2001) Use of surface treated aerogels derived from various silica precursors in translucent insulation panels. J Non Cryst Solids 285:264–271

Aegerter MA, Leventis N, Koebel MM (eds) (2011) Aerogels handbook. Springer, New York

Alaoui AH, Woignier T, Scherer GW, Phalippou J (2008) Comparison between flexural and uniaxial compression tests to measure the elastic modulus of silica aerogel. J Non Cryst Solids 354:4556–4561

Allen MP, Tildesley DJ (1989) Computer simulation of liquids. Oxford University Press, New York

Benoit C, Rahmani A, Jund P, Jullien R (2001) Numerical study of the dynamic properties of silica aerogels. J Phys Condens Mat 13:5413–5426

Bhattacharya S, Kieffer J (2005) Fractal dimensions of silica gels generated using reactive molecular dynamics simulations. J Chem Phys 122:094715

Burchell MJ, Graham G, Kearsley A (2006) Cosmic dust collection in aerogel. Annu Rev Earth Planet Sci 34:385–418

Campbell T, Kalia RK, Nakano A, Shimojo F, Tsuruta K, Vashishta P, Ogata S (1999) Structural correlations and mechanical behavior in nanophase silica glasses. Phys Rev Lett 82:4018–4021

Cantin M, Casse M, Koch L, Jouan R, Mestreau P, Roussel D, Bonnin F, Moutel J, Teichner SJ (1974) Silica aerogels used as Cherenkov radiators. Nucl Inst Methods 118:177–182

Coquil T, Fang J, Pilon L (2011) Molecular dynamics study of the thermal conductivity of amorphous nanoporous silica. Int J Heat Mass Transf 54:4540–4548

Cotana F, Pisello AL, Moretti E, Buratti C (2014) Multipurpose characterization of glazing systems with silica aerogel: in-field experimental analysis of thermal-energy, lighting and acoustic performance. Build Environ 81:92–102

Cross J, Goswin R, Gerlach R, Fricke J (1989) Mechanical properties of SiO_2 – aerogels. J Phys Colloq 50(C4):185–190

Duer K, Svendsen S (1998) Monolithic silica aerogel in superinsulating glazings. Sol Energy 63:259–267

Ebert H-P (2011) Thermal properties of aerogels. In: Aegerter MA, Leventis N, Koebel MM (eds) Aerogels handbook. Springer, New York, pp 537–564

Ferreiro-Rangel CA, Gelb LD (2013) Investigation of the bulk modulus of silica aerogel using molecular dynamics simulations of a coarse-grained model. J Phys Chem B 117:7095–7105

Ferreiro-Rangel CA, Gelb LD (2015) Computational study of uniaxial deformations in silica aerogel using a coarse-grained model. J Phys Chem B 119:8640–8650

Fesmire JE (2006) Aerogel insulation systems for space launch applications. Cryogenics 46: 111–117

Feuston BP, Garofalini SH (1990) Oligomerization in silica sols. J Phys Chem 94:5351–5356

Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications, 2nd edn. Academic, San Diego

Fricke J (1988) Aerogels – highly tenuous solids with fascinating properties. J Non Cryst Solids 100:169–173

Fricke J, Emmerling A (1992) Aerogels. J Am Ceram Soc 75:2027–2036

Fricke J, Lu X, Wang P, Buttner D, Heinemann U (1992) Optimization of monolithic silica aerogel insulants. Int J Heat Mass Transf 35:2305–2309

Gao T, Jelle BP, Ihara T, Gustavsen A (2014) Insulating glazing units with silica aerogel granules: the impact of particle size. Appl Energy 128:27–34

Garofalini S (1983) A molecular dynamics simulation of the vitreous silica surface. J Chem Phys 78:2069–2072

Garofalini SH, Martin G (1994) Molecular simulations of the polymerization of silicic acid molecules and network formation. J Phys Chem 98:1311–1316

Garofalini SH, Melman H (1986) Applications of molecular dynamics simulations to sol-gel processing. MRS Proc 73:497

Gelb LD (2007) Simulating silica aerogels with a coarse-grained flexible model and langevin dynamics. J Phys Chem C 111:15792–15802

Gelb LD (2011) Simulation and modeling of aerogels using atomistic and mesoscale methods. In: Aegerter MA, Leventis N, Koebel MM (eds) Aerogels handbook. Springer, New York, pp 565–581

Gonçalves W, Morthomas J, Chantrenne P, Perez M, Foray G, Martin CL (2016) Molecular dynamics simulations of amorphous silica surface properties with truncated Coulomb interactions. J Non Cryst Solids 447:1–8

Gronauer, M, Kadur, A, Fricke, J (1986) Mechanical and acoustic properties of silica aerogel. Aerogels: Proceedings of the first international symposium, Würzburg, Fed Rep Germany, 23–25 Sept 1985, Springer, Berlin Heidelberg, pp 167–173

Gross J, Reichenauer G, Fricke J (1988) Mechanical-properties of SiO_2 aerogels. J Phys D Appl Phys 21:1447–1451

Guissani Y, Guillot B (1996) A numerical investigation of the liquid-vapor coexistence curve of silica. J Chem Phys 104:7633–7644

Haile J, Johnston I, Mallinckrodt AJ, McKay S (1993) Molecular dynamics simulation: elementary methods. Comput Phys 7:625–625

Heinemann U, Caps R, Fricke J (1996) Radiation conduction interaction: an investigation on silica aerogels. Int J Heat Mass Transf 39:2115–2130

Herrmann G, Iden R, Mielke M, Teich F, Ziegler B (1995) On the way to commercial production of silica aerogel. J Non Cryst Solids 186:380–387

Horbach J, Kob W (1999) Static and dynamic properties of a viscous silica melt. Phys Rev B 60:3169–3181

Horbach J, Kob W, Binder K (1999) Specific heat of amorphous silica within the harmonic approximation. J Phys Chem B 103:4104–4108

Hrubesh LW (1998) Aerogel applications. J Non Cryst Solids 225:335–342

Jelle BP, Baetens R, Gustavsen A (2011) Aerogel insulation for building applications: a state-of-the-art review. Energ Buildings 43:761–769

Jones SM (2006) Aerogel: space exploration applications. J Sol Gel Sci Technol 40:351–357

Jund P, Jullien R (1999) Molecular-dynamics calculation of the thermal conductivity of vitreous silica. Phys Rev B 59:13707–13711

Kharzheev YN (2008) Use of silica aerogels in Cherenkov counters. Phys Part Nucl 39:107–135

Kieffer J, Angell CA (1988) Generation of fractal structures by negative-pressure rupturing of SiO_2 glass. J Non Cryst Solids 106:336–342

Kieffer J, Bhattacharya S (2008) Molecular dynamics simulation study of growth regimes during polycondensation of silicic acid: from silica nanoparticles to porous gels. J Phys Chem C 112:1764–1771

Kistler SS (1931) Coherent expanded aerogels and jellies. Nature 127:741–741

Kistler SS (1932) Coherent expanded aerogels. J Phys Chem 36:52–64

Koebel MM, Rigacci A, Achard P (2011) Aerogels for superinsulation: a synoptic view. In: Aegerter MA, Leventis N, Koebel MM (eds) Aerogels handbook. Springer, New York, pp 607–633

Kramer GJ, Farragher NP, van Beest BW, van Santen RA (1991) Interatomic force fields for silicas, aluminophosphates, and zeolites: derivation based on ab initio calculations. Phys Rev B 43:5068–5080

Lei J, Liu Z, Yeo J, Ng TY (2013) Determination of the Young's modulus of silica aerogels – an analytical-numerical approach. Soft Matter 9:11367–11373

Lemay JD, Tillotson TM, Hrubesh LW, Pekala RW (1990) Microstructural dependence of aerogel mechanical properties. MRS Proc 180:321

Liu M, Qiu L, Zheng X, Zhu J, Tang D (2014) Study on the thermal resistance in secondary particles chain of silica aerogel by molecular dynamics simulation. J Appl Phys 116:093503

Lu X, Caps R, Fricke J, Alviso CT, Pekala RW (1995) Correlation between structure and thermal-conductivity of organic aerogels. J Non Cryst Solids 188:226–234

Mahajan SS, Subbarayan G, Sammakia BG (2007) Estimating thermal conductivity of amorphous silica nanoparticles and nanowires using molecular dynamics simulations. Phys Rev E 76:056701

McGaughey AJH, Kaviany M (2004) Thermal conductivity decomposition and analysis using molecular dynamics simulations part II. Complex silica structures. Int J Heat Mass Transf 47:1799–1816

Muller-Plathe F, Bordat P (2004) Reverse non-equilibrium molecular dynamics. Lecture Notes in Physics, vol 640. Springer, Berlin Heidelberg, pp 310–326

Munetoh S, Motooka T, Moriguchi K, Shintani A (2007) Interatomic potential for Si-O systems using Tersoff parameterization. Comput Mater Sci 39:334–339

Murillo JSR, Bachlechner ME, Campo FA, Barbero EJ (2010) Structure and mechanical properties of silica aerogels and xerogels modeled by molecular dynamics simulation. J Non Cryst Solids 356:1325–1331

Nakano A, Bi L, Kalia RK, Vashishta P (1993) Structural correlations in porous silica: molecular dynamics simulation on a parallel computer. Phys Rev Lett 71:85

Nakano A, Bi LS, Kalia RK, Vashishta P (1994) Molecular-dynamics study of the structural correlation of porous silica with use of a parallel computer. Phys Rev B 49:9441–9452

Neugebauer A, Chen K, Tang A, Allgeier A, Glicksman LR, Gibson LJ (2014) Thermal conductivity and characterization of compacted, granular silica aerogel. Energ Buildings 79: 47–57

Ng TY, Yeo JJ, Liu ZS (2012) A molecular dynamics study of the thermal conductivity of nanoporous silica aerogel, obtained through negative pressure rupturing. J Non Cryst Solids 358:1350–1355

Ng TY, Joshi SC, Yeo JJ, Liu ZS (2016) Effects of nanoporosity on the mechanical properties and applications of aerogels in composite structures. In: Meguid SA (ed) Advances in nanocomposites. Springer International Publishing, Cham, Switzerland pp 97–126

Patil SP, Rege A, Sagardas, Itskov M, Markert B (2017) Mechanics of nanostructured porous silica aerogel resulting from molecular dynamics simulations. J Phys Chem B 121(22):5660–5668

Pohl PI, Faulon JL, Smith DM (1995) Molecular-dynamics computer-simulations of silica aerogels. J Non Cryst Solids 186:349–355

Rahmani A, Benoit C, Poussigue G (1994) A fractal model for silica aerogels. J Phys Condens Matter 6:1483

Rapaport DC (2004) The art of molecular dynamics simulation. Cambridge University Press, Cambridge

Rubin M, Lampert CM (1983) Transparent silica aerogels for window insulation. Sol Energy Mater 7:393–400

Scherer GW, Smith DM, Qiu X, Anderson JM (1995) Compression of aerogels. J Non Cryst Solids 186:316–320

Shell MS, Debenedetti PG, Panagiotopoulos AZ (2002) Molecular structural order and anomalies in liquid silica. Phys Rev E Stat Nonlin Soft Matter Phys 66:011202

Stillinger F, Rahman A (1978) Revised central force potentials for water. J Chem Phys 68:666–670

Tabata M, Imai E, Yano H, Hashimoto H, Kawai H, Kawaguchi Y, Kobayashi K, Mita H, Okudaira K, Sasaki S, Yabuta H, Yokobori S-i, Yamagishi A (2014) Design of a silica-aerogel-based cosmic dust collector for the Tanpopo mission aboard the International Space Station. Trans Jpn Soc Aeronaut Space Sci, Aerosp Technol Jpn 12:Pk_29–Pk_34

Tersoff J (1988) New empirical approach for the structure and energy of covalent systems. Phys Rev B 37:6991–7000

Tillotson TM, Hrubesh LW (1992) Transparent ultralow-density silica aerogels prepared by a two-step sol-gel process. J Non Cryst Solids 145:44–50

Tsuneyuki S, Tsukada M, Aoki H, Matsui Y (1988) First-principles interatomic potential of silica applied to molecular dynamics. Phys Rev Lett 61:869–872

Vacher R, Woignier T, Pelous J, Courtens E (1988a) Structure and self-similarity of silica aerogels. Phys Rev B 37:6500–6503

Vacher R, Woignier T, Phalippou J, Pelous J, Courtens E (1988b) Fractal structure of base catalyzed and densified silica aerogels. J Non Cryst Solids 106:161–165

van Beest BW, Kramer GJ, van Santen RA (1990) Force fields for silicas and aluminophosphates based on ab initio calculations. Phys Rev Lett 64:1955–1958

Vashishta P, Kalia RK, Rino JP, Ebbsjo I (1990) Interaction potential for SiO_2 – a molecular-dynamics study of structural correlations. Phys Rev B Condens Matter Mate Phys 41:12197–12209

Wei G, Liu Y, Zhang X, Yu F, Du X (2011) Thermal conductivities study on silica aerogel and its composite insulation materials. Int J Heat Mass Transf 54:2355–2366

Wittwer V (1989) Translucent insulation for passive solar energy utilization in buildings. Sol & Wind Technol 6:419–426

Woignier T, Phalippou J (1989) Scaling law variation of the mechanical properties of silica aerogels. J Phys Colloq 50(C4):179–184

Woignier T, Pelous J, Phalippou J, Vacher R, Courtens E (1987) Elastic properties of silica aerogels. J Non Cryst Solids 95:1197–1202

Woignier T, Phalippou J, Vacher R, Pelous J, Courtens E (1990) Different kinds of fractal structures in silica aerogels. J Non Cryst Solids 121:198–201

Woignier T, Reynes J, Hafidi Alaoui A, Beurroies I, Phalippou J (1998) Different kinds of structure in aerogels: relationships with the mechanical properties. J Non Cryst Solids 241:45–52

Woignier T, Primera J, Alaoui A, Etienne P, Despestis F, Calas-Etienne S (2015) Mechanical properties and brittle behavior of silica aerogels. Gels 1:256–275

Yeo JJ (2014) Modeling and simulation of the structural evolution and thermal properties of ultralight aerogel and 2D materials. Doctoral dissertation, Nanyang Technological University, Singapore

Yeo JJ, Liu ZS, Ng TY (2013) Enhanced thermal characterization of silica aerogels through molecular dynamics simulation. Model Simul Mate Sci Eng 21:075004

Zeng JSQ, Stevens PC, Hunt AJ, Grief R, Daehee L (1996) Thin-film-heater thermal conductivity apparatus and measurement of thermal conductivity of silica aerogel. Int J Heat Mass Transf 39:2311–2317

Toughening and Strengthening Mechanisms in Bamboo from Atoms to Fibers

65

Sina Askarinejad, Sina Youssefian, and Nima Rahbar

Contents

Abstract

Bamboo, a grass, with stiffness-to-weight and strength-to-weight ratios higher than steel and concrete is a wonder of nature. In this chapter we present a comprehensive review of the mechanical properties of bamboo from nano- to macroscale. After presenting the main toughening mechanisms in bamboo, we present a summary of the atomistic simulations of bamboo microfibril which is the main sources of stiffness and strength in bamboo fibers. The role of humidity in bamboo microfibril is also presented at small scale. It is shown experimentally and numerically that the addition of 1% moisture content to the system does not change density and increases the elastic modulus in the longitudinal direction. This is probably due to the relatively larger fractional free volume in the

S. Askarinejad · S. Youssefian · N. Rahbar (✉)
Department of Civil and Environmental Engineering, Worcester Polytechnic Institute, Worcester, MA, USA
e-mail: nrahbar@wpi.edu

structure of lignin, and subsequently LCC, at low moisture contents. This interesting phenomenon is important in understanding the mechanics of bamboo microfibrils.

1   Introduction and Background

Bamboos (Graminaceae) include numerous species that grow in all continents, except for Europe and Antarctica, in tropical, subtropical, and temperate regions of the world. They are one of the swiftest growing plants on Earth in terms of rate of mass accrued during their growth phase (Zachariah et al. 2016). Bamboo's attractive combinations of high strength-to-weight ratio, high stiffness-to-weight ratio, and shape factor (Ghavami 2005; Wegst 2011; Xiao et al. 2010; Gibson 2012; Huang and Netravali 2009) have led to an increasing interest in the use of bamboo as an eco-friendly highly efficient structural material (Wegst and Ashby 2004). Such unique properties have resulted in many applications of bamboo such as in bicycles, housing in rural and urban environments (Janssen 1981), and microscopes (Basu 2007). In countries such as China, bamboo has been widely used as structural material for centuries (Yu et al. 2011a). Bamboo has also been used to fabricate structural elements and as alternative to steel in reinforced concrete (Geymayer and Cox 1970). Recent economic data also show significant growth in innovative materials from bamboo that are helping a new industry to sprout (Watching Grass Grow 2018).

Excluding the outer layer (epidermal layer) of bamboo stalk, its structure consists of fibers and vascular bundles that are surrounded by parenchyma cells. Researchers attribute the outstanding mechanical properties of bamboo to the presence of the fibers which are nonuniformly distributed in the cross section of bamboo (Tan et al. 2011; Lo et al. 2008). The nonuniform distribution and the orientation of these fibers make bamboo an orthotropic material with high strength along the fibers and low strength transversal to the fibers (Ghavami 1995; Lakkad and Patel 1980; Lo et al. 2004). This evolutionary designed structure establishes bamboo as a great representative of a naturally occurring functionally graded material (FGM) (Nogata and Takahashi 1995; Habibi et al. 2015; Ray et al. 2004; Ghavami et al. 2003). FGMs Suresh and Mortensen (1988) are generally characterized by their variation in composition and structure over volume. As the most dominant loading on a bamboo plant is bending along the main axis of the plant due to wind, the profile of bending stress along the cross section of the bamboo varies from the highest at the outer layer to the lowest at the inner later. Evolutionary, this stress profile has resulted in the corresponding variation in the material properties in a continuous manner that reduces the stress concentration and increases bending overall strength (Shao et al. 2009; Janssen 1991).

The structure of bamboo is a typical example of plants with highly developed architectures in which bamboo culm is divided by the nodes (Fig. 1). Taylor et al. studied the role of bamboo nodes and found out that beyond being a point of strength, the node may be a point of weakness when loaded in tension. However,
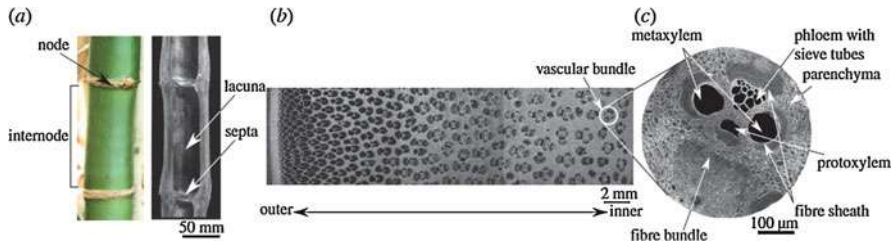
Fig. 1 Hierarchical microstructure of bamboo (Mannan et al. 2017)

the spacing of nodes is too large to affect the stiffness and strength of the tube as a whole and also greater than the critical crack length for brittle fracture. Thus, the diaphragm and ridge structure of the node are possibly evolved to reinforce a biologically essential feature which would otherwise be a point of weakness (Taylor et al. 2015). It has a hierarchical structure that makes a typical natural functionally graded bio-composite (Amada and Untao 2001). In this structure, there are no tree rings in the culm's cross section. Instead, bamboo comprises three fundamental tissues named epidermis, parenchyma, and vascular bundles (Tan et al. 2011).

The epidermis is the outermost layers of bamboo which protects the internal components from invaders. The attractive green color of bamboo culm is due to the chlorophyll of epidermis. Smooth surface of the bamboo, also, exists because of silica, a wax layer, and other substances of epidermis. The silica also facilitates CO_2 uptake of plants and strengthens the outer epidermis for prevention of external attacks (Chang et al. 1998; Chung et al. 2014).

Parenchyma tissue is the non-woody structure of bamboo. The ground parenchyma tissues of crop species such as rice and maize have only one type. Bamboo, however, has two types of parenchyma cells named short and long parenchyma cells that have differences in lignification and the hemicellulose distribution. These differences reflect the differences in function of the two types of parenchyma cells (He et al. 2002). Parenchyma ground tissue forms a matrix in which vascular bundles are placed.

The vascular bundles mainly consist of xylem, phloem, and fiber caps which are playing a decisive role in bamboo physiological growth and biomechanical function. The vascular bundle inhomogeneous distribution is dense in the outer region and sparse in the inner region (Amada et al. 1997). This special dissemination is a perfect structural adaptation toward a high bending stiffness and strength of the bamboo culm. The conducting elements (vessels and phloem) provide a channel for the transport of water and nutrients. The fibers, which are accounting for approximately 40% of a culm by volume, primarily carry out the function of mechanical support.

The fibers around the phloem do not have the same thickness. The immature fibers close to the vessels and the phloem complete their wall thickening first, whereas those at the periphery of the fiber caps are in a transition state. Therefore, a distinct density gradient can be observed from the fibers adjacent to the vessels

toward the fibers connected to the surrounding parenchymatic tissue (Wang et al. 2012a). The six different fiber wall types could be distinguished in the bundle fibers close to phloem that all have mainly hemicellulose-lignin matrix surrounding the cellulose microfibrils, in their secondary walls. The degree of lignification varies remarkably across the fiber wall, with a higher lignin content present in the narrow layers (Wang et al. 2012a, b).

There are numerous studies on material properties of bamboo, including measuring strength, Young's modulus of matrix and fiber, and detailed analysis of microstructures and fiber distribution (Nogata and Takahashi 1995; Amada et al. 1997; Tan et al. 2011; Bai et al. 1999; Lee et al. 1997). Dixon et al. measured the flexural properties of moso bamboo in the axial direction, along with the compressive strengths in the axial and transverse directions. Based on the microstructural variations and solid cell wall properties, analytical models of mechanical properties of bamboo which describe the experimental results were developed (Dixon et al. 2014).

Considering that bamboo is an inhomogeneous orthotropic material, numerous methods have been used to model its macroscale properties numerically (Askarinejad et al. 2015). Moreover, in order to estimate how the microstructure influences the effective properties of these materials, multiscale methods such as homogenization have been previously used (Silva et al. 2006).

The fracture toughness of bamboo culms was studied by many such as Amada et al. (2001). The average value obtained was 56.8 MPa.m$^{1/2}$. This value is higher than Al alloys. The fracture property of bamboo as a natural composite material depends on where the crack initiates. Using multiscale mechanical characterizations coupled with advanced environmental electron microscopy (ESEM), Habibi and Lu showed that fiber interfacial areas along with parenchyma cell boundaries were preferred routes for crack growth in both radial and longitudinal directions (Habibi and Lu 2015). The occurrence of interfacial fractures within the cellular parenchyma matrix along with honeycomb structure of fibers in both radial and longitudinal directions is mainly responsible for the remarkable fracture toughness of bamboo by reducing the overall effective stress intensity factor. Keogh et al. performed fatigue tests on bamboo culm (Keogh et al. 2015). Bamboo culm samples display fatigue failure when loaded in compression across their diameters, simulating a type of deformation which commonly occurs as a result of culm bending. However, they observed no fatigue behavior when samples are loaded in axial compression.

2   Toughening Mechanisms in Bamboo

The main toughening mechanisms in bamboo are observed to be fiber bridging (Tan et al. 2011; Habibi and Lu 2015). Habibi and Lu investigated the asymmetric flexural responses of natural bamboo strips under different loading configurations, during elastic bending and fracture failure stages, with their respective deformation mechanisms at microstructural level (Habibi and Lu 2015). They found fiber

bridging and crack deflection are also partly responsible for the gradient and remarkable flexural strength and toughness of bamboo strips. Hence, as the overall volume fraction of fiber decreased (thickness of the strips approached the entire thickness of bamboo culm), the contribution of triggered toughening mechanisms, due to the presence of fibers, also diminished which resulted in minimizing the flexural strength and toughness of bamboo strips. They performed the flexural tests in modes A and B. In the mode A, the load was exerted on the higher fiber density side and the mode B on the lower fiber density side of bending configurations. For mode A fracture, the crack bridging, due to the fiber bridging, alongside its resultant fiber debonding is the most significant toughening mechanism. Accordingly, the layers with very large volume fraction of fibers result in very low extent of crack bridging and fiber debonding in mode B, because delamination occurred more often. This considerably reduces the flexural strength.

Constant moment crack growth resistance tests were previously performed on single-edge-notched bend (SENB) specimens (Tan et al. 2011; Askarinejad et al. 2015). The specimens were characterized into two categories, according to the placement of the notch. Specimens notched in the outer side of longitudinal(L)/tangential (θ) orientation are hereby referred to as outside cracked, while those notched on inside part of the L/θ orientation are referred to as inside cracked.

Three-dimensional finite element simulations were performed to extract the exact values of crack-driving forces. Fiber volume fractions as a function of thickness were extracted from the optical microscopy measurements. Later, classical rule of mixture was applied to estimate the actual graded elastic properties of bamboo microstructure as a function of distance from the outer layer. The estimate of bamboo Young's modulus as a function of the distance from the outer layers was applied to bamboo cross section in the finite element simulation, Askarinejad et al. (2015).

The cracks in both inside- and outside-notched samples grow parallel to the longitudinal direction of fibers. In all the specimens tested, the crack does not grow straight. All cracks kink 90° with respect to fiber orientation and advance parallel to the fiber direction. It was observed that the bamboo toughness appears to depend on the placement of the notch, with higher value for inside-notched specimen than the outside-notched specimens. After pre-cracking further stable crack growth occurs at a higher load.

The fracture toughness of a composite structure, toughened by the mechanism of crack bridging, was first modeled by Budiansky et al. using elastic-ideally plastic springs (Budiansky et al. 1988). This model analytically estimates the toughening due to small-scale bridging (SSB). In this case the bridge length is relatively small compared to crack length, specimen dimensions, and distances from the crack to the specimen boundaries. The toughening due to SSB, $\triangle K_{SSB}$, is given by,

$$\triangle K_{SSB} = \sqrt{\frac{2}{\pi}} V_f \int_0^L \frac{\alpha \sigma_y}{\sqrt{x}} dx \qquad (1)$$

where α is a constraint/triaxiality factor, V_f is the volume fraction of ductile phase, σ_y represents the uniaxial yield stress, x is the distance from the crack tip, and L is the bridging length, which is equal to the distance from the crack tip to the last uncracked reinforcement.

A large-scale bridging model developed by Bloyer et al. (1998, 1999) was used to estimate the toughening due to large-scale bridging (LSB) effects. This model that depends on the traction distribution, crack length, and specimen geometry employs a weight function $h(a, x)$ to calculate the stress intensity diminution caused by the shielding zone. For LSB the toughening due to ligament bridging $\triangle K_{LSB}$ is given by,

$$\triangle K_{LSB} = V_f \int_0^L \alpha \sigma_y h(a, x) dx \qquad (2)$$

The weight function for the bridging tractions is given by Fett and Munz (1994). Hence, the fracture toughness of the composite bamboo structure is the summation of crack initiation toughness K_m and the added toughness, $\triangle K$, due to SSB or LSB. K_m is referring to mixed mode matrix toughness before fiber bridging happens and is found through experiments.

$$K_c = K_m + \triangle K \qquad (3)$$

Once K_c is obtained, the energy release rate can be calculated.

For the case of bamboo, α can be taken to be 1.0 (Budiansky et al. 1988), and the volume fractions of fibers for outside and inside cracks are 0.45 and 0.2, respectively. The uniaxial yield stress for outside crack with denser fibers is around 620 MPa, whereas for inside crack is around 220 MPa. The Young's moduli are around 16 and 7.5 GPa for outside and inside cracks, respectively.

The resistance curve for the bamboo structure as a function of crack (kinked) length is presented in Fig. 2. It is noted that the R-curve of bamboo does not exhibit saturation behavior. However, the curve rises gradually up to point of instability. This instability point can be considered as a point of failure for any given notch position. In general, it can be said that this corresponds to a crack size of ~12 mm. The differences in the fracture resistance of the bamboo in different notch orientations can be explained as a result of the graded structure of bamboo. In general, bamboo derives its strength from the vascular fibers that can sustain higher stresses during deformation. In the presence of notches or cracks, however, the high energy involved with the deformation and bending of regions with higher density of fibers causes high stress concentrations at the notch/crack tips, which leads to high crack-driving forces at crack tips for a unit load. Meanwhile, the regions with the least fiber density do not result in high stress concentrations at crack tips, but rather tend to dissipate the energy by some inelastic deformation. It could be said that the much higher loads observed for the inside-notched sample can be explained by two following reasons: (a) the higher level of strain energy of the uncracked outer layer
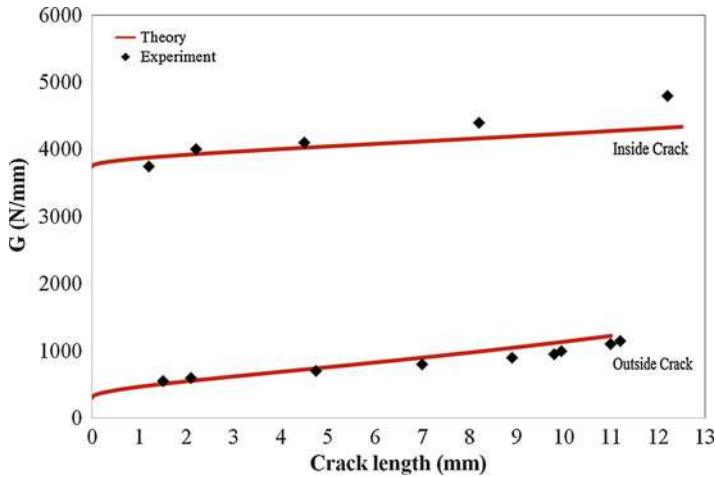
Fig. 2 Fiber bridging toughening mechanism model and R-curve behavior of bamboo for two possible crack configurations

to applied bending loads and (b) the energy-dissipating capability of lesser fiber-density notched region. The combination of these two conditions results in higher crack initiation/propagation loads for a sample with an inside crack. For the case of outside-notched sample, the opposite occurs. That is, the part in contact with the applied load (the bottom part of the bamboo) is weak and bends easily under lower applied loads. The resulting bending stress is tensile in the region that contains the notch. As a result of the high fiber density, which causes high stress concentrations at the crack tip, much lower stresses are magnified to levels equivalent to crack-driving force. Hence, lower loads for crack initiation/propagation are observed for the outside-notched samples.

The main effect of mixed mode on toughness of a bamboo is postulated to be on the matrix toughness, K_m. Afterward, the large-scale mode-I fiber bridging formulations can precisely predict the toughening behavior of bamboo. Hence, as the formulation clearly shows, the contribution of mode-II toughening mechanism to intrinsic toughness of bamboo is negligible.

It is quite clear from Fig. 2 that the bamboo exhibits R-Curve behavior in both inside and outside cracked directions. However, this behavior is more pronounced for samples with an outside crack than samples with an inside crack. The results show that the crack starts from two corners of the notch due to the relatively high stress concentrations at these points. However, the load at which crack growth starts depends on the side of the bamboo that contains the notch, being highest for the inside-notched specimen and least for the outside-notched specimen. The important observation from these experiments was that the direction of crack growth or the crack growth path deviates from that of the original notch, which is perpendicular to fiber longitudinal direction.

Another observation from crack/microstructure interaction is that the length and width of the crack increases with applied loads, with bridging across the faces of the cracks. Such bridges absorb energy and, thus, reduce the effective stress intensities at the crack tip. Hence, higher applied loads are required for crack growth, leading to rising observed crack-resistance curves (R-curves) of Fig. 2. A comparison of crack/microstructure interaction reveals that the effects of crack bridging on toughness in the inside-notched specimen are larger and more pronounced, due to the lower fiber density. This shows that the inner side of the bamboo culm is tougher than the outer side, which in turn is stronger than the inner part.

3 Multiscale Mechanical Properties of Bamboo Fibers

Fibers are a fundamental structural component of the bamboo culm as they provide mechanical support. The structure of the fibers in which cellulose fibrils are surrounded by a matrix of mainly lignin and hemicellulose has been widely investigated (Youssefian and Rahbar 2015; Rao et al. 2007). Between 20% and 30% of the cross-sectional area of the stalk is made of longitudinal fibers which are heavily concentrated near the exterior layer (Silva et al. 2006). To gain better understanding of the principles of mechanical optimization in bamboo fibers, a single-fiber tensile test at four different span lengths for fibers of the bamboo species was performed by Osorio et al. (2011). The average strength values of 800 MPa and Young's modulus of 43 GPa were obtained. The strength and elastic modulus of bamboo fibers was reported to be around 610 MPa and 46 GPa, respectively (Amada et al. 1997; Amada and Untao 2001). Nogata and Takahashi compared the strength and elastic modulus of bamboo fibers with different steel alloys, aluminum alloys, and ceramics. They showed that the bamboo fiber has a specific elastic modulus and strength which are much higher than that of steel, Al alloys, and ceramics (Nogata and Takahashi 1995).

Microstructures of bamboo fiber have been previously investigated using scanning electron microscope (Phong et al. 2011). The results showed the fibers were longitudinally aligned in the flat-wise direction. They were together bonded by hemicelluloses and lignin with different bonding levels. Mechanical properties of fiber bundle and their interfacial shear strength (IFSS) with typical resins such as unsaturated polyester and polypropylene are related to lignin and hemicelluloses content in chemical composition. The experimental results also proved that the IFSS of alkaline-treated fibers were better than that of steam-exploded and untreated fibers. Moisture absorptions of bamboo fibers at 50%, 60%, 70%, 80%, and 90% relative humidity at 200 °C were also determined. The results showed that the difference of moisture content of the alkaline-treated and the steam-exploded fibers was much lower than that of the untreated fibers at all relative humidity levels.

The chemical composition of bamboo fiber consists mainly of cellulose, hemicelluloses, and lignin. These components are actually the same high-glycans and make up about 90% of total weight of bamboo fiber. The other constituents are protein, fat, pectin, tannins, pigments, and ash. These constituents play important

role in physiological activity of bamboo, and they are found in cell cavity or special organelles (Khalil et al. 2012). The lignin is considered to provide stiffness and yellow color to bamboo fibers. Different treatments cannot remove all the lignin content of the bamboo fibers, as lignin has been found quite resistant to various alkalis. Noncellulosic components have enough contribution to fiber properties such as strength, flexibility, moisture, and even density (Li et al. 2010). The unidirectional arrangement of bamboo fibers in tissues and cell wall structure of bamboo is one of unique properties of bamboo.

Bamboo fiber is different from the other cellulose fibers in morphology, crystalline structure, and molecular structure. In SEM images, the cross section of the bamboo fiber is round with a small lumen, and there is no node in the longitudinal surface (Yueping et al. 2009). Most of the bamboo fibers have multilamellar cell walls with various layers. By means of viscometry, Yueping et al. showed that the bamboo fiber has a small molecular mass and degree of polymerization (Yueping et al. 2009). Wegst et al. reported that bamboo cellulose microfibrils imbedded in a lignin-hemicellulose matrix shaped into hollow prismatic cells of varying wall thickness (Wegst et al. 2014). Natural fibers themselves are cellulose fiber-reinforced materials, and the microfibril angle and cellulose content determine the mechanical behavior of the fiber (Jain et al. 1992).

3.1   Nanostructure of Bamboo Microfibrils

The bulk properties of materials are dictated not only by their elemental composition and molecular connectivity but also by their macromolecular architecture. A principal example of this is observed in plant cell walls, which are composite materials comprising biopolymers designed to serve multiple functions including structural support, microbial defense, and transport of water and nutrients throughout the organism.

Ciesielski et al. introduced a computational methodology to extract nanoscale geometry of cellulose microfibrils within thermochemically treated biomass directly from electron tomographic data sets. Computational analysis of the tomographic data is used to extract mathematical descriptions for longitudinal axes of cellulose microfibrils from which we calculate their nanoscale curvature (Ciesielski et al. 2013). They have extracted mathematical representations for the nanoscale geometry of cellulose microfibrils in pretreated biomass from tomographic data sets. The resulting parametric equations were used to calculate the curvature of plant cellulose microfibrils for the first time. They also found that orienting microfibril models to bend about the 100 crystal planes reduced the formation of kink defects after energy minimization (Ciesielski et al. 2013). The structure of the cellulose microfibrils of spruce wood was investigated by Fernandes et al., using a range of spectroscopic methods coupled to small-angle neutron and wide-angle X-ray scattering (Fernandes et al. 2011). Disorder in chain packing and hydrogen bonding was shown to increase outward from the microfibril center. The extent of disorder blurred the distinction between the Iα and Iβ allomorphs. Chains at the surface were

distinct in conformation, with high levels of conformational disorder at C-6, less intramolecular hydrogen bonding, and more outward-directed hydrogen bonding.

Fernandes et al. suggest that cellulose microfibrils with about 24 chains possibly twisted and with considerable disorder increasing toward the surfaces. Less extensive disordered regions probably exist within the core of each microfibril. Tight lateral binding is facilitated by the hydrogen-bonding pattern of the surface chains, at only a small cost in tensile stiffness due to the loss of intramolecular hydrogen bonding between $O-_2$ and $O-_6$. The surface location of these disordered chains means that a parallel mechanical model with tight lateral bonding to the ordered core is more appropriate than a series model, consistent with Fourier-transform infrared spectroscopy (FTIR) observations. The twisting of adjacent cellulose microfibrils may explain why they do not come into crystallographic register with one another over enough of their length to merge (Fernandes et al. 2011). A rectangular model is favored for spruce microfibrils, differing from more crystalline celluloses in having one hydrophobic and one hydrophilic face exposed. It is not clear how other polymers restrict access of celluloses to the hydrophobic surfaces, but impacts of the structure and aggregation of microfibrils on the recalcitrance of woody biomass are to be expected.

Kulasinski et al. investigate the influence of adsorbed water on amorphous cellulose structure and properties, within the full range of moisture content from the dry state to saturation, by molecular dynamics simulation (Kulasinski et al. 2014a). They found out that the properties of amorphous cellulose undergo a significant change about 10% of moisture content.

3.2   Molecular Origin of Strength and Stiffness in Bamboo Microfibrils: Molecular Dynamics Simulations

COMPASS (Sun et al. 1998) (cff91 ver. 2.644) was chosen as a proper force field for accurate atomistic simulations. COMPASS has been used successfully by many researchers for modeling the structural and mechanical properties of cellulose (Miyamoto et al. 2009; Moon et al. 2011; Sturcov et al. 2005; Eichhorn and Davies 2006; Eichhorn et al. 2005). Its accuracy for modeling cellulose molecular structure is proven.

In COMPASS, the nonbonded energies include van der Waals and electrostatic energies, with hydrogen bonds being a natural consequence of electrostatic energies. To study the hydrogen bond energies between particles, the following criteria have been used:

1. The maximum distance between the hydrogen and the acceptor atom for which hydrogen bonding is possible is 2.5 Å.
2. The minimum angle between the donor, hydrogen, and acceptor atoms in degrees for which hydrogen bonding is possible was chosen as 120°.

We have used a CHARMM-like MacKerell et al. (1998) hydrogen-bonding potential such as Eq. 4 to calculate the hydrogen bond energies between particles,

$$E_{hb} = D_{hb}\left[5\left(\frac{r_{hb}}{r_{DA}}\right)^{12} - 6\left(\frac{r_{hb}}{r_{DA}}\right)^{10}\right]\cos^4(\theta_{DHA}) \tag{4}$$

where θ_{DHA} is the bond angle between hydrogen donor (D) and the hydrogen (H) and the hydrogen acceptor (A). r_{DA} is the distance between the donor and acceptor. The values of D_{hb} and r_{hb} were adopted from the literature (Mayo et al. 1990).

To study the nanostructures of the models, radial distribution function (RDF) was used. RDF gives a measure of the probability of finding an atom, within a spherical shell of infinitesimal thickness at a distance, r, from the reference atom. To calculate elastic modulus, the periodic structures were expanded along each direction to the maximum strain amplitude of 0.01 in ten steps. In each step the stresses were obtained from virial stress expression which is commonly used to relate the computed stress in molecular dynamics to continuum stresses.

The unique mechanical properties of bamboo fibers come from their composite structure, in which cellulose fibrils are surrounded by a matrix of mainly lignin and hemicellulose (Zou et al. 2009; Wegst et al. 2014; Fernandes et al. 2011). Prior researchers have investigated the mechanical properties of cellulose fibers, lignin, and hemicellulose and their interactions (Chung et al. 2002; Moon et al. 2011; Sathitsuksanoh et al. 2010; Lahiji et al. 2010; Iwamoto et al. 2009; Sinko et al. 2013; Kulasinski et al. 2014b; Cousins 1976, 1978; Bergander and Salmen 2002; Salmen 2004; Dri et al. 2013). Among those previous efforts, Besombes and Mazeau (2005a, b) performed molecular dynamics simulations on the assembly of a threo guaiacyl β-O-4 dimer of lignin and different surfaces of cellulose. They showed that the adsorption of lignin onto cellulose is surface-dependent. They also found a major contribution of van der Waals interactions onto (100) face and a major influence of hydrogen bond interactions in the adsorption of lignin onto (110) and ($\bar{1}$10) faces. Linder et al. (2013) performed non-equilibrium molecular dynamics simulations of lignin and cellulose. They found that lignin strongly associates with the cellulose nanofibril. Nevertheless, few attempts have been made to use the cross-linked structure of lignin-hemicellulose to study the mechanical behavior of bamboo microfibrils and the interfaces of its matrix with cellulose nanofibrils.

Youssefian and Rahbar presented the role of each component in the mechanical properties of bamboo microfibrils as well as the underlying mechanisms of interactions between the matrix and cellulose nanofibrils (Youssefian and Rahbar 2015). With this aim, molecular models of lignin, bamboo hemicellulose, and a cross-linked structure of these two materials as representative matrix materials were developed. Molecular dynamics techniques described above were used to elucidate the structures, mechanical properties of the lignin, hemicellulose, and the matrix. It is important to note that the properties of the systems under study are known to be sensitive to the percentage of water molecules (Chung et al. 2002). Therefore, initially, no water molecule was added to the simulations to remove this effect from the results.

The most abundant carbohydrate in bamboo is cellulose with the volumetric percentage ranging from 40% to 45%. Figure 3 presents the structure of cellulose
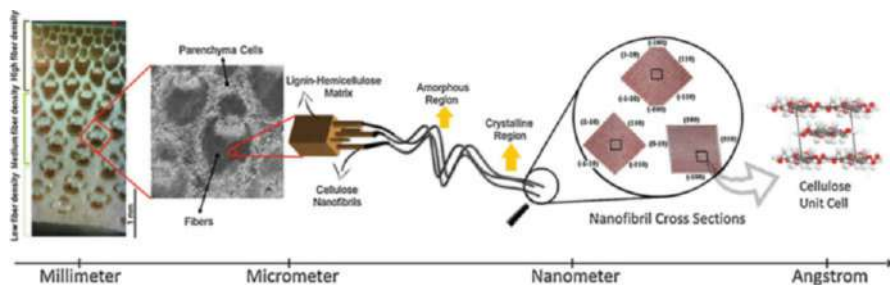
Fig. 3 Hierarchal structure of bamboo fiber Youssefian and Rahbar (2015)

at different scales down to its building unit cell. Cellulose nanofibrils are formed by assembling linear chains of aldehyde sugars often referred to as glucose molecules, to make either rectangular or hexagonal cross sections with diameters of 3 to 5 nm (Lahiji et al. 2010; Ding and Himmel 2006). If the hydrogen bonds between the hydroxyl groups form in an order, highly ordered (crystalline) regions are formed. However, if random hydrogen bonds form, disordered (amorphous) regions develop (Lahiji et al. 2010; Pakzad and Yassar 2013). The positions of the hydroxyl groups determine the crystal system. These can be either triclinic or monoclinic unit cells (α or β type, respectively) with latter being the building block of plants such as bamboo. In a bamboo fiber, cellulose nanofibrils are surrounded by lignin-carbohydrate complex (LCC) matrices that mainly contain lignin and hemicellulose with volumetric percentages that ranging from 20% to 30% and 15% to 25% of the bamboo structure, respectively.

Lignin is a natural phenolic macromolecule that mainly presents in the plant secondary cell wall. It is made up of three main phenylpropanoid subunits, namely, p-hydroxyphenyl (H-type), guaiacyl (G-type), and syringyl (S- type) units (Malherbe and Cloete 2002). The biosynthesis of lignin occurs from different polymerizations of these three subunits. Hence, there are many possible bonding patterns between the individual units. Advancements in spectroscopic methods, however, have enabled scientists to elucidate the leading structural features of lignin (Pu et al. 2008). They have also enabled scientists to propose different models for the molecular structure of lignin (Adler 1977; Brauns 1962; Forss et al. 1966; Nimz 1974; Glasser and Glasser 1974). In this study, a structural model with 28 subunits of lignin proposed by Sakakibara (1980) has been used (Sakakibara 1980). In this model, the value of the structural units and the number of protons per C9 structural units are close to that of spruce milled wood lignin reported by other researchers.

In order to be able to provide a rigid support and shape for plants, lignin polyphenols are linked together in a three-dimensional cross-linked structures by covalently bonding to hemicellulose (Vanholme et al. 2008; Xiao et al. 2001; Sun et al. 2000). Hemicelluloses are a heterogeneous group of polysaccharides that, unlike the cellulose, frequently have side chain groups. They are essentially amorphous with little strength (Scheller et al. 2010). The two major categories

of hemicelluloses are glucomannans and xylans. Bamboo hemicellulose has been shown to be a xylan and further characterized as a β-(1→4)-linked-xylopyranosyl backbone, with the presence of L-arabinofuranose and 4-O-methyl-D-glucuronic acid as single side chains (4-O-methy1-D-glucurono- arabino-xylan (Maekawa 1976)) that are arranged in an irregular manner. Therefore, the positions of the side chains are not fully determined. With the ratio of uronic acid/arabinose/xylose of 1:3:32 reported (Shi et al. 2013), we consider hemicellulose structure as hybrid chains of two extreme positions of the side chains, one is a hemicellulose structure in which two functional groups are attached to the adjacent xylans and are in the closest possible distance (CPD). The other structure has the functional groups in the furthest possible distance (FPD). A preliminary study of the energy of different configurations revealed that the total energies were of the same order of magnitude. Hence, from the energy point of view, they are both acceptable. Hemicellulose molecules bond to lignin by a variety of different chemical bonds; however, most of the evidence refers to ether and ester bonds. Jeffries proposed structures for ester and ether linkages for lignin/uronic acid and lignin/arabinoxylan groups, respectively (Jeffries 1990). These linkage models have been used to create the bonds between lignin and hemicellulose in the cross-linked LCC network. Based on the structures discussed here, molecular models of the typical lignin, hemicellulose, LCC, and cellulose nanofibrils were created. Atomistic simulation techniques were used to study the structure, thermodynamic and mechanical properties, and interactions of the bamboo microfibril materials.

Thermodynamic properties such as density and glass transition temperature are good physical properties for evaluating the realism of the conformations of atomistic models and the reliability of the COMPASS force field in simulating mechanical behavior of lignin and hemicellulose. The computed glass transition temperature for LCC, presented in Table 1, falls between the obtained values for hemicellulose and lignin which are in good agreements with the experimental data. The final relaxed conformations of lignin and hemicellulose have average densities of 1.26 $\pm$ 0.02 and 1.45 $\pm$ 0.03 g/cc, respectively. These are comparable with experimental results, in Table 1. Hence the lignin and hemicellulose models estimate the real densities within 5.2% and 4.6%, respectively. These differences are partly attributed to the vacuum condition assumed in the simulation that does not occur in the experiments. The density of LCC relaxed conformation was determined to be 1.34 g/cc. Implementing the simple rule of mixture density, in which two materials are mixed physically without any changes in mass or volume, the LCC density, ρ_{LCC}, was estimated to be about 1.35 g/cc, using densities obtained from the lignin and hemicellulose simulations. This value is close to the density of the relaxed LCC model. Therefore, the molecular dynamics calculations showed the volume fractions of hemicellulose and lignin do not significantly change in the process of LCC creation. Hence, by using the same equation and the available experimental values for lignin and hemicellulose densities, the true value of LCC density can be estimated at 1.41 g/cc (Table 1).

The Young's moduli of the hemicellulose, lignin, and LCC are presented in Table 1. The average Young's moduli of lignin and hemicellulose were estimated to

Table 1 Density (g/cc), glass transition temperature, T_g, (°C), and elastic modulus of lignin, hemicellulose, and LCC obtained from the molecular dynamics simulations (Youssefian and Rahbar 2015) and experiments (Gibson 2012)

Density			Glass trans. temp.		Elastic modulus	
Material	Simul.	Exp.	Simul.	Exp.	Simul.	Exp.
Lignin	1.26	1.33	140.26	97–171	5.90	2–6.7
Hemicell.	1.45	1.52	186.06	140–180	8.40	3.5–8.0
LCC	1.34	N/A	166.11	N/A	6.93	N/A

be 5.90 ± 0.37 and 8.40 ± 0.15 GPa, respectively. These are in good agreement with the respective experimental measurements of 6.7 and 8.0 GPa for nearly dry Pinus radiata lignin and hemicellulose (consisting of arabino-4-O-methylglucuronoxylan which have a close structure to bamboo hemicellulose) (Cousins 1976, 1978). Other studies suggest respective values in the range of 1 to 2 and 3.5 to 7 GPa for different types of lignin and hemicellulose (Bergander and Salmen 2002; Salmen 2004). Hence, the results from the simulations and experiments show that hemicellulose has better mechanical properties than lignin. Performing the same procedure on the LCC model resulted in an average Young's modulus of 6.93 ± 0.31 GPa, which is between the Young's modulus of hemicellulose and lignin.

The elastic modulus of a material is defined by its resistance to the tension or compression of bonds. It is affected by factors such as interatomic and intermolecular energies per unit volume. For the amorphous materials, such as LCC, lignin, and hemicellulose, applied stresses are mostly used to overcome the nonbonded energies between the molecules. They are also used to unwind the chains rather than directly struggle with the potential energies between atoms or the crosslinking between chains. Therefore, at small strains, nonbonded energies play more important roles in determining the elastic moduli of amorphous materials. However, the effects of nonbonded energies are not equal. According to the Lennard-Jones function, nonbonded energies with lower equilibrium energies or shorter equilibrium distances exhibit greater resistance to stretching or compressing. Thus, they result in higher Young's moduli. The lower minimum potential energy creates a greater curvature (second derivative) of its energy-distance curve at the point of equilibrium and thus leads to a greater force-distance slope. The same situation holds true for the case where the minimum energies are the same, but equilibrium distances are different. In this case, the shorter equilibrium distance results in a greater curvature, which increases the slope of the force-distance curves. Therefore, the strong short-range hydrogen bonds play more important roles in determining the Young's modulus than the other weak long-range nonbonded interactions.

3.2.1 Adhesive Interactions and Mechanisms of Microfibril Strengthening

In the composite structure of a bamboo microfibril, knowledge of adhesive interactions between the different layers determines its strength. Applied stresses on a microfibril are carried either by cellulose nanofibrils, the LCC matrix, or the

interfaces of these two regions. To investigate the adhesion energies at these interfaces, 24 different assemblies of lignin, hemicellulose, and LCC on top of cellulose substrates were created, each of which simulates the interaction between one of the materials and one face of the eight possible faces of cellulose nanofibrils. The adhesion energies were computed from final trajectories of the simulations and presented in Fig. 4a. Although these results indicate that the overall adhesion energies for these materials are different, their tendencies to adhere to nanofibril faces exhibit almost the same pattern. For each material, the interaction energies of (100) and ($\bar{1}$00) faces are the lowest, whereas the energies of other faces vary around an average value. The average adhesion energy between lignin and nanofibril faces was about 152 mJ/m^2. This is higher than adhesion energy between LCC and nanofibril faces which is about 133 mJ/m^2. Hemicellulose with average adhesion energy of around 83 mJ/m^2 shows the lowest adherence to nanofibril among the three materials. This adhesion trend, also, has been shown by Hosoya et al. in the pyrolysis of hemicellulose and lignin with cellulose where lignin-cellulose interactions were significant compared to low hemicellulose-cellulose interactions (Hosoya et al. 2007). Therefore, lignin with greater overall adhesion energy to cellulose is responsible for providing strong interaction between LCC matrix and cellulose nanofibrils to create strong bamboo microfibrils.

To understand the mechanism of interactions between these materials and cellulose, we computed the electrostatic and van der Waals energies, accountable for the adhesion, as shown in Fig. 4b, 4c. These results suggest that the van der Waals energies do not change significantly over the nanofibril faces, whereas the electrostatic energies of (100) and ($\bar{1}$00) faces are less than that of other faces. Hence, the electrostatic energy is responsible for reduction of adhesion energy between cellulose (100) or ($\bar{1}$00) faces and hemicellulose, LCC, and lignin.

One of the major components of electrostatic energies at the interface of cellulose nanofibril and hemicellulose, LCC, and lignin is hydrogen bonding which are illustrated in Fig. 4d. Regardless of the slightly higher average hydrogen bond energy between lignin and cellulose, all three materials have almost the same hydrogen bond interaction energies with cellulose nanofibrils. The hydrogen bond energies show a similar pattern to the electrostatic energies presented in Fig. 4c. In other words, (100) and ($\bar{1}$00) faces have the lowest hydrogen bond energies which are similar to that of the electrostatic interactions. This suggests that different levels of hydrogen bond energies at the interface of nanofibril and hemicellulose, LCC, and lignin are the main reason for different electrostatic energies from one face to another, causing the adhesion energies between the matrix and (100) and ($\bar{1}$00) faces of nanofibril to drop.

A closer look at the (100) and ($\bar{1}$00) face provides some insights into the weak hydrogen bonds between the faces and the matrix. Simulation result shows that the surfaces of (100) and ($\bar{1}$00) are covered with hydrogen atoms (white spheres). The hydrogen atoms are bonded either to oxygen or carbon. The hydrogen atoms connected to oxygen are less exposed because they stay closer to the surface, while the hydrogen atoms connected to carbon face outward and are configured for contact with adjacent layers. Therefore, since the accessibility of the hydrogen atoms in OH
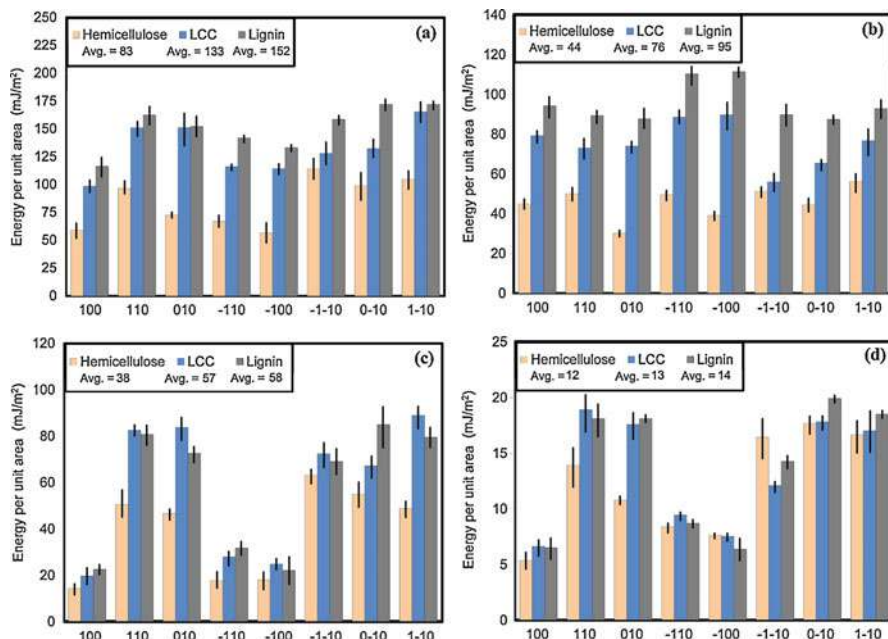
Fig. 4 (**a**) The adhesion energy per unit area between different cellulose nanofibril faces and hemicellulose, LCC, and lignin computed from the molecular dynamics simulations. (**b**) The van der Waals energy per unit area between different cellulose nanofibril faces and hemicellulose, LCC, and lignin. (**c**) The electrostatic energy per unit area between different cellulose nanofibril faces and hemicellulose, LCC and lignin. (**d**) The hydrogen bond energy per unit area between different cellulose nanofibril faces and hemicellulose, LCC, and lignin

groups is less than that in CH groups, the overall hydrogen bond energy between these two faces and the matrix diminishes.

This adhesion study proves that the weakest interaction between cellulose nanofibril and LCC occurs at the (100) and ($\bar{1}$00) faces. However, the question of which interface is the weakest link in a microfibril still remains. To answer this question, we need to calculate the other possible locations where defects may occur. Figure 5 shows a schematic of a nanofibril in a matrix of LCC that is woven around both crystalline and amorphous regions. In this structure, the stress can either detach layers of LCC/LCC, interfaces of LCC/crystalline cellulose, and LCC/amorphous cellulose or fracture the cellulose nanofibril.

The weakest adhesion energy between cellulose nanofibrils and LCC was found to be around $106 \pm 8\,\text{mJ/m}^2$. This energy is less than the energy between the two layers of LCC in the matrix, which was estimated at about $160 \pm 29\,\text{mJ/m}^2$. The adhesion energy between LCC and the amorphous region of nanofibrils was close to the adhesion energy of amorphous cellulose/amorphous cellulose interface which were estimated to be 54 ± 10 and $44 \pm 6\,\text{mJ/m}^2$, respectively. Therefore, the amorphous region of nanofibril has the lowest adhesion energy in the system.
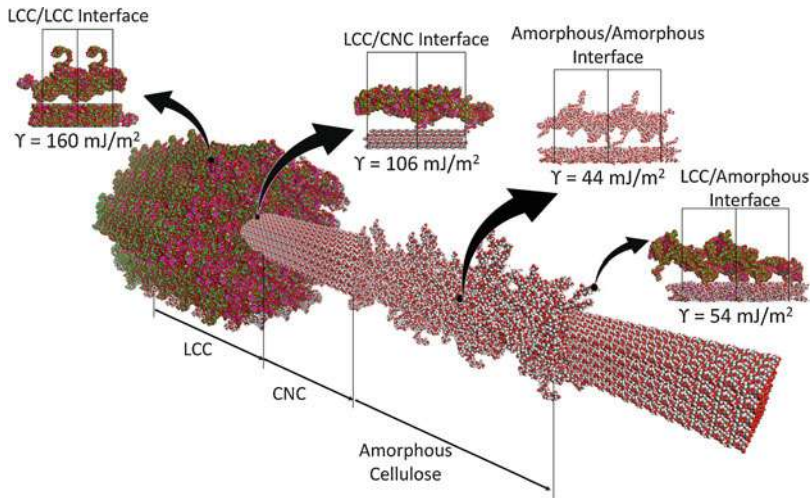
Fig. 5 The adhesion energy per unit area between different interfaces, present in a possible nanostructure of bamboo microfiber. The adhesive interaction energy at the interface of LCC layers is the highest among all the regions. The amorphous regions exhibit the lowest adhesive interactions; hence, their interface strength is likely to determine the strength of overall strength of bamboo microfibrils

4   Effect of Humidity on Mechanical Properties

Another component of the cell walls is water, which is known to have profound effects on the mechanical properties of wood. At the cell wall level, nanoindentation studies have shown that the longitudinal elastic modulus of S2 secondary layers in wood cell wall decreases with increasing moisture content (MC) (Yu et al. 2011b; Bertinetti et al. 2015). The results in these studies support the assumption that the moisture-sensitive matrix is responsible for the change in modulus of elasticity. In order to fully understand the plasticization effect of moisture on the individual wood cell wall constituents, Cousins measured the elastic moduli of extracted lignin and hemicellulose (Cousins 1976, 1977, 1978). He observed a small increase in elastic modulus of lignin from MC of 0.0% to 3.6% and linear regression at the higher MC. For hemicellulose, up to 10% MC, the loss of modulus is small, but from 10% to nearly 70%, MC the elastic modulus decreases by nearly three orders of magnitude. In another attempt, to understand the mechanism of water interactions with wood cell walls, Kulasinski et al. employed molecular dynamics simulations and showed that increasing MC above 10% decreases the stiffness of amorphous cellulose, noticeably (Kulasinski et al. 2015). Their results suggest that the moisture is adsorbed in hemicellulose and the interfaces of cellulose/hemicellulose, while no sorption occurs in the ordered cellulose (Kulasinski et al. 2014a).

In order to study the effects of humidity, nanoindentation was used to investigate the anisotropic moisture-dependent elastic moduli of bamboo fiber cell

walls (Youssefian and Rahbar 2015). Fiber wall elastic moduli were assessed using a Hysitron (Minneapolis, Minnesota, USA) TriboIndenter equipped with a Berkovich probe following the procedures in Jakes et al. (2015). Specimens were conditioned in absorption inside the enclosure for at least 36 h at room temperature (24–26 °C) under each RH, and the RH was maintained during the experiments.

Youssefian and Rahbar (2015) showed that the nanoindentation elastic moduli and the differences between the transverse and longitudinal surfaces agree well with previous work performed under ambient lab conditions (Jager et al. 2011). These results show that the longitudinal elastic modulus remained considerably higher over the range of MCs tested. The longitudinal elastic modulus is strongly influenced by longitudinal modulus of cellulose microfibrils, which provides higher modulus of elasticity due to its partly crystalline structure.

Distinguishable differences were also observed in the shapes of the elastic modulus curves with MC, and these differences are of most interest in the paper to gain new insights into the fiber wall structure-property relationships. The transverse elastic modulus continuously decreases almost linearly with increasing moisture content at the rate of 0.35 GPa per 1% moisture content from 7.9 to 2.6 GPa. In contrast, the longitudinal elastic modulus initially increases from an average value of 19.1 GPa at 0.5% moisture content to maximum of 19.4 GPa at approximately 3% MC and then decreases to 14.6 GPa at 14.5% with a rate of 0.4 GPa per 1% moisture content. A slight increase at low MCs has also been observed in bulk wood elastic moduli parallel to the grain, but elastic moduli perpendicular to the grain decrease consistently with increasing MC even at low MCs (Green et al. 1994).

This simplified definition of the stiffness of the potential energy indicates that strong short-range potential energies connect atoms with stiffer interactions than weak long-range potential energies. As a result, covalent bonds connect atoms with higher stiffness than hydrogen bonds and hydrogen bonds with higher stiffness than other nonbonded potential energies. In materials such as ordered cellulose in which external loads predominantly struggle against C-C and C-O, covalent bond potential energies dominate the elasticity and escalate the elastic modulus to relatively high value ($E_X = 11$–57 GPa, $E_Y = 50$–57 GPa and $E_Z = 110$–200 GPa) (Oliver et al. 2004; Jager et al. 2011; Green et al. 1994; Buehler et al. 2008). For amorphous materials such as the fiber cell wall matrix constituents, covalent bonds do not play significant roles, whereas hydrogen bonds primarily influence the elasticity, causing relatively low modulus elasticity of 6.93 GPa. Since water molecules with capabilities of interfering with hydrogen bond network are more likely to alter the elastic response of the matrix than the ordered regions of cellulose microfibrils, we focus our study on the influence of water molecules on hydrogen bond energies and nanostructures of bamboo fiber matrix and its constituents. Simulation results show that lignin and LCC have lower number of hydroxyl groups and are less hydrophilic compared to hemicellulose (Diddens et al. 2008). Therefore, the influence of water molecules on the hydrogen bond network of lignin, LCC, and hemicellulose should be different. In order to explore the different responses

of the matrix constitutes to absorbed moisture, atomistic simulations have been used as a powerful tool to elucidate the variation of the molecular interactions and nanostructures.

4.1   Variation of Elastic Modulus as a Function of Humidity

Figure 6 elucidates the variations of the density and modulus of elasticity of hemicellulose, lignin, and LCC with MC. Hemicellulose with higher density than lignin and LCC exhibits more sensitivity to water molecules. Introducing a small
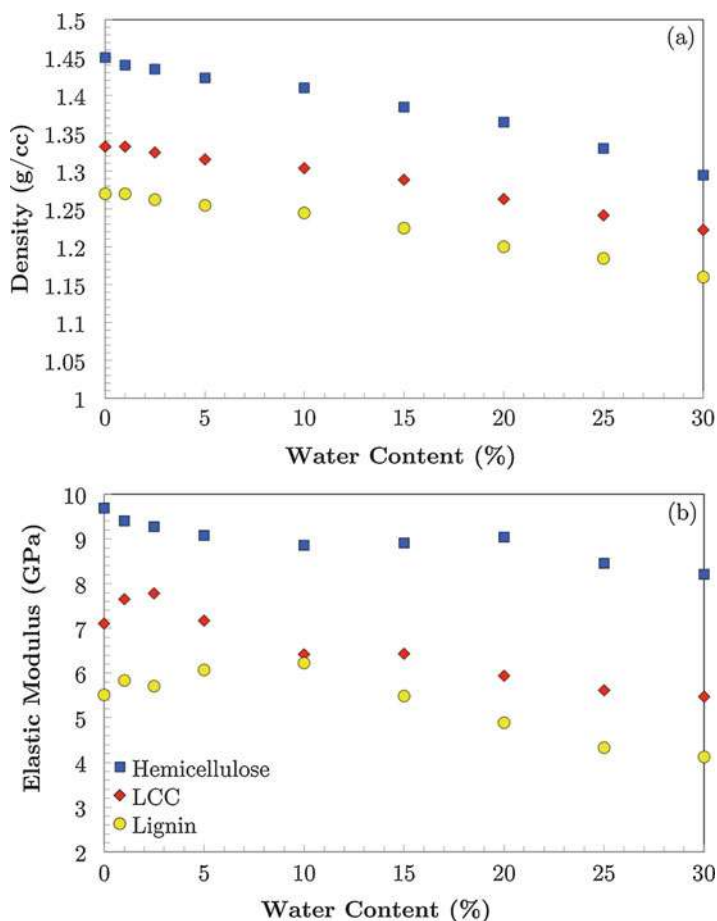
Fig. 6 Variation of density and elastic modulus of hemicellulose (blue), lignin (yellow), and LCC (red) with increasing MC using molecular dynamics simulations. Hemicellulose density constantly decreases with growing MC, whereas LCC and lignin are resistant to the variation of density at very low MC. Hemicellulose elastic modulus constantly decreases with increasing MC, whereas LCC and lignin elastic moduli increase at low MC and then decrease (Youssefian et al. 2017)

amount of water (less than 2.5%) into dry LCC and lignin does not change their densities significantly, whereas hemicellulose density decreases notably. As the water molecules continue permeating into these polymers, their densities decrease almost linearly. The elastic modulus of hemicellulose drops almost linearly with increasing water content, whereas lignin and LCC elastic moduli increase at low water content (less than 10%) and decrease afterward. The peaks in elastic moduli are at 2.5% and 10% MC for LCC and lignin, respectively. When water molecules permeate into the polymer, they enhance the elastic modulus of the material. Once the number of water molecules inside the material reaches a certain level, the effect of water molecules becomes regressive, and the modulus of elasticity starts to decrease.

Comparison between the shapes of the moisture-dependent elastic moduli from the nanoindentation experiments and simulation results reveals interesting similarities. The transverse nanoindentation elastic moduli have a similar shape to the hemicellulose simulations, and the longitudinal nanoindentation elastic moduli have similar shape to the LCC and lignin simulations. This suggests that the nanostructure of fiber is organized in such a way that when compressed in the direction perpendicular to the cellulose microfibrils, hemicellulose, or possibly the similarly structured amorphous cellulose, has more influence on the elastic response than lignin. Conversely, in the direction parallel to the cellulose microfibrils, the lignin has a larger effect than the hemicellulose on the elastic response.

The intermolecular energies between particles in a material play an important role in determining its molecular structure that intimately dictates the physical properties such as density and elastic modulus. Since hydrogen bond interactions are both short range and angle-dependent, their impacts on structuring of polymers are significant. In hemicellulose and lignin, due to their high number of hydroxyl groups, the consequences of hydrogen bond energies are amplified, especially when interacting with water molecules. Studying variation of hydrogen bonds with water content allows us to capture important physics of water-polymer interactions with a minimal need for experimental inputs. Hence, we start with investigating the total hydrogen bond energies of hemicellulose, LCC, and lignin in different MC. All three materials exhibit almost the same progressive trend of total hydrogen bond energies when the water content increases. Nevertheless, hemicellulose shows higher total hydrogen bond energy than LCC and lignin due to the larger number of hydroxyl groups.

In an effort to investigate the conformational change of the hemicellulose, LCC, and lignin with increasing MC, the hydrogen bond energies between merely polymer chains are calculated. At low MC, the diminution of hydrogen bond energies indicates that hydrogen bonds are being broken by water molecules. Higher than 10% MC, the lignin and LCC curves go to almost a plateau, indicating a saturation point for hydrogen bond rupture process. Higher than 10% MC, the growth of hydrogen bond energies between water molecules increases significantly. This suggests that water molecules tend to create hydrogen bond with other water molecules and aggregate at high water content. The salient growing trend of this graph below 10% MC shows that the water molecules tend to form hydrogen bonds

with polymer chains at low MC. Taken all together, the results suggest that at low MC the hydrogen bonds between the polymer chains and water molecules seem to be more favorable than water-water hydrogen bond interactions. At higher MC, however, water molecules tend to make hydrogen bond with other water molecules and form nano-droplets inside the materials.

In order to analyze the evolution of hydrogen bond characteristics with increasing water molecules, radial distribution functions (RDFs) between water molecules and hydroxyl groups of hemicellulose, LCC, and lignin were calculated. The observed variations in dihedral conformation and internal structural variables are the results of the dynamics of nonbonded interactions like hydrogen bonding, electrostatic, and van der Waals interactions. The results suggest as the MC increases, water molecules tend to form hydrogen bonds with other water molecules rather than the polymer chains. Nonetheless, there is a significant difference between hemicellulose and lignin RDFs. Unlike hemicellulose, lignin peaks drop suddenly after 10% water content. This suggests that lignin nanostructure is saturated after 10% MC. LCC also follows the lignin behavior.

The other difference between lignin and hemicellulose response to water molecules is the height of the peaks. The higher peaks on lignin diagram indicate higher probability of finding the water molecules around the reaction site on lignin chains than that of hemicellulose. Therefore, lignin tends to interact with a larger fraction of introduced water molecules than hemicellulose. This is due to the position of functional groups adjacent to the hydroxyl groups. The lignin hydroxyl groups are mostly parts of well-spread hydroxymethyl groups that are extended out from the main chains. The hemicellulose hydroxyl groups, however, are parts of xylose groups, where two hydroxyl groups attach directly to the main chains and are localized on pyranose rings. Lignin hydroxyl groups are more exposed to and more accessible by water molecules than those of hemicellulose hydroxyl groups. Hence, although the hemicellulose structure contains more hydroxyl groups, lignin hydroxyl groups are more efficient in making hydrogen bonds.

Studying the fractional free volumes (FFVs) of a material can bring a deeper insight into understanding the nanostructure and the transport behavior of nonporous amorphous polymers. FFV can be defined as the ratio of the specific free volume (v_f) to the total specific volume (v_T). v_f can be calculated by subtracting the occupied volume (v_{OC}) from the total volume, $v_f = v_T - v_{OC}$. Therefore, the FFV can be calculated from, $FFV = 1 - v_{OC}/v_T$.

To calculate fractional free volume of lignin, hemicellulose, and LCC, Bondi has been used to compute the occupied volumes by the particles. Bondi estimated the occupied volume from the packing density of molecular crystals at $0\,K$ to be approximately 1.3 times the van der Waals volume (v_W) of the repeat unit of the polymer which can be reliably estimated from group contributions (Thran et al. 1999; Bondi 1966). Here, we have used a method developed by Zhao et al. for fast calculation of van der Waals volume as a sum of atomic and bond contributions (Zhao et al. 2003). In this method, which is termed as Atomic and Bond Contributions of van der Waals volume (VABC), the only information needed

for calculating the v_W is the atomic contributions and the number of atoms, bonds, and rings as follows:

$$v_W = 14.71N_O + 20.58N_C + 7.24N_H - 5.92N_{\text{Bonds}}$$
$$- 14.7N_{\text{ARings}} - 3.8N_{\text{NARings}} \quad (5)$$

where N_O, N_C, N_H, N_{Bonds}, N_{ARings}, and N_{NARings} are numbers of oxygen, carbon, hydrogen, bonds, aromatic rings, and nonaromatic rings, respectively. The oxygen, carbon, and hydrogen atoms have been assumed to be represented by hard spheres with a radius derived from the van der Waals radius of respective atoms ($r_C = 1.7\,\text{Å}$, $r_H = 1.2\,\text{Å}$, $r_O = 1.55\,\text{Å}$) (Batsanov 2001).

The number of bonds can be simply calculated by,

$$N_{\text{Bonds}} = N - 1 + N_{\text{ARings}} + N_{\text{NARings}} \quad (6)$$

where N (the total number of atoms) is the summation of all atoms in the system. Therefore, Eq. 5 can be modified to,

$$v_W = 8.79N_O + 14.93N_C + 1.32N_H - 20.62N_{\text{ARings}}$$
$$- 9.72N_{\text{NARings}} + 5.92. \quad (7)$$

Although VABC approach has proven to provide useful correlative and even predictive capabilities, its capabilities have not been fully proven for organic compounds. Since in our models water molecules have a major contribution in the expansion of the nanostructure, we have employed a more accurate and more complicated theory for computing the occupied volume by Connolly (1985). Connolly method analytically calculates the solvent-excluded volume of a molecule by a direct determination of the volume enclosed by the solvent accessible surface. In this method, the molecules are modeled as a static collection of hard spheres (with van der Waals radii) which completely exclude a spherical probe representing a solvent molecule. As the probe rolls over the van der Waals surface, it creates a Connolly surface that is a combination of three kinds of surfaces: convex, saddle, and concave. The total solvent-excluded volume of the molecule is the summation of all these surface pieces. Here, the volume of each polymer was calculated with Connolly method using a probe radius of $0\,\text{Å}$ which returns the van der Waals volume of carbohydrates.

Figure 7 shows the variation of FFV with MC for lignin, LCC, and hemicellulose, calculated by VABC and Connolly methods. Although at higher MCs, VABC estimates FFV at higher values than Connolly, both approaches show the same trend for all three materials. The lignin FFV in dry condition is about 14%. Introducing water into lignin decreases the free volume down to 13% at 2.5% MC, followed by a linear increase to 20% around 30% MC. Hemicellulose, however, exhibits different
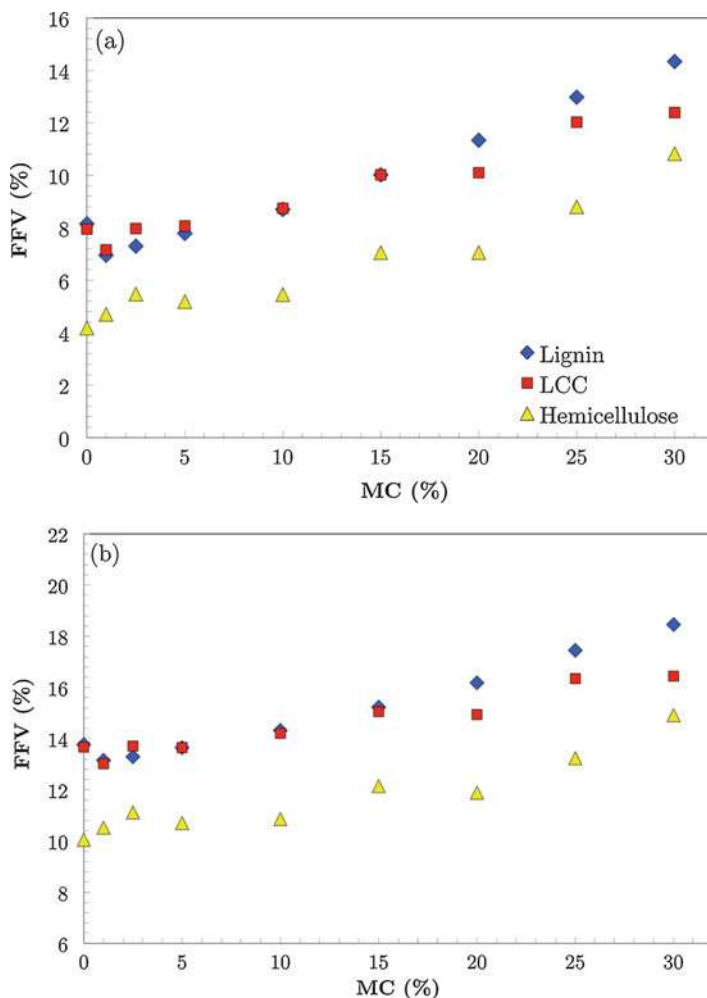
Fig. 7 Variation of fractional free volume (FFV) of hemicellulose, lignin, and LCC calculated by (**a**) VABC, (**b**) Connolly method using the results of the molecular dynamics simulations. Lignin and LCC have more free space for water molecules to fit in. As the water molecules are introduced to the system, they fit into the free volume between polymer chains and reduce FFV at low MC. Hemicellulose has denser structure. Adding water molecules to this system expands the volume and increases the portion of free volume, even at low MC (Youssefian et al. 2017)

response to increasing water molecules. Dry hemicellulose has an FFV around 10%, indicating a denser structure than lignin and LCC. Introducing water into the hemicellulose structure increases the portion of free volume to 12% rapidly followed by a linear increase up to 17% at 30% MC. The FFV value of LCC drops to 13% once the water is introduced into the system and then increases linearly with growing MC. This behavior seems to be influenced more by lignin than hemicellulose.

The different trends of variation of lignin and hemicellulose FFV with increasing MC indicate different influence of water molecules on the nanostructure of these materials. Hemicellulose with a dense hydrogen bond network has denser molecular structure than lignin. Upon the permeation of water molecules into hemicellulose molecular structure, free volume of this material increases because the system expands rapidly to create enough space for water molecules to fit in. Lignin with lower number of hydroxyl groups forms a different molecular structure with more free space available for water molecules to fit to. At low MCs, water molecules take spaces between the lignin polymer chains and decrease the free volume of this material. After 10% MC, however, the nano-droplets are formed in lignin, and free volume linearly increases with increasing MC.

5 Summary

Bamboo, a grass, with stiffness-to-weight and strength-to-weight ratios higher than steel and concrete is a wonder of nature. In this chapter we presented a comprehensive analysis of the mechanical properties of bamboo along with toughening mechanisms and models. We show at the molecular level that the bamboo microfibrils that are the main source of strength and stiffness. The microfibril nanostructure is composed of crystalline and amorphous cellulose in the matrix of lignin-carbohydrate complex (LCC). In this nanocomposite, hemicellulose provides the stiffness, while lignin binds strongly to cellulose faces and provides the strength. We also show that the role of humidity is significant in bamboo microfibrils. Addition of large volume of water (more than 10% water content) reduces the overall density and stiffness. However, addition of 1% water molecules to the system does not change density at nanoscale and increases the elastic modulus of in the longitudinal direction. We have shown this is probably due to the relatively larger fractional free volume in the structure of lignin, and subsequently LCC, at low moisture contents. This universal phenomenon is important in understanding the mechanics of these wood-based materials. It must be noted that more accurate force field can indeed improve the quality of these prediction and understanding of the physics of the complex bamboo system. Hence, development of new reliable force field that can more accurately capture the dynamics of cellulose, hemicellulose, and lignin especially at finite temperature can certainly advance this frontier in this field.

References

Adler E (1977) Lignin chemistry–past, present and future. Wood Sci Technol 11:169–218
Amada S, Untao S (2001) Fracture properties of bamboo. Compos Part B Eng 32(5):451–459
Amada S, Ichikawa Y, Munekata T, Nagase Y, Shimizu H (1997) Fiber texture and mechanical graded structure of bamboo. Compos Part B Eng 28(1):13–20

Askarinejad S, Kotowski P, Shalchy F, Rahbar N (2015) Effects of humidity on shear behavior of bamboo. Theor Appl Mech Lett 5(6):236–243

Bai XS, Lee AWC, Thompson LL, Rosowsky DV (1999) Finite element analysis of moso bamboo-reinforced southern pine OSB composite team. Wood Fiber Sci 31:403

Basu P (2007) Microscopes made from bamboo bring biology into focus. Nat Med 13:1128

Batsanov SS (2001) Van der Waals radii of elements. Inorg Mater 37:871–885

Bergander A, Salmen LJ (2002) Cell wall properties and their effects on the mechanical properties of fibers. Mater Sci 37:151–156

Bertinetti L, Hangen UD, Eder M, Leibner P, Fratzl P, Zlotnikov I (2015) Characterizing moisture-dependent mechanical properties of organic materials: humidity-controlled static and dynamic nanoindentation of wood cell walls. Philos Mag 95:1992–1998

Besombes S, Mazeau K (2005a) The cellulose/lignin assembly assessed by molecular modeling. Part 1: adsorption of a threo guaiacyl β-O-4 dimer onto a Iβ cellulose whisker. Plant Physiol Biochem 43:299–308

Besombes S, Mazeau K (2005b) The cellulose/lignin assembly assessed by molecular modeling. Part 2: seeking for evidence of organization of lignin molecules at the interface with cellulose. Plant Physiol Biochem 43:277–286

Bloyer DR, Rao KTV, Ritchie RO (1998) Fracture toughness and R-curve behavior of laminated brittle-matrix composites. Metall Mater Trans A 29A(10):2483–2496

Bloyer DR, Rao KTV, Ritchie RO (1999) Fatigue-crack propagation behavior of ductile/brittle laminated composites. Metall Mater Trans A 30A(3):633–642

Bondi A (1966) Van der Waals volumes and radii of metals in covalent compounds. J Phys Chem 70:3006–3007

Brauns F (1962) Soluble native lignin, milled wood lignin, synthetic lignin, and the structure of lignin. Holzforschung Int J Biol Chem Phys Technol Wood 16:97–102

Budiansky B, Amazigo JC, Evans AG (1988) Small-scale crack bridging and the fracture toughness of particulate-reinforced ceramics. J Mech Phys Solids 36(2):167–187

Buehler M (2008) Atomistic modeling of materials failure. Springer Science & Business Media, Boston

Chang ST, Wang SY, Wu JH (1998) Rapid extraction of epidermis chlorophyll of moso bamboo (Phyllostachys pubescens) culm using ultrasonics. J Wood Sci 44(1):78–80

Chung KF, Yu WK, Chan SL (2002) Mechanical properties and engineering data of structural bamboo for bamboo scaffolding. Eng Struct 24:429–442

Chung MJ, Cheng SS, Lee CJ, Chang ST (2014) Novel methods for dyeing the epidermis of bamboo culms and their colour fastness. Color Technol 130(2):112–119

Ciesielski PN, Matthews JF, Tucker MP, Beckham GT, Crowley MF, Himmel ME, Donohoe BS (2013) 3D electron tomography of pretreated biomass informs atomic modeling of cellulose microfibrils. ACS Nano 7(9):8011–8019

Connolly ML (1985) Computation of molecular volume. J Am Chem Soc 107:1118–1124

Cousins W (1976) Elastic modulus of lignin as related to moisture content. Wood Sci Technol 10:9–17

Cousins WJ (1977) Elasticity of isolated lignin: Young's modulus by a continuous indentation method. N Z J For Sci 7:107–112

Cousins W (1978) Young's modulus of hemicellulose as related to moisture content. Wood Sci Technol 12:161–167

Diddens I, Murphy B, Krisch M, Muller M (2008) Anisotropic elastic properties of cellulose measured using inelastic X-ray scattering. Macromolecules 41:9755–9759

Ding SY, Himmel ME (2006) The maize primary cell wall microfibril: a new model derived from direct visualization. J Agric Food Chem 54:597–606

Dixon PG, Gibson LJ (2014) The structure and mechanics of Moso bamboo material. J R Soc Interface 11(99):20140321

Dri F, Hector GJ, Moon R, Zavattieri P (2013) Anisotropy of the elastic properties of crystalline cellulose from first principles density functional theory with van der Waals interactions. Cellulose 20:2703–2718

Eichhorn SJ, Young RJ, Davies GR (2005) Modeling crystal and molecular deformation in regenerated cellulose fibers. Biomacromolecules 6(1):507–513

Fernandes AN, Thomas LH, Altaner CM, Callow P, Forsyth VT, Apperley DC, Jarvis MC (2011) Nanostructure of cellulose microfibrils in spruce wood. Proc Natl Acad Sci 108(47):E1195–E1203

Fett T, Munz D (1994) Stress intensity factors and weight functions for one dimensional cracks, Report KfK 5290. Karlsruhe: Institut fur material forschung

Forss K, Fremer KE, Stenlund B (1966) Spruce lignin and its reactions in sulfite cooking. I. Structure of lignin. Pap Puu-pap Och Tra 48:565–574

Geymayer HG, Cox FB (1970) Bamboo reinforced concrete. ACI J Proc 67(10):841–846

Ghavami K (1995) Ultimate load behavior of bamboo-reinforced lightweight concrete beams. Cem Concr Compos 17:281

Ghavami K (2005) Bamboo as reinforcement in structural concrete elements. Cem Concr Compos 27:637–649

Ghavami K, Rodrigues CS, Paciornik S (2003) Bamboo – functionally graded composite material. Asian J Civil Eng 4:1

Gibson LJ (2012) The hierarchical structure and mechanics of plant materials. J R Soc Interface 9(76):2749–2766

Glasser WG, Glasser HR (1974) Simulation of reactions with lignin by computer (Simrel). II. A model for softwood lignin. Holzforschung Int J Biol Chem Phys Technol Wood 28:5–11

Green DW, Kretschmann DE (1994) Moisture content and the properties of clear southern pine. Res Pap FPL–RP–531, Madison

Habibi MK, Lu Y (2015) Crack propagation in bamboo's hierarchical cellular structure. Sci Rep 4:5598

Habibi MK, Samaei AT, Gheshlaghi B, Lu J, Lu Y (2015) Asymmetric flexural behavior from bamboo's functionally graded hierarchical structure: underlying mechanisms. Acta Biomater 16:178–186

He XQ, Suzuki K, Kitamura S, Lin JX, Cui KM, Itoh T (2002) Toward understanding the different function of two types of parenchyma cells in bamboo culms. Plant Cell Physiol 43(2):186–195

Hosoya T, Kawamoto H, Saka S (2007) Cellulose–hemicellulose and cellulose–lignin interactions in wood pyrolysis at gasification temperature. J Anal Appl Pyrolysis 80:118–125

Huang X, Netravali A (2009) Biodegradable green composites made using bamboo micro/nano-fibrils and chemically modified soy protein resin. Compos Sci Technol 69(7):1009–1015

Iwamoto S, Kai W, Isogai A, Iwata T (2009) Elastic modulus of single cellulose microfibrils from tunicate measured by atomic force microscopy. Biomacromolecules 10:2571–2576

Jager A, Hofstetter K, Buksnowitz C, Gindl-Altmutter W, Konnerth J (2011) Identification of stiffness tensor components of wood cell walls by means of nanoindentation. Compos A Appl Sci Manuf 42:2101–2109

Jain S, Kumar R, Jindal UC (1992) Mechanical behavior of bamboo and bamboo composite. J Mater Sci 27(17):4598–4604

Jakes JE, Hunt CG, Yelle DJ, Lorenz L, Hirth K, Gleber SC, Frihart CR (2015) Synchrotron-based X-ray fluorescence microscopy in conjunction with nanoindentation to study molecular-scale interactions of phenol? Formaldehyde in wood cell walls. ACS Appl Mater Interfaces 7:6584–6589

Janssen JJA (1981) Bamboo in building structures. Doctoral dissertation, Technische Hogeschool Eindhoven

Janssen JJA (1991) Mechanical properties of bamboo. Kluwer Academic Publishers, Dordrecht

Jeffries TW (1990) Biodegradation of lignin-carbohydrate complexes. Biodegradation 1:163

Keogh L, O'Hanlon P, O'Reilly P, Taylor D (2015) Fatigue in Bamboo. Int J Fatigue 75:51–56

Khalil HA, Bhat IUH, Jawaid M, Zaidon A, Hermawan D, Hadi YS (2012) Bamboo fibre reinforced biocomposites: a review. Mater Des 42:353–368

Kulasinski K, Keten S, Churakov SV, Guyer R, Carmeliet J, Derome D (2014a) Molecular mechanism of moisture-induced transition in amorphous cellulose. ACS Macro Lett 3(10):1037–1040

Kulasinski K, Keten S, Churakov SV, Derome D, Carmeliet J (2014b) A comparative molecular dynamics study of crystalline, paracrystalline and amorphous states of cellulose. Cellulose 21:1103–1116

Kulasinski K, Guyer R, Derome D, Carmeliet J (2015) Water adsorption in wood microfibril-hemicellulose system: role of the crystalline-amorphous interface. Biomacromolecules 16:2972–2978

Lahiji RR et al (2010) Atomic force microscopy characterization of cellulose nanocrystals. Langmuir 26:4480–4488

Lakkad SC, Patel JM (1980) Mechanical properties of bamboo, a natural composite. Fibre Sci Tech 14:319

Lee AWC, Bai XS, Bangi AP (1997) Flexural properties of bamboo-reinforced southern pine OSB beams. Forest Prod J 47:74

Li LJ, Wang YP, Wang G, Cheng HT, Han XJ (2010) Evaluation of properties of natural bamboo fiber for application in summer textiles. J Fiber Bioeng Inform 3(2):94–99

Lindner B, Petridis L, Schulz R, Smith JC (2013) Solvent-driven preferential association of lignin with regions of crystalline cellulose in molecular dynamics simulation. Biomacromolecules 14:3390–3398

Lo TY, Cui HZ, Leung HC (2004) The effect of fiber density on strength capacity of bamboo. Mater Lett 58:2595

Lo TY, Cui HZ, Tang PWC, Leu HC (2008) The effect of fiber density on strength capacity of bamboo. Construct Build Mater 22(7):1532–1535

MacKerell AD Jr et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616

Maekawa E (1976) Studies on hemicellulose of bamboo. Wood Res-Slovakia 59:153–179

Malherbe S, Cloete T (2002) Lignocellulose biodegradation: fundamentals and applications. Rev Environ Sci Biotechnol 1:105–114

Mannan S, Paul Knox J, Basu S (2017) Correlations between axial stiffness and microstructure of a species of bamboo. R Soc Open Sci 4:160412

Mayo SL, Olafson BD, Goddard WA (1990) DREIDING: a generic force field for molecular simulations. J Phys Chem 94:8897–8909

Miyamoto H, Umemura M, Aoyagi T, Yamane C, Ueda K, Takahashi K (2009) Structural reorganization of molecular sheets derived from cellulose II by molecular dynamics simulations. Carbohydr Res 344(9):1085–1094

Moon RJ, Martini A, Nairn J, Simonsen J, Youngblood J (2011) Cellulose nanomaterials review: structure, properties and nanocomposites. Chem Soc Rev 40(7):3941–3994

Nimz H (1974) Beech lignin – proposal of a constitutional scheme. Angew Chem Int Ed Engl 13:313–321

Nogata F, Takahashi H (1995) Intelligent functionally graded material: bamboo. Compos Eng 5(7):743–751

Oliver WC, Pharr GM (2004) Measurement of hardness and elastic modulus by instrumented indentation: advances in understanding and refinements to methodology. J Mater Res 19:3–20

Osorio L, Trujillo E, Van Vuure AW, Verpoest I (2011) Morphological aspects and mechanical properties of single bamboo fibres and flexural characterization of bamboo/epoxy composites. J Reinf Plast Compos 30:0731684410397683

Pakzad A, Yassar R (2013) New frontiers of nanoparticles and nanocomposite materials. Adv Struct Mater 4:233–263

Phong NT, Fujii T, Chuong B, Okubo K (2011) Study on how to effectively extract bamboo fibers from raw bamboo and wastewater treatment. J Mater Sci Res 1(1):144

Pu Y, Zhang D, Singh PM, Ragauskas AJ (2008) The new forestry biofuels sector. Biofuels Bioprod Biorefin 2:58–73

Rao KMM, Rao KM (2007) Extraction and tensile properties of natural fibers: vakka, date and bamboo. Compos Struct 77(3):288–295

Ray AK, Das SK, Mondal S, Ramachandrarao P (2004) Microstructural characterization of bamboo. J Mater Sci 39:1055

Sakakibara A (1980) A structural model of softwood lignin. Wood Sci Technol 14:89–100

Salmen L (2004) Micromechanical understanding of the cell-wall structure. C R Biol 327:873–880

Sathitsuksanoh N, Zhu Z, Ho TJ, Bai MD, Zhang YHP (2010) Special issue on lignocellulosic bioethanol: current status and perspectives. Bioresour Technol 101:4926–4929

Scheller HV, Ulvskov P (2010) Hemicelluloses. Plant Biol 61:263–288

Shao ZP, Fang CH, Tian GL (2009) Mode I interlaminar fracture property of moso bamboo (Phyllostachys pubescens). Wood Sci Technol 43:527–36

Shi Z et al (2013) Isolation and structural exploration of hemicelluloses from the largest bamboo species: dendrocalamus sinicus. Bioresources 8:5036–5050

Silva ECN, Walters MC, Paulino GH (2006) Modeling bamboo as a functionally graded material: lessons for the analysis of affordable materials. J Mater Sci 41(21):6991–7004

Sinko R, Mishra S, Ruiz L, Brandis N, Keten S (2013) Dimensions of biological cellulose nanocrystals maximize fracture strength. ACS Macro Lett 3:64–69

Sturcov A, Davies GR, Eichhorn SJ (2005) Elastic modulus and stress-transfer properties of tunicate cellulose whiskers. Biomacromolecules 6(2):1055–1061

Sun H (1998) COMPASS: an ab initio force-field optimized for condensed-phase applications: overview with details on alkane and benzene compounds. J Phys Chem B 102:7338–7364

Sun RC, Fang J, Tomkinson J (2000) Delignification of rye straw using hydrogen peroxide. Ind Crop Prod 12:71–83

Suresh S, Mortensen A (1988) Fundamentals of functionally graded materials. In: Processing and thermomechanical behavior of graded metals and metal-ceramic composites, IOM Communications, London

Tan T, Rahbar N, Allameh SM, Kwofie S, Dissmore D, Ghavami K, Soboyejo WO (2011) Mechanical properties of functionally graded hierarchical bamboo structures. Acta Biomater 7(10):3796–3803

Taylor D, Kinane B, Sweeney C, Sweetnam D, O'Reilly P, Duan K (2015) The biomechanics of bamboo: investigating the role of the nodes. Wood Sci Technol 49:345–357

Thran A, Kroll G, Faupel F (1999) Correlation between fractional free volume and diffusivity of gas molecules in glassy polymers. J Polym Sci B Polym Phys 37:3344–3358

Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Curr Opin Plant Biol 11:278–285

Wang SG, Lin SY, Pu XL, Ding YL, Wan XC (2012b) Developmental changes in cell wall of bundle sheath fibers close to phloem of Fargesia yunnanensis. Bot Stud 53(3):353–362

Wang X, Ren H, Zhang B, Fei B, Burgert I (2012a) Cell wall structure and formation of maturing fibres of moso bamboo (Phyllostachys pubescens) increase buckling resistance. J R Soc Interface 9(70):988–996

Watching Grass Grow The Economist (2018) Innovative materials from bamboo are helping a new industry to sprout.

Wegst UGK (2011) Bending efficiency through property gradients in bamboo, palm, and wood-based composites. J Mech Behav Biomed Mater 4(5):744–755

Wegst UGK, Ashby MF (2004) The mechanical efficiency of natural materials. Philos Mag 84(21):2167–2186

Wegst UG, Bai H, Saiz E, Tomsia AP, Ritchie RO (2014) Bioinspired structural materials. Nat Mater 14:23–36

Xiao B, Sun X, Sun R (2001) Chemical, structural, and thermal characterizations of alkali-soluble lignins and hemicelluloses, and cellulose from maize stems, rye straw, and rice straw. Polym Degrad Stab 74:307–319

Xiao Y, Zhou Q, Shan B (2010) Design and construction of modern bamboo bridges. J Bridge Eng 15(5):533–541

Youssefian S, Rahbar N (2015) Molecular origin of strength and stiffness in bamboo fibrils. Sci Rep 5:11116

Youssefian S, Jakes JE, Rahbar N (2017) Variation of nanostructures, molecular interactions, and anisotropic elastic moduli of lignocellulosic cell walls with moisture. Sci Rep 7:2054

Yu D, Tan H, Ruan Y (2011a) A future bamboo-structure residential building prototype in China: life cycle assessment of energy use and carbon emission. Energ Buildings 43(10):2638–2646

Yu Y, Fei B, Wang H, Tian G (2011b) Longitudinal mechanical properties of cell wall of Masson pine (Pinus massoniana Lamb) as related to moisture content: a nanoindentation study. Holzforschung 65:121–126

Yueping W, Ge W, Haitao C, Genlin T, Zheng L, QunFeng X, Xiangqi Z, Xiaojun H, Xushan G (2009) Structures of natural bamboo fiber for textiles. Text Res J 80:334–343

Zachariah EJ, Sabulal B, Nair DNK, Johnson AJ, Kumar CSP (2016) Carbon dioxide emission from bamboo culms. Plant Biol 18:400–405

Zhao YH, Abraham MH, Zissimos AM (2003) Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds. J Org Chem 68:7368–7373

Zou L, Jin H, Lu WY, Li X (2009) Nanoscale structural and mechanical characterization of the cell wall of bamboo fibers. Mater Sci Eng C 29:1375–1379

Multiscale Modeling of Lignocellulosic Biomass

66

Francisco J. Martin-Martinez, Jose A. Dobado,
Angel Sanchez-Gonzalez, Diego López Barreiro, and
Markus J. Buehler

Contents

Abstract

Lignocellulosic biomass is a key feedstock for the development of future sustainable technologies that utilize the potential of biomass as a source of carbon-based

F. J. Martin-Martinez (✉) · M. J. Buehler
Laboratory for Atomistic and Molecular Mechanics (LAMM), Department of Civil and
Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: fmartinm@mit.edu; mbuehler@mit.edu

D. L. Barreiro
Laboratory for Atomistic and Molecular Mechanics Civil and Environmental Engineering,
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: dlopezb@mit.edu

J. A. Dobado · A. Sanchez-Gonzalez
Facultad de Ciencias, Departamento de Química Orgánica, Universidad de Granada, Granada,
Spain
e-mail: dobado@ugr.es; angels@ugr.es

materials. Lignocellulosic biomass is the main component of cell walls in plants, and it is basically composed by cellulose, hemicellulose, and lignin. Among the three, lignin is of special interest, since it is the second most abundant natural polymer after cellulose and the main natural source of aromatic compounds. Effective lignin valorization is therefore a key challenge in the development of integrated biorefineries, and many strategies are being researched to successfully utilize it. Within this framework, this chapter presents a revision of current state of the art in multiscale modeling of lignocellulosic biomass (especially lignin), as well as an overview of molecular simulations that have helped understanding lignocellulosic biomass structure to develop new depolymerization strategies to enable the production of biomass-derived materials.

1   Introduction

Lignocellulosic biomass is an essential natural source for producing sustainable chemicals, materials, and fuels in a circular economy. Unlocking the potential of such resource requires however to address the fundamental challenge of economically degrading biomass into its basic constituents, i.e., cellulose, hemicellulose, and lignin. Once this is achieved, cellulose and hemicellulose can be decomposed into sugars for further processing and fermentation, while lignin arises as a raw material for the production of aromatic chemicals and biomass-derived materials. The current state of the art of biomass technology points to the valorization of lignin as the bottleneck for the successful development of integrated biorefineries. Therefore, many efforts are focused on developing sustainable lignin-related technologies, and accordingly lignin is the main goal of the current chapter.

   Lignin is the second most abundant polymer provided by nature after cellulose and the most abundant natural material composed mostly of aromatic moieties. It is a highly heterogeneous biopolymer, mainly constituted by three types of aromatic monomers known as monolignols: hydroxyphenyl (H), guaiacyl (G), and syringyl (S). These monolignols generate the 3D amorphous lignin structure by means of oxidative coupling reactions (Beckham et al. 2016). Monolignols derive from the bioconversion of L-phenylalanine through a series of conversion steps that involve multiple enzymes. The final result is a hard and glassy material at room temperature, which becomes soft above its glass transition temperature (80–100 °C) (Petridis et al. 2011b). It is very recalcitrant, with several vital roles in lignocellulosic biomass, such as providing mechanical strength and rigidity, participating in the water conduction system, or as defensive barrier against microbial, enzymatic, and chemical degradation (Sangha et al. 2012b). Lignin is usually obtained as a waste product from the pulp and paper industry, and with the expected advent of lignocellulosic biorefineries, the amount of waste lignin available is expected to increase dramatically. Due to the unique chemical properties of lignin, mainly devoted to its aromatic backbone, much interest has been placed on the production of materials from it. Some potential high-value lignin-derived

products are low-cost carbon fibers, plastics, composites, additives for cement, dust suppression agents, drilling fluids for oil recovery, polymeric foams, phenolic resins, biodispersants, surfactants, and membranes or aromatic chemicals (Calvo-Flores et al. 2015).

Understanding lignin structure and composition is crucial for deciphering biosynthetic pathways, producing plants with more easily degradable cell walls, and developing better technologies for depolymerization (see Fig. 1). However, decomposing lignocellulosic biomass into its basic components to produce such a wide range of materials requires a deep fundamental understanding of biomass structure, as well as polymerization and degradation mechanisms. Furthermore, isolating and characterizing lignin is one of the major challenges in current biomass technology, and therefore, complementary theoretical approaches can greatly assist to reach beyond experimental limitations. Modeling and simulations at different lengths from nano- to macroscale have helped indeed the development of depolymerization strategies to produce lignin-derived materials. Multiscale modeling can be used to predict the physicochemical properties of lignocellulosic biomass, as well as its components, including lignin. It has helped to provide insight into lignin structure, reactivity, and intermolecular interactions with lignocellulosic biomass components, to shed light into better degradation processes (Sangha et al. 2012b).

The choice of an appropriate computational method always depends on the specific molecular system under study, as well as on those questions to be answered. For dynamic processes that imply relatively large molecular systems and long (in a molecular context) time scales, molecular dynamics (MD) simulations can provide useful information from nanoscopic to mesoscopic length scale. When the study implies reactivity, reaction mechanisms, or electronic properties, quantum mechanical (QM) calculations are required, and the computational cost significantly increases. When it comes to molecular modeling and simulations, there is always an intrinsic trade-off between the desired level of accuracy and the computational cost of the calculations. This usually limits the size of the model system under study, and it is, in the end, the main criteria for selecting a computational method. A scheme summarizing the different methods that are used in lignocellulose multiscale modeling, along different time and length scales, is shown in Fig. 2.

The following sections provide an overall view of the different studies that have contributed to a better understanding of lignocellulose structure, properties, and degradation. Furthermore, it covers extensively how these methods, and the results derived from them, are helping the development of degradation strategies for the production of materials from lignocellulosic biomass. We cover both QM calculations and MD simulations of lignocellulosic biomass, with special focus on lignin, since we understand that this material is the cornerstone that could unlock the development of future integrated biorefineries. In conjunction with multiscale modeling techniques, we also review the experimental approaches that are currently applied and investigated to degrade biomass, and lignin, to its fundamental building blocks.
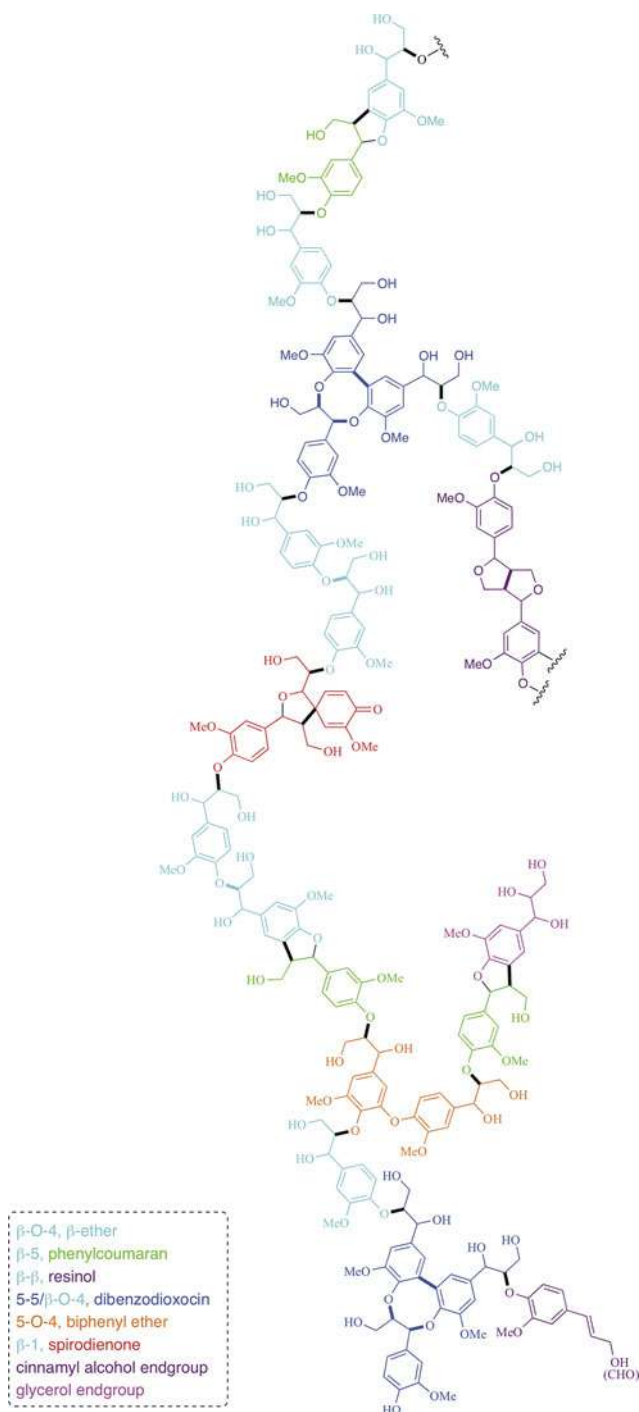
Fig. 1 Standard chemical model of lignin with different structural motifs (colored)
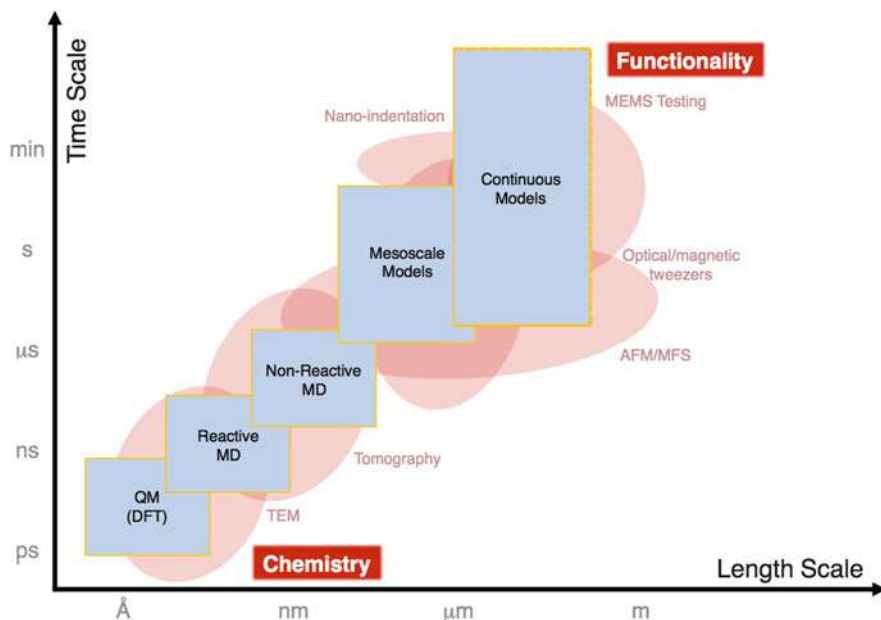
Fig. 2 Overview of the available modeling methodologies for lignocellulosic biomass and examples of some of the processes that have been studied with them

2   Ab Initio and Computational Chemistry Modeling of Fundamental Building Blocks of Lignocellulose

In the case of lignocellulosic biomass, as it happens with any materials system under study, large part of the challenge when performing computational chemistry calculations comes from the selection of the method that provides accurate results in a reasonable amount of time. Computational chemistry is much about adequated approximations, and that also includes modeling biomass components. As many of us already know, ab initio methods are the most accurate, but also the most computationally demanding approaches. In this regard, it is commonly accepted nowadays that density functional theory (DFT) is almost as accurate yet less computationally demanding than purely ab initio methods to calculate the electronic structure of molecules and materials. In addition, semiempirical QM methods also provide an efficient approach for calculating molecular structures and their properties. Below, we review some of these DFT approaches that have been undertaken to understand the structure of lignocellulosic biomass and especially of lignin.

Within DFT, B3LYP hybrid functional has been the most widely used across all areas of computational chemistry, including lignocellulosic biomass (Mostaghni et al. 2013), although a relatively new generation of DFT methods with that

improve accuracy and dispersion corrections are nowadays taking over B3LYP's overwhelming domain, e.g., ωB97X-D or Minnesota density functionals.

Within this context, QM calculations have been used to investigate the structure, electronic properties, and reactivity of monolignols, dehydrodilignols, and other lignin-like compounds. Back in 1996, Garver (1996) applied semiempirical MO methods to calculate minimum energy structures and energy barriers for bond rotations of lignin models represented by trimer acetoguaiacyl-dehydro-diisoeugenol methyl ether. This was complemented by Dos Santos (2001) in a study that used semiempirical methods to perform a conformational analysis of similar lignin model systems. Later in 2000, Russell et al. (2000) studied lignin formation mechanism by comparing the properties of lignin to synthetically produced polymers. In a similar direction, Durbeej and Eriksson (2005b) study dehydrogenated dilignols. They reported spin populations to characterize the electronic distribution of unpaired electrons and to correlate such distribution with the observed linkage ratios. Their results clearly pointed out a prevalence of the radical in the phenolic oxygen, partially explaining the prevalence of β–O–4 linkage in nature. In later works (Durbeej and Eriksson 2003, 2005a), the authors described the geometry and stability of dilignols considering some frequent linkages, β–O–4, β–5, β–β, β–1, 5–5, and 5–O–4, highlighting the intramolecular hydrogen bonding as a key factor for stabilizing the different conformers. These studies were the first attempts to describe the reaction mechanism for dilignols formation. Similarly, Parthasarathi et al. (2011) analyzed the main dissociation tendencies of a diverse set of lignin linkages encompassing 65 lignin model compounds. More recently, radical coupling reactions for different lignin linkages were investigated in an extensive computational study performed by Sangha et al. (2012a). Natural bond orbital (NBO) analysis was used to quantify the spin densities, and this was completed by providing reaction enthalpies of monolignol dimerization for the main linkages in lignin, β–O–4, β–5, β–β, β–1, 5–5, and 5–O–4. Additional studies implemented Monte Carlo conformational search to find the lowest energy conformers for each dilignol, including solvent effects. These results showed that β–O–4, β–5, β–β linkages are the most energetically favorable. In the same direction, Watts et al. (2011) studied lignin polymerization with focus on potential mechanisms that lead to the formation of α–linkages. They also employ B3LYP hybrid functional to calculate thermodynamic data, pointing out that the most probable mechanism for the formation of α–linkages occurs by the formation of α radical monolignols hydroxylated in the C$_{\beta}$– and in the presence of H$_2$O$_2$.

Moreover, taking into account that lignin structures are rich in aromatic functional groups, computational studies where monolignols and dilignols interact with each other require the implementation of electron dynamic correlation to describe π–orbitals. The development of dispersion-corrected functionals enabled further studies in the field. In 2006, there was a valuable advance in DFT methods with the publication of M05-2X hybrid functional that takes into account the noncovalent interactions. Chen and Sarkanen (2010) used the dispersion-corrected M05-2X hybrid functional to study the possible arrangements between a radical monolignol and a lignin model. Their results pointed out that noncovalent forces rule the

polymerization steps, where a co-facial arrangement between the radical monolignol and lignin substructures is preferred due to π–stacking between aromatic groups. The authors also highlight that such noncovalent interactions prevail over hydrogen bonding, substituent interactions, and solvent effects. With focus in the lignin β–O–4 linkage bond, Lim et al. (2013) studied model compounds to estimate the efficiencies of single electron transfer-promoted photochemical and enzymatic C–C bond cleavage reactions. Their results show that the radical cations derived from less alkoxy-substituted dimeric β–O–4 linkage models undergo more rapid C_1–C_2 bond cleavage than those of more alkoxy-substituted analogues. Further insight was also provided by Sánchez-González et al. (2017), by analyzing the formation mechanism of β–O–4 linkages with quantum theory of atoms in molecules (QTAIM) and noncovalent index (NCI) analysis to quantify noncovalent interactions between reactants. They pointed out the importance of hydrogen bonding and π-stacking interactions in the stabilization of transition states that control the polymerization mechanism of lignin.

The interactions among the three main components, cellulose, hemicellulose, and lignin, were studied to examine the hydrogen bonding network of cellulose–hemicellulose systems and the covalent bond linkages of hemicellulose–lignin systems (Zhang et al. 2011). In combination with experimental data from X-ray photoelectron spectroscopy (XPS), Haensel et al. (2012) applied DFT to estimate the valence band structure of cellulose and lignin. They proved valence band analysis supported by DFT as a powerful basis for a detailed interpretation of spectroscopic data. The comparison with experimental, and in particular with sum frequency generation–vibrational spectroscopy (SFG-VS) for crystalline cellulose inside intact lignocellulose, showed that dispersion-corrected DFT-D2 underestimates the distances between hydrogen-bonded oxygen atoms compared to the experimentally determined values (Lee et al. 2013).

As expected by the limitations imposed by the computational cost of the methods, all these studies have been mainly focused on the structure, electronic properties, and reactivity of monolignols, dehydrodilignols, and other small lignin-like model compounds.

3 Molecular Dynamics and Larger-Scale Models of Lignocellulosic Components

As we discussed already, a proper description of lignocellulosic biomass requires a multiscale approach that is able to describe all key fundamental features of such hierarchical material along different length scales. This multiscale description implies not only the abovementioned ab initio and DFT calculations but also larger-scale models that help in understanding the structure of cellulose, hemicellulose, and lignin, as well as their integration into lignocellulosic biomass. Ab initio and DFT methods are computationally too expensive to simulate the behavior of molecular systems in such larger scales, and therefore MD simulations arise here as the most

suitable choice to reach larger model sizes that still retain the details required to achieve the intended atomistic description of fundamental features.

MD simulations of wood have helped in understanding the structure and molecular deformation mechanisms (Jin 2015). In the context of MD simulations of complexes between lignocellulose components, Shevchenko and Bailey (1996) studied the lignin–carbohydrate complex (LCC) of living plants and soil humic materials by considering model linear chains of lignin, in the presence of carbohydrate linear structures. MD simulations also require the selection of models systems, yet larger than those employed in QM calculations. The results pointed out by these authors show that a helical conformation of lignin chains can accommodate the carbohydrate ribbon structure inside, stabilizing the LCC without covalent bonds. A similar work was presented by Houtman and Atalla (1995), where the adsorption of lignin oligomers in a polysaccharide surface matrix was studied. This work pointed that lignin can be adsorbed at the surface of cellulose microfibrils, leading the polymerization of the lignin–polysaccharide complex. Additionally, the nature of the polysaccharide matrix was shown to induce a selective binding with preferred orientations, affecting the architecture of the hemicellulose in the cell wall. In connection with hemicellulose, one of the most useful advances for modeling lignin–hemicellulose complexes was achieved through the development of WallGe, a Fortran program to build virtual cellulose-hemicellulose networks by stochastic self-assembly. This approach also provided a way to predict the mechanical properties finite element analysis (Kha et al. 2010). Within a similar context, a modeling of lignified secondary plant cell walls adapted to grass was applied to systems composed of cellulose, xylan, water, and lignin (Charlier and Mazeau 2012). More specific to xylan, as the most relevant component in hemicellulose, MD simulations help understand the mechanism of thermal decomposition of hemicellulose (Huang et al. 2013).

Despite these studies on LLC, lignin–cellulose and lignin–hemicellulose complexes, stand-alone models for lignin are of special importance due to the relevance of this lignocellulosic component, and therefore, the amount of literature on the topic is especially significant. Some of the first MD studies on lignin were performed by modeling β–O–4-linked dimeric and oligomeric model compounds (Simon and Eriksson 1995). Similar methods were used to characterize the conformational behavior of dilignols formed by the most abundant linkage present in lignin, the β–O–4 linkage.

In 2009, Petridis and Smith (2009) presented a first parametrization of molecular mechanics force field for lignin, based on QM data of model compounds. Partial atomic charges were derived using the electrostatic potential fitting from methoxybenzene–water interactions. Dihedral parameters were optimized by fitting to critical rotational potentials, and bonded parameters were obtained by optimizing vibrational frequencies and normal modes. This parametrization paid special attention to properly describe the β–O–4 linkage by considering the partial atomic charges of the atoms involved in it. This work opened new perspectives for computational simulations of lignocellulosic biomass, since, together with the

existing force field for polysaccharides, it enabled more accurate full simulations of lignocellulose (Petridis and Smith 2009). The same authors used neutron scattering experiments and MD simulations to reveal that lignin aggregates are characterized by a surface fractal dimension that is invariant under change of scale from ∼1 to 1000 Å (Petridis et al. 2011a). This work and a complementary study on the temperature-dependent structure and dynamics of individual softwood lignin polymers in aqueous solution helped to understand the role of lignin as residual that hampers an effective hydrolysis of plant biomass in cellulosic ethanol production (Petridis et al. 2011b). In a similar direction, replica-exchange MD have been used to characterize the globule-to-coil transition for a softwood-like lignin biopolymer in a tetrahydrofuran (THF)–water cosolvent system, at temperatures at which the cosolvent undergoes a de-mixing transition (Smith et al. 2016).

Furthermore, Besombes and Mazeau (2004) used CHARMM force field to study the influence of hydrogen bonding and steric effects of methoxide groups in the conformational characteristics of such linkage and the relative stability of *erythro-* and *threo-* forms. This study was later complemented by the same authors to further analyze the assembly of cellulose and lignin (Besombes and Mazeau 2005a). They consider a molecular model of a cellulose crystal corresponding to the allomorph Iβ. The model was selected to exhibit different surfaces that mimic the carbohydrate matrix present in native wood cell wall. In a second part, they extended their previous computational investigations by considering more complex systems, looking for evidence of organization of lignin molecules at the interface with cellulose (Besombes and Mazeau 2005b). Surface coverage of cellulose, molecular mass, and lignin structural features were some of the parameters taken into account to examine their influence on the associative interactions between cellulose and lignin. An additional step was taken by Beste (2014) by selecting the most common linkages to perform MD simulations with a reactive force field (ReaxFF) to describe lignin degradation into carbon fibers.

Going up in the scale of simulations, it is worth mentioning the implementation of experimental measurement of the intra-ring properties of loblolly pine (*Pinus taeda*) to provide inputs to finite element models using both solid layers and cellular layers to represent longitudinal tracheids. This was used to investigate the effect of considering wood strands as two-layer composites consisting of earlywood and latewood, or intra-ring, layers (Hindman and Lee 2007).

The complexity of the hierarchical structure along different large scales, together with computational limitations, has called for the implementation of new methodologies in the field, as, for example, the use of quantum mechanics/molecular mechanics (QM/MM) that has been proved successful in simulating different kinds of biomolecular systems (Senn and Thiel 2009). Combined QM/MM approaches have become indeed the method of choice for modeling reactions in many biomolecular systems. As it happens with the previously discussed DFT methods, the limitations imposed by the computational cost MD simulations, when it comes to the size the models, have not hampered the production of very useful information to support better degradation strategies as we discuss below.

4 Downstream: Multiscale Modeling to Assist Pyrolysis, Hydrothermal Liquefaction, Depolymerization, and Upgrading

From previous sections on structural QM and MD modeling of lignocellulose and lignin at different length scales, it has been shown that a plethora of approaches have provided significant amount of information on the structure, chemical bonding, and molecular-scale features of lignocellulosic biomass. All this fundamental knowledge has been proved very useful in developing the polymerization processes that we will describe in this part of the chapter. These experimental techniques, largely supported by modeling and simulation, help the actual valorization of those materials that are produced by degrading biomass to simpler building blocks, i.e., downstream processing. This section aims at connecting these ab initio and MD simulations with state-of-the-art experimental approaches for depolymerizing lignocellulosic biomass, as well as providing a quick introduction to those upgrading strategies that follow biomass degradation. It also includes some references to the kinetic modeling that has been done to simulate mass and heat transfer that occur during pyrolysis and hydrothermal processes.

4.1 Valorizing Lignocellulosic Biomass

As we have already pointed out, prior to the valorization of lignocellulosic biomass into any kind of material, it is necessary to break down its structure into its main biochemical fractions: cellulose, hemicellulose, and lignin (see Fig. 3). Effective methods to separate biomass into pure streams of its biochemical constituents are important for the valorization of any lignocellulosic biopolymer. For example, during the enzymatic hydrolysis of cellulose to produce bioethanol, 50% of the enzymes tend to bind nonproductively to lignin, thus decreasing the degree of conversion of cellulose. Thus, several methodologies are available for the isolation of lignin streams, e.g., acid hydrolysis, alkaline hydrolysis, hydrothermal degradation, or ionic liquids, and some modeling have been devoted to understand the fundamental mechanisms that control lignin–cellulose and lignin–hemicellulose interactions. Once a lignin stream is obtained, it can be used as raw feedstock for producing lignin-derived materials, e.g., carbon fibers or graphitic porous carbons. In many other cases, the valorization of lignin into usable materials requires additional depolymerization steps that yield monomers or small oligomers, hence the importance of modeling these smaller fragments that can be recovered and directly used, i.e., catechol, phenol, or be subjected to a second valorization step by means of biological or chemical methods, in order to upgrade the depolymerized fragments into high-value materials.

Lignin depolymerization is a challenging process, due to the presence of a handful of chemical linkages with different strengths in the structure of lignin, as it has
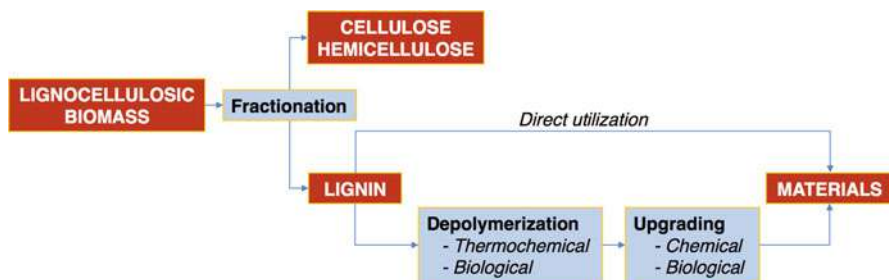
Fig. 3 The pathway from lignocellulosic biomass to lignin-derived materials

been discussed in the previous sections. Conventional depolymerization methods for lignin can be generally divided into (thermo)chemical or biological. The main (thermo)chemical routes include pyrolysis and hydrothermal/solvothermal processing (sometime aided by base catalysts), whereas biological depolymerization is mainly related to the use of extracellular enzymes secreted by ligninolytic organisms (mainly fungi) capable of degrading the aromatics structure of lignin (see Fig. 4).

When it comes to modeling, most of the work in the field has been done concerning the pretreatment, drying (Zhao et al. 2015), and pyrolysis of lignocellulosic biomass, which usually implies large-scale empirical modeling of the kinetic mechanisms, in the solid phase for biomass pyrolysis and in gas phase for secondary reactions of released products. This includes mathematical models of porous reactive mediums to study the processes of heat and mass exchange, drying, and ignition (Loboda et al. 2015), as well as to simulate heat and mass transfer in liquid, vapor, and bound water movement during drying processes (Jia 2008). Some multiscale models of hydrothermal pretreatment methods have been also developed, starting from the microscale level where chemical reaction takes place followed by the mesoscale level, in which the diffusion in lignocellulosic material is the main process, and finally the macroscale level where the mixing and bulk diffusion are important (Ali-Hosseini and Nilay Shah 2009). There are also some multiscale computational models that even scale up to the design and optimization of industrial plants that are devoted to the pretreatment of biomass (Colin et al. 2016). Moreover, jointly with plant bioengineering, the pathways for the biosynthesis of lignin can be understood and modified to design novel, artificial lignin variants that are more easily converted into materials, either directly or through strategies that involve depolymerization and subsequent upgrading. For instance, it has been reported that the side chains of the monolignols can be varied without compromising the lignification process (Vanholme et al. 2012).

To understand the enzymatic deconstruction of cellulose, Beckham et al. (2011) have provided extensive work on understanding the molecular-level mechanisms that enzymes employ to deconstruct plant cell walls. These authors discussed path
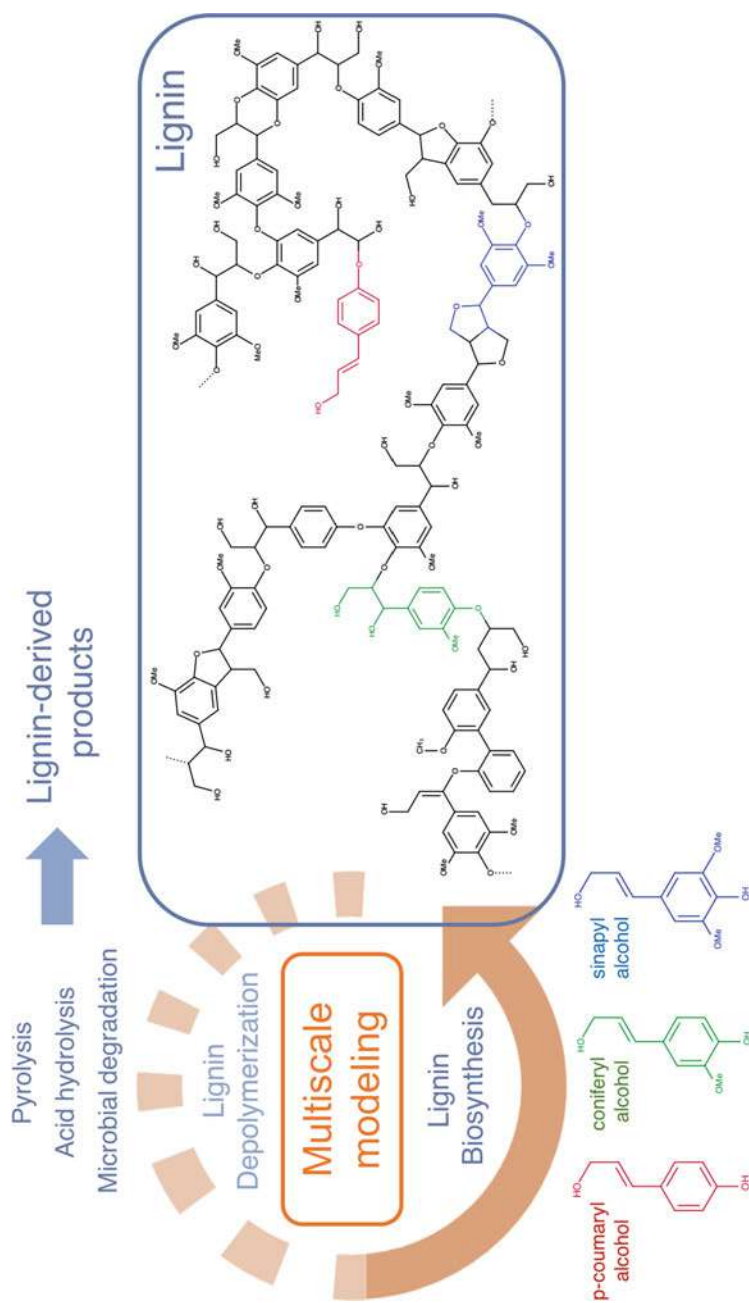
Fig. 4 Multiscale modeling guides the design of strategies for lignin depolymerization as well as new insights about lignin biosynthesis

sampling methods to measure rates of elementary steps of enzyme action, coarse-grained modeling for understanding macromolecular, cellulosomal complexes, methods to screen for enzyme improvements, and studies of cellulose at the molecular level. They have pointed out simulations as a complementary tool to understand carbohydrate-active enzymes and plant cell walls, which will enable industrial processes for the production of advanced, renewable fuels.

4.2   Direct Conversion of Lignin to Materials

Lignin can be directly subjected to chemical processes in order to produce several types of materials, such as carbon fibers for composite reinforcement or porous activated carbons with applications in energy storage, water purification, or CO_2 capture and sequestration. Lignin-derived carbon fibers can be obtained through different methods, like dry and wet spinning or melt spinning. This is typically followed by a stabilization process of the spinned lignin fibers (normally in an oxygen atmosphere) followed by a high-temperature treatment in an inert atmosphere (Beste 2014). Prior to these, lignin can be conditioned to improve the properties of the final carbon fibers, either chemically (e.g., acetylation, hydrogenation, or phenolation) or biologically. Moreover, the different monolignols that constitute lignin present different thermostabilization properties when forming carbon fibers, and thus the mechanical properties of the carbon fibers can be improved by adjusting the blending ratio of the various monolignols. Another way of valorizing directly lignin after its isolation is its hydrothermal carbonization (HTC). Here, mixtures of lignin and water are subjected to high pressures and temperatures. Under such conditions, water acquires properties very different from those at atmospheric conditions: i.e., high ionic product or reduced electrical conductivity and density. Despite the high thermal stability of lignin, HTC turns it into a solid material with a more condensed aromatic character and with properties that resemble those from brown coal. The resulting product has also a higher glass transition temperature and a higher thermal stability. During HTC, the β–O–4 linkages are mostly hydrolyzed, and the propyl side chains from the monolignol units are partially cleaved. Dealkylation and demethylation also take place, and more catechol-like units are obtained. The addition of catalysts (e.g., H_2SO_4 or Fe^{2+}) can enhance these reactions.

At this stage, modeling can be a useful tool here to understand the oxidative thermal conversion of lignin (Beste 2014) and how the different linkages between monolignols behave under this process (see Fig. 5). Theoretical studies of bond dissociation energies for lignin model compounds have been also performed for six prevalent linkages (β–O–4, α–O–4, 4–O–5, β–1, α–1, and 5–5) by using density functional theory methods (Huang et al. 2015). Some quantum MD studies have helped understanding fragment distribution of thermal decomposition of lignin monomer (Endo et al. 2008) as well as monomer and dimers of lignin from thermal decomposition (Kato et al. 2008).

Fig. 5 Lignin-derived products are currently produced through biological and thermochemical processes
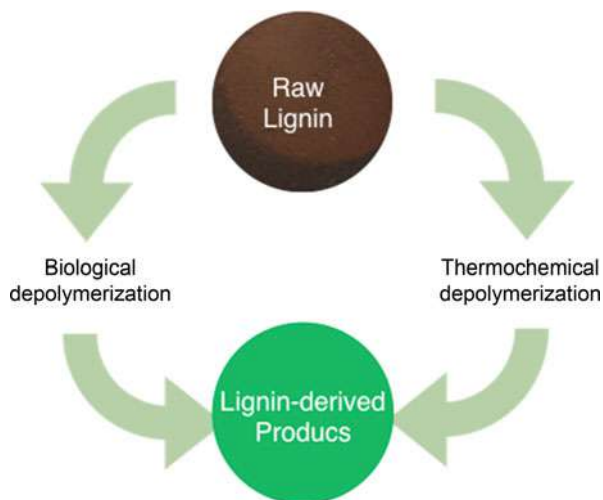
4.3   Thermochemical Depolymerization Processes

Lignin pyrolysis is carried out at high temperatures during short reaction times in an inert atmosphere, turning lignin into a liquid product with a complex mixture of hundreds of phenolic monomers and oligomers (Kawamoto 2017). It can be performed in the presence of catalysts, such as zeolite, in order to increase the selectivity toward certain products, like BTX (Li et al. 2015). It behaves differently, depending on the biomass type used. Softwood generates guaiacol-like phenols as main products, whereas hardwood tends to generate syringol and phenolic monomers and oligomers. Pyrolysis is a radical reaction, and hence the pyrolysis products tend to present stability issues and be prone to repolymerization. Hydrothermal processes take advantage of the special properties of water under sub- and supercritical conditions to break down lignin into smaller (even monomeric) aromatic molecules, such as catechol or guaiacol. Many of these molecules are useful building blocks for a bio-based chemical industry, perfumes, pharmaceuticals, or building blocks in organic synthesis. Under such conditions, water becomes highly reactive and breaks down the lignin structure. The presence of a liquid medium causes a dilution in the biomass feedstock, thus reducing the repolymerization of the reaction intermediates. A variation of this process includes the addition of a base to the aqueous or organic solvent (Erdocia et al. 2014), which has proved to lead to higher yields of the liquid product. Other organic solvents (e.g., acetone, ethanol, tetralin) (Breunig et al. 2018) can be used instead of water. The process is termed then solvothermal conversion. These organic solvents act as an in situ source of hydrogen or as hydrogen shuttle for the reaction. For instance, formic acid is reported to decompose under high pressure and temperature, yielding atomic hydrogen (Gasson et al. 2012). Under these conditions, lignin is quickly depolymerized, and the presence of atomic hydrogen in the reaction medium can cap the radicals generated during the process, thus avoiding subsequent polymerization reactions. Variations of the hydrothermal/solvothermal

process might include the use of heterogeneous catalysts or high-pressure hydrogen atmospheres (Breunig et al. 2018).

Modeling of pyrolysis is a multicomponent, multiphase, and multiscale problem of kinetic modeling of pyrolysis and combustion of biomass, which includes modeling of bio-oil and therefore a comprehensive description of kinetic and transport processes, both at the particle and the reactor scale (Ranzi et al. 2017). The models can provide information on heat of reaction, reaction products concentration, and reactants accumulation (Roman et al. 2011).

Blasi (2008) reviewed the state of the art in modeling chemical and physical processes of wood and biomass pyrolysis, indicating critical issues that require further investigation. Additional and very detailed reviews on biomass pyrolysis were provided also by Sharma et al. (2015), where they described operational and environmental challenges associated with pyrolysis at different scales, e.g., particle level, multiphase reacting flow, reactor performance, process integration and control, and by Anca-Couce (2016) with a focus on individual reaction mechanisms of cellulose, hemicellulose, and lignin and recent advances in understanding the fundamental reaction pathways. They demonstrated that a multiscale consideration of pyrolysis on multiple levels – specifically, on molecular, particle, and reaction levels – is required to accurately describe biomass pyrolysis.

More fundamental knowledge on reaction mechanisms during pyrolysis has been achieved by DFT calculations of model compounds (Kawamoto 2017), reaching a better understanding of the molecular mechanisms involved in lignin pyrolysis/gasification and guiding the development of better systems that overcome issues with low product selectivity.

4.4   Biological Depolymerization Processes

The first thing to remember when discussing depolymerization strategies is that the chemical bonds in lignin are rather stable, making them difficult to degrade by microbes. However, there are a few microorganisms in nature that encode metabolic pathways to break down lignin. These organisms are typically fungi that degrade lignin through two types of extracellular enzymes: laccases and peroxidases. These enzymes break down the lignin structure into aromatic radicals that subsequently cleave other lignin linkages (C_α–C_β, alkyl-aryl demethoxylation, and aromatic ring cleavage) by means of nonenzymatic reactions (Beckham et al. 2016). It has been reported that bacteria can also express those enzymes and degrade lignin (Bugg et al. 2011). However, both fungi and bacteria degrade lignin very slowly, which hinders their industrial implementation. Tools provided by synthetic biology and metabolic engineering (e.g., directed evolution or enzyme resurrection) (Alcalde 2015) can be of use to enhance the lignin degradation rate, making these microbes more prone for industrial implementation. Other issues of concern for the biological depolymerization of lignin are the need of a redox environment and aerobic conditions to operate, which demands a continuous supply of H_2O_2 and O_2.

From a modeling standpoint, the enzymatic biodegradation of lignin in soil by various microorganisms has been reviewed recently by Datta et al. (2017), with

focus on fungi and bacteria capable of producing ligninolytic enzymes. They state that very little is known about the degradation of lignin in soil, and they envision new efforts to identify new lignin-degrading microbes and thorough study of their genomics, biochemistry, and proteomics in the coming years.

As it happens with other protein-related areas, modeling can be applied to optimize the interaction of lignin with the enzymes (laccases, peroxidases) secreted by microbes. This can provide on the one hand guidelines for tailoring the lignin substrate to optimize its enzymatic degradation and, on the other hand, for tailoring the enzymes for optimized bioconversion or for expanding the substrate specificity for biological funneling during the upgrading of depolymerized lignin liquors (Beckham et al. 2016).

In any case, it is clear that understanding lignin-degrading reactions of ligninolytic enzymes is a challenge in which modeling has provided useful inputs by calculating binding affinities, intermediates, reaction energy profiles (Chen et al. 2011), as well as oxidation mechanisms using QM calculations (Elder 1997).

In the case of peroxidase, MD simulations have been the most used method to provide information on peroxidase behaviors in solution (Gerini et al. 2003), and the structure of peroxidase (Banci et al. 1994), as well as the effect of calcium ions on it (Banci et al. 1996). More recently, QM/MM methods have shed light into the intrinsic mechanisms that rule the interaction of tryptophan radicals in the long-range electron transfer of lignin peroxidase and lignin peroxidase-like systems (Bernini et al. 2012).

For plant and fungal laccases, some molecular docking and MD analyses have helped in unraveling the differential enzymatic catalysis with respect to lignin biosynthesis and degradation (Awasthi et al. 2015). Also, the factors affecting product distribution in laccase-catalyzed oxidation reactions have been studied by a combination of experimental and computational approaches (Lahtinen et al. 2013).

4.5   Chemical and Biological Upgrading of Depolymerized Lignin

The depolymerization products from lignin can be upgraded to other materials in two different ways: through chemical and biological processes. In both cases, the main challenge is that lignin depolymerization usually yields a complex slurry of aromatic compounds (Beckham et al. 2016) with very different reactivities.

Chemical processes for upgrading depolymerized lignin include, among others, classical chemical reactions such as alkylation, acylation, carboxylation, halogenation, or oxidation. There is even an industrially established process for the production of vanillin from lignin through chemical oxidation (Araújo et al. 2010). Lignin can be used as additive or substitutive feedstock for the production of adhesives, such as polyurethane foams or urea-formaldehyde resins. But lignin presents two issues that hinder their use as additives for adhesives: its wide range of molecular weights and its relatively low reactivity due to the presence of methoxy groups that suppress the reactivity of the aromatic rings. Most of the research is being devoted to enhance the reactivity of those aromatic rings, mainly by demethylating monolignols to produce polyols (Ferhan et al. 2013).

Some research is being done on the modeling side for studying the properties of lignin-derived adhesives (Lopez-Albarran et al. 2017). Using ReaxFF simulations, it was shown that the type of mono- and oligolignol precursors used to generate the adhesives affects their mechanical properties (i.e., the Young's modulus), and their results correlated quite well with experimental values.

The biological upgrading of depolymerized lignin offers the possibility of obtaining different types of products: molecules that maintain the aromatic ring from the phenolic monomers (e.g., vanillin (Lahtinen et al. 2013)), ring-opened species generated through aromatic catabolic pathways, or products obtained through central metabolic routes, where the microbes use "upper pathways" to funnel the carbon flux of the handful of aromatic materials coming from lignin depolymerization and convert them into central intermediates (e.g., protocatechuate or catechol) that then can be used by the microbes to produce energy storage compounds, such as TAGs or PHAs (Beckham et al. 2016). An interesting approach is to develop a one-pot system in which the enzymes that depolymerize the lignin into low molecular weight compounds work in concert with microbes that take up those low molecular weight compounds and convert them into high-value products. This is being developed at NREL under the name of Lignin Consolidated Bioprocessing (Salvachua et al. 2015). It is a process intensification strategy that reduces the number of unit operations needed and the capital expenses, due to the synergistic interaction between extracellular enzymes and microbes. Moreover, the intracellular accumulation of high-value compounds facilitates their recovery from the broth, as those products do not need complex separation processes from the lignin depolymerization liquor. Conversely, cells can be separated from the fermentation broth and subsequently lysed to recover their intracellular metabolites. Several studies have proved that bacterial species (e.g., *Pseudomonas putida* strains, acinetobacter ADP1 or *Rhodococcus jostii*) (Salvachua et al. 2015) can make use of the liquor obtained after depolymerizing lignin to produce polyhydroxy alkanoates (PHAs). These can be used as source of various materials, such as bioplastics for films, coating, biocompatible drug-delivery systems, and biomedical materials, among others (Linger et al. 2014).

5 Concluding Remarks

As we have seen, understanding lignin structure and composition, as well as its integration with cellulose and hemicellulose into the hierarchical structure of lignocellulosic biomass, is crucial for deciphering biosynthetic pathways, producing plants with more easily degradable cell walls, and developing better technologies for biomass depolymerization.

Computational chemistry calculations and molecular simulations along different scales of lignocellulosic biomass provide atomistic details, yet scalable predictions, on the structure energetics of reactions, and depolymerization processes involving lignin, cellulose, and hemicellulose, many of which still elude current experimental techniques.

Large part of the challenge when performing computational chemistry calculations comes from the selection of the correct methodology that provides accurate results in a reasonable amount of time. Thus, QM calculations have been mostly used to investigate the structure, electronic properties, and reactivity of monolignols, dehydrodilignols, and other small lignin-like model compounds. Nevertheless, the limited size of the models has not hampered the production of many useful information on the early stages of lignin polymerization, as well as the mechanisms for its degradation. The same statement applies to MD simulations to support better degradation strategies. Therefore, QM calculations and MD have provided extensive fundamental insight that result in improved technologies for biomass conversion.

In the future, the development of better molecular mechanics force fields, the use of more advanced QM/MM methods, and most importantly, the increased computational power provided by cloud computing, together with the implementation of new machine learning algorithms for speeded-up multiscale modeling and simulations, are expected to dramatically increase our understanding of lignocellulosic biomass structure, biosynthesis, and degradation. This includes atomistic models of the material but also large-scale empirical modeling of the kinetic mechanisms to study the processes of heat and mass exchange, drying, and ignition, as well as to simulate heat and mass transfer.

Unlocking the potential of lignocellulosic biomass is crucial to develop integrated biorefineries in a circular economy. However, it requires addressing fundamental challenges in economically degrading biomass into its basic constituents. A synergetic combination of multiscale modeling and process engineering is most definitely the way to achieve it.

References

Alcalde M (2015) Engineering the ligninolytic enzyme consortium. Trends Biotechnol 33(3):155–162. https://doi.org/10.1016/j.tibtech.2014.12.007

Ali-Hosseini S, Nilay Shah N (2009) Multiscale modelling of biomass pretreatment for biofuels production. Chem Eng Res Des 87(9):1251–1260. https://doi.org/10.1016/j.cherd.2009.04.018

Anca-Couce A (2016) Reaction mechanisms and multi-scale modelling of lignocellulosic biomass pyrolysis. Prog Energy Combust Sci 53:41–79. https://doi.org/10.1016/j.pecs.2015.10.002

Araújo JDP, Grande CA, Rodrigues AE (2010) Vanillin production from lignin oxidation in a batch reactor. Chem Eng Res Des 88(8):1024–1032. https://doi.org/10.1016/j.cherd.2010.01.021

Awasthi M, Jaiswal N, Singh S, Pandey VP, Dwivedi UN (2015) Molecular docking and dynamics simulation analyses unraveling the differential enzymatic catalysis by plant and fungal laccases with respect to lignin biosynthesis and degradation. J Biomol Struct Dyn 33(9):1835–1849. https://doi.org/10.1080/07391102.2014.975282

Banci L, Carloni P, Savellini G (1994) Molecular dynamics studies on peroxidases: a structural model for horseradish peroxidase and a substrate adduct. Biochemistry 33(41):12356–12366. https://doi.org/10.1021/bi00207a002

Banci L, Carloni P, Diaz A, Gori Savellini G (1996) Molecular dynamics calculations on peroxidases: the effect of calcium ions on protein structure. J Biol Inorg Chem 1(3):264–272. https://doi.org/10.1007/s007750050052

Beckham G, Bomble Y, Bayer E, Himmel M, Crowley M (2011) Applications of computational science for understanding enzymatic deconstruction of cellulose. Curr Opin Biotechnol 22(2):231–238. https://doi.org/10.1016/j.copbio.2010.11.005

Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR (2016) Opportunities and challenges in biological lignin valorization. Curr Opin Biotechnol 42(Supplement C):40–53. https://doi.org/10.1016/j.copbio.2016.02.030

Bernini C, Pogni R, Basosi R, Sinicropi A (2012) The nature of tryptophan radicals involved in the long-range electron transfer of lignin peroxidase and lignin peroxidase-like systems: insights from quantum mechanical/molecular mechanics simulations. Proteins Struct Funct Bioinf 80(5):1476–1483. https://doi.org/10.1002/prot.24046

Besombes S, Mazeau K (2004) Molecular dynamics simulations of a guaiacyl β-O-4 lignin model compound: examination of intramolecular hydrogen bonding and conformational flexibility. Biopolymers 73(3):301–315. https://doi.org/10.1002/bip.10587

Besombes S, Mazeau K (2005a) The cellulose/lignin assembly assessed by molecular modeling. Part 1: adsorption of a threo guaiacyl β-O-4 dimer onto a Iβ cellulose whisker. Plant Physiol Biochem 43(3):299–308. https://doi.org/10.1016/j.plaphy.2005.02.005

Besombes S, Mazeau K (2005b) The cellulose/lignin assembly assessed by molecular modeling. Part 2: seeking for evidence of organization of lignin molecules at the interface with cellulose. Plant Physiol Biochem 43(3):277–286. https://doi.org/10.1016/j.plaphy.2005.02.004

Beste A (2014) ReaxFF study of the oxidation of lignin model compounds for the most common linkages in softwood in view of carbon fiber production. J Phys Chem A 118(5):803–814. https://doi.org/10.1021/jp410454q

Blasi CD (2008) Modeling chemical and physical processes of wood and biomass pyrolysis. Prog Energy Combust Sci 34(1):47–90. https://doi.org/10.1016/j.pecs.2006.12.001

Breunig M, Gebhart P, Hornung U, Kruse A, Dinjus E (2018) Direct liquefaction of lignin and lignin rich biomasses by heterogenic catalytic hydrogenolysis. Biomass Bioenergy 111:352–360. https://doi.org/10.1016/j.biombioe.2017.06.001

Bugg TD, Ahmad M, Hardiman EM, Singh R (2011) The emerging role for bacteria in lignin degradation and bio-product formation. Curr Opin Biotechnol 22(3):394–400. https://doi.org/10.1016/j.copbio.2010.10.009

Calvo-Flores F, Dobado J, Isac-Garcia J, Martin-Martinez F (2015) Lignin and lignans as renewable raw materials: chemistry, technology and applications. Wiley series in renewable resource. Wiley, Chichester

Charlier L, Mazeau K (2012) Molecular modeling of the structural and dynamical properties of secondary plant cell walls: influence of lignin chemistry. J Phys Chem B 116(14):4163–4174. https://doi.org/10.1021/jp300395k

Chen M, Zeng G, Tan Z, Jiang M, Li H, Liu L, Zhu Y, Yu Z, Wei Z, Liu Y, Xie G (2011) Understanding lignin-degrading reactions of ligninolytic enzymes: binding affinity and interactional profile. PLoS One 6(9):e25647. https://doi.org/10.1371/journal.pone.0025647

Chen YR, Sarkanen S (2010) Macromolecular replication during lignin biosynthesis. Phytochemistry 71(4):453–462. https://doi.org/10.1016/j.phytochem.2009.11.012

Colin J, Rémond R, Perré P (2016) Design and optimization of industrial woody biomass pretreatment addressed by drykiln_crp, a multiscale computational model: particle, bed and dryer levels. Drying Technol 34(15):1820–1830. https://doi.org/10.1080/07373937.2016.1143483

Datta R, Kelkar A, Baraniya D, Molaei A, Moulick A, Meena RS, Formanek P (2017) Enzymatic degradation of lignin in soil: a review. Sustainability 9(7):1163. https://doi.org/10.3390/su9071163

Dos Santos H (2001) Conformational analysis of lignin models. Quim Nova 24(4):480–490. https://doi.org/10.1590/S0100-40422001000400009

Durbeej B, Eriksson L (2003) A density functional theory study of coniferyl alcohol inter-monomeric cross linkages in lignin – three-dimensional structures, stabilities and the thermodynamic control hypothesis. Holzforschung 57(2):150–164. https://doi.org/10.1515/HF.2003.024

Durbeej B, Eriksson L (2005a) Formation of β-O-4 lignin models – a theoretical study. Holzforschung 57(5):466–478. https://doi.org/10.1515/HF.2003.070

Durbeej B, Eriksson L (2005b) Spin distribution in dehydrogenated coniferyl alcohol and associated dilignol radicals. Holzforschung 57(1):59–61. https://doi.org/10.1515/HF.2003.009

Elder T (1997) Oxidation of a lignin model compound by the veratryl alcohol cation radical results from molecular orbital calculations. Holzforschung 51(1):47–56. https://doi.org/10.1515/hfsg.1997.51.1.47

Endo K, Matsumoto D, Kato K, Takagi Y, Ida T, Mizuno M, Saito K, Fukushima K, Kato N (2008) Fragment distribution of thermal decomposition for lignin monomer by QMD calculations using the excited and charged model molecules. Appl Surf Sci 255(4):1048–1051. https://doi.org/10.1016/j.apsusc.2008.05.277

Erdocia X, Prado R, Corcuera MÁ, Labidi J (2014) Base catalyzed depolymerization of lignin: influence of organosolv lignin nature. Biomass and Bioenergy 66(Supplement C):379–386. https://doi.org/10.1016/j.biombioe.2014.03.021

Ferhan M, Yan N, Sain M (2013) A new method for demethylation of lignin from woody biomass using biophysical methods. J Chem Eng Process Technol 4(5):1000160. https://doi.org/10.4172/2157-7048.1000160

Garver T (1996) Conformational analysis and 2D NMR assignment strategies for lignin model compounds. The structure of acetoguaiacyl-dehydro-diisoeugenol methyl ether. Can J Chem 74(2):173–184. https://doi.org/10.1139/v96-021

Gasson JR, Forchheim D, Sutter T, Hornung U, Kruse A, Barth T (2012) Modeling the lignin degradation kinetics in an ethanol/formic acid solvolysis approach. Part 1. Kinetic model development. Ind Eng Chem Res 51(32):10595–10606. https://doi.org/10.1021/ie301487v

Gerini M, Roccatano D, Baciocchi E, Di Nola A (2003) Molecular dynamics simulations of lignin peroxidase in solution. Biophys J 84(6):3883–3893. https://doi.org/10.1016/S0006-3495(03)75116-9

Haensel T, Reinmöller M, Lorenz P, Beenken W, Krischok S, Ahmed SU (2012) Valence band structure of cellulose and lignin studied by XPS and DFT. Cellulose 19(3):1005–1011. https://doi.org/10.1007/s10570-012-9681-9

Hindman DP, Lee JN (2007) Modeling wood strands as multi-layer composites: bending and tension loads. Wood Fiber Sci 39(4):515–516

Houtman CJ, Atalla RH (1995) Cellulose-lignin interactions (a computational study). Plant Physiol 107(3):977–984. https://doi.org/10.1104/pp.107.3.977

Huang J, Wu D, Tong H, Li W (2013) Molecular dynamic simulation study on pyrolytic behaviour of xylan. J Energy Inst 86(4):189–193. https://doi.org/10.1179/1743967113z.00000000058

Huang J, Wu S, Cheng H, Lei M, Liang J, Tong H (2015) Theoretical study of bond dissociation energies for lignin model compounds. J Fuel Chem Technol 43(4):429–436. https://doi.org/10.1016/S1872-5813(15)30011-6

Jia D, Afzal MT (2008) Modeling the heat and mass transfer in microwave drying of white oak. Drying Technol 26:1103–1111. https://doi.org/10.1080/07373930802266058

Jin K, Qin Z, Buehler MJ (2015) Molecular deformation mechanisms of the wood cell wall material. J Mech Behav Biomed Mater 42:198–206. https://doi.org/10.1016/j.jmbbm.2014.11.010

Kato K, Endo K, Matsumoto D, Ida T, Saito K, Fukushima K, Kato N (2008) Simulation of SIMS for monomer and dimer of lignin under the assumption of thermal decomposition using QMD method. Appl Surf Sci 255(4):1040–1043. https://doi.org/10.1016/j.apsusc.2008.05.074

Kawamoto H (2017) Lignin pyrolysis reactions. J Wood Sci 63(2):117–132. https://doi.org/10.1007/s10086-016-1606-z

Kha H, Tuble SC, Kalyanasundaram S, Williamson RE (2010) Wallgen, software to construct layered cellulose-hemicellulose networks and predict their small deformation mechanics. Plant Physiol 152:774–786. https://doi.org/10.1104/pp.109.146936

Lahtinen M, Heinonen P, Oivanen M, Karhunen P, Kruus K, Sipilä J (2013) On the factors affecting product distribution in laccase-catalyzed oxidation of a lignin model compound vanillyl alcohol: experimental and computational evaluation. Org Biomol Chem 11(33):5454–5464. https://doi.org/10.1039/c3ob40783g

Lee CM, Mohamed NMA, Watts HD, Kubicki JD, Kim SH (2013) Sum-frequency-generation vibration spectroscopy and density functional theory calculations with dispersion corrections (DFT-D2) for cellulose Iα and Iβ. J Phys Chem B 117(22):6681–6692. https://doi.org/10.1021/jp402998s

Li C, Zhao X, Wang A, Huber GW, Zhang T (2015) Catalytic transformation of lignin for the production of chemicals and fuels. Chem Rev 115(21):11559–11624. https://doi.org/10.1021/acs.chemrev.5b00155

Lim SH, Nahm K, Ra CS, Cho DW, Yoon UC, Latham JA, Dunaway-Mariano D, Mariano PS (2013) Effects of alkoxy groups on arene rings of lignin β-O-4 model compounds on the efficiencies of single electron transfer-promoted photochemical and enzymatic C-C bond cleavage reactions. J Org Chem 78(18):9431–9443. https://doi.org/10.1021/jo401680z

Linger JG, Vardon DR, Guarnieri MT, Karp EM, Hunsinger GB, Franden MA, Johnson CW, Chupka G, Strathmann TJ, Pienkos PT, Beckham GT (2014) Lignin valorization through integrated biological funneling and chemical catalysis. Proc Natl Acad Sci USA 111(33):12013–12018. https://doi.org/10.1073/pnas.1410657111

Loboda EL, Kasymov DP, Yakimov AS (2015) Modeling of thermophysical processes in the ignition of a small wooden plank. J Eng Phys Thermophys 88(1):111–121. https://doi.org/10.1007/s10891-015-1173-y

Lopez-Albarran P, Pizzi A, Navarro-Santos P, Hernandez-Esparza R, Garza J (2017) Oligolignols within lignin-adhesive formulations drive their Young's modulus: a ReaxFF-MD study. Int J Adhes Adhes 78:227–233. https://doi.org/10.1016/j.ijadhadh.2017.08.003

Mostaghni F, Teimouri A, Mirshokraei SA (2013) Synthesis, spectroscopic characterization and DFT calculations of β-O-4 type lignin model compounds. Spectrochim Acta Part A 110:430–436. https://doi.org/10.1016/j.saa.2013.03.075

Parthasarathi R, Romero R, Redondo A, Gnanakaran S (2011) Theoretical study of the remarkably diverse linkages in lignin. J Phys Chem Lett 2(20):2660–2666. https://doi.org/10.1021/jz201201q

Petridis L, Smith J (2009) A molecular mechanics force field for lignin. J Comput Chem 30(3):457–467. https://doi.org/10.1002/jcc.21075

Petridis L, Pingali SV, Urban V, Heller WT, O'Neill HM, Foston M, Ragauskas A, Smith JC (2011a) Self-similar multiscale structure of lignin revealed by neutron scattering and molecular dynamics simulation. Phys Rev E 83:061911. https://doi.org/10.1103/PhysRevE.83.061911

Petridis L, Schulz R, Smith J (2011b) Simulation analysis of the temperature dependence of lignin structure and dynamics. J Am Chem Soc 133(50):20277–20287. https://doi.org/10.1021/ja206839u

Ranzi E, Debiagi PEA, Frassoldati A (2017) Mathematical modeling of fast biomass pyrolysis and bio-oil formation. Note I: kinetic mechanism of biomass pyrolysis. ACS Sustain Chem Eng 5(4):2867–2881. https://doi.org/10.1021/acssuschemeng.6b03096

Roman M, Bobasu E, Selisteanu D (2011) Modelling of biomass combustion process. Energy Procedia 6:432–440. https://doi.org/10.1016/j.egypro.2011.05.050

Russell W, Forrester A, Chesson A (2000) Predicting the macromolecular structure and properties of lignin and comparison with synthetically produced polymers. Holzforschung 54:505–510. https://doi.org/10.1515/HF.2000.086

Salvachua D, Karp EM, Nimlos CT, Vardon DR, Beckham GT (2015) Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria. Green Chem 17(11):4951–4967. https://doi.org/10.1039/c5gc01165e

Sánchez-González Á, Martín-Martínez FJ, Dobado JA (2017) The role of weak interactions in lignin polymerization. J Mol Model 23(3):80. https://doi.org/10.1007/s00894-017-3257-4

Sangha AK, Parks JM, Standaert RF, Ziebell A, Davis M, Smith JC (2012a) Radical coupling reactions in lignin synthesis: a density functional theory study. J Phys Chem B 116(16):4760–4768. https://doi.org/10.1021/jp2122449

Sangha AK, Petridis L, Smith JC, Ziebell A, Parks JM (2012b) Molecular simulation as a tool for studying lignin. Environ Prog Sustain Energy 31(1):47–54. https://doi.org/10.1002/ep.10628

Senn H, Thiel W (2009) QM/MM methods for biomolecular systems. Angew Chem Int Ed 48(7):1198–1229. https://doi.org/10.1002/anie.200802019

Sharma A, Pareek V, Zhang D (2015) Biomass pyrolysis—a review of modelling, process parameters and catalytic studies. Renew Sust Energ Rev 50:1081–1096. https://doi.org/10.1016/j.rser.2015.04.193

Shevchenko S, Bailey G (1996) The mystery of the lignin-carbohydrate complex: a computational approach. J Mol Struct Theochem 364(2–3):197–208

Simon JP, Eriksson KEL (1995) A molecular mechanics investigation of lignin structure. I. Conformational analysis of 1-phenyl-2-phenoxy-1,3-propanediol using MM3. Holzforschung 49(5):429–438. https://doi.org/10.1515/hfsg.1995.49.5.429

Smith MD, Petridis L, Cheng X, Mostofian B, Smith JC (2016) Enhanced sampling simulation analysis of the structure of lignin in the THF-water miscibility gap. Phys Chem Chem Phys 18(9):6394–6398. https://doi.org/10.1039/c5cp07088k

Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W (2012) Metabolic engineering of novel lignin in biomass crops. New Phytol 196(4):978–1000. https://doi.org/10.1111/j.1469-8137.2012.04337.x

Watts HD, Mohamed MNA, Kubicki JD (2011) Evaluation of potential reaction mechanisms leading to the formation of coniferyl alcohol [small alpha]-linkages in lignin: a density functional theory study. Phys Chem Chem Phys 13(47):20974–20985. https://doi.org/10.1039/c1cp21906e

Zhang X, Yang W, Blasiak W (2011) Modeling study of woody biomass: interactions of cellulose, hemicellulose, and lignin. Energy Fuels 25(10):4786–4795. https://doi.org/10.1021/ef201097d

Zhao J, Fu Z, Jia X, Cai Y (2015) Modeling conventional drying of wood: inclusion of a moving evaporation interface. Drying Technol 34(5):530–538. https://doi.org/10.1080/07373937.2015.1060999

Simple Asphaltene Thermodynamics, Oilfield Reservoir Evaluation, and Reservoir Fluid Geodynamics

67

Oliver C. Mullins, Julian Y. Zuo, Hadrien Dumont, Vinay K. Mishra, and Li Chen

Contents

Abstract

Reservoir crude oils consist of dissolved gases, liquids, and dissolved solids, the asphaltenes. For 40 years, gas-liquid equilibria of crude oils have been treated with cubic equations of state, variants of the van der Waals EoS. However, asphaltenes had eluded a similar thermodynamic treatment. Recently, the molecular and nanocolloidal structures of asphaltenes in crude oil have been developed and codified in the Yen-Mullins model. With sizes known, the gravity term clarifies and can be added to the Flory-Huggins polymer solution theory, yielding the Flory-Huggins-Zuo (FHZ) EoS. Measurement of gradients of crude oils vertically and laterally in reservoirs is best accomplished with "downhole fluid analysis" (DFA). Interpretation of asphaltene gradients with the FHZ EoS allows determination of the extent of thermodynamic equilibrium of reservoir crude oils. Equilibrated crude oils imply flow connectivity in reservoirs, addressing one of the most important uncertainties in reservoirs. Disequilibrium fluid gradients in reservoirs imply ongoing fluid processes in geologic time. Specification of many of these processes in reservoir case studies has led to

O. C. Mullins (✉) · J. Y. Zuo · H. Dumont · V. K. Mishra · L. Chen
Schlumberger Oilfield Services, Houston, TX, USA
e-mail: mullins1@slb.com; julianzuo@hotmail.com; HDumont@slb.com; vkmishra@slb.com; LChen5@slb.com

1649

a newly formalized discipline, "reservoir fluid geodynamics" (RFG). Many reservoir concerns are addressed utilizing RFG analyses including reservoir connectivity, viscosity gradients, solution gas gradients, variable productivity index, tar mat formation, and sluggish aquifer support. The combination of new science embodied in the Flory-Huggins-Zuo EoS and the Yen-Mullins model and new technology, DFA, has led to an explosion of applications through reservoir fluid geodynamics.

1   Introduction

Reservoir crude oils are a complex mixture consisting of dissolved gases, liquids, and dissolved (or colloidally suspended) solids, the asphaltenes as conveyed in Fig. 1. As such, the physical and chemical properties of reservoir hydrocarbons can vary enormously, for example, from dry natural gas to immobile tar (Mullins 2008). The economic value and the methods of production and processing all depend on the composition of the oil especially regarding these three phases, gas, liquid, and solid. In addition, there can be important and deleterious chemical constituents of the reservoir fluid such as CO2, H2S, and Hg. It is essential to know the properties of the crude oil in order to know whether to produce the reservoir fluid. For production, a variety of concerns must be addressed to develop production methods, gas handing facilities, flow assurance processes, and methods of reservoir pressure support. All of these concerns depend critically on the properties of the reservoir fluid. In addition, the reservoir fluid properties can change dramatically vertically and laterally in the reservoir. Such fluid variation must also be understood for implementation of optimal production planning. Moreover, remote sensing such as seismic imaging is ineffective in the determination of properties of the reservoir hydrocarbons with the exception of some ability to identify light hydrocarbon fluids in the reservoir. Consequently, it is necessary to acquire fluid samples prior to production; there is no substitute.

Figure 2 shows a series of "dead" oil samples from a single reservoir from deepwater Gulf of Mexico (Elshahawi et al. 2009). This figure shows clear differences in fluid colors, with lighter colors near the top of the reservoir and darker colors near the bottom. The color difference reflects large differences in asphaltene

Fig. 1 Reservoir crude oils consist of dissolved gas, liquid, and dissolved (or colloidally suspended) solids, the asphaltenes (Mullins 2008). All economics and all production methods depend on the composition of the hydrocarbon
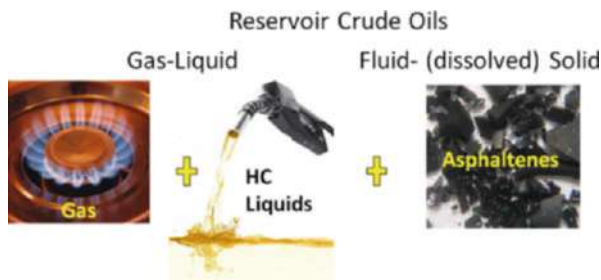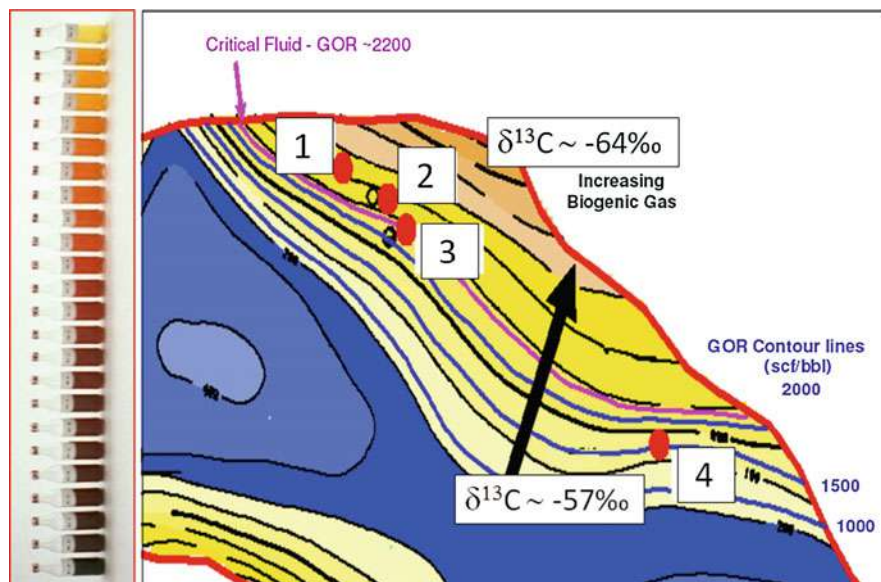
Fig. 2 (Left) Huge color and GOR gradients in crude oil composition are observed in oil samples within a single connected reservoir, deepwater Gulf of Mexico (Elshahawi et al. 2009). (Right) The areal map of the reservoir is shown with contour lines which also represent different gas-oil ratios (GORs). Four well locations are shown. The methane stable carbon isotope ratios indicate a greater fraction of thermogenic gas lower in the reservoir and a greater fraction of biogenic gas toward the top (Elshahawi et al. 2009)

content of the fluids. In addition, gas-oil ratios (GORs) of the corresponding live oil samples vary enormously from 8000 to 1800 scf/bbl (standard cubic feet of gas per barrels of oil at 1 atm and 60 °F), further demonstrating the wide range of fluid composition encountered in a single reservoir(Elshahawi et al. 2009; Zuo et al. 2016). (Dead crude oils have been flashed to 1 atmosphere thereby losing dissolved gases; live crude oil contains its dissolved gases as in the reservoir at elevated pressures.)

Figure 2 shows that the carbon isotope varies throughout the reservoir, indicating that there are two sources of methane in this reservoir and that primary biogenic methane in this reservoir is substantial. (Primary) biogenic methane is generally enriched in ^{12}C typically about $\delta^{13}C = -68‰$, while thermogenic methane resulting from the catagenesis of kerogen is typically in the range of $\delta^{13}C = -40‰$ (Hantschel and Kauerauf 2009; Tissot and Welte 1984). Consequently, this reservoir experienced a mixture of biogenic methane and crude oil. (Primary biogenic methane corresponds to biodegradation of organic matter in sediments; secondary biogenic methane is from biodegradation of crude oil.)

In addition to fluid complexities, the reservoir can possess innumerable complexities associated with the rocks or "formations" which comprise the hydrocarbon-bearing formations and the sealing formations which trap the hydrocarbon. The
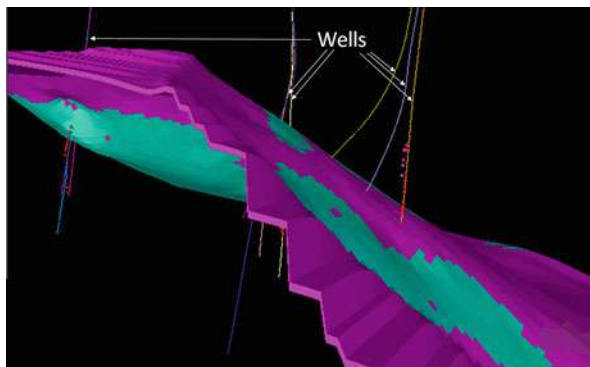
Fig. 3 The upper and lower surfaces of an oil reservoir, deepwater Gulf of Mexico (Mullins et al. 2007a). Wells are drilled to obtain required information about fluid and rock formation properties for consideration of possible economic production of crude oil. Reservoirs are large, and well penetrations are few when major decisions must be made. Methods of interpolation and extrapolation for fluid and formation properties are required

overall size of the reservoir is of immediate concern and is addressed by seismic imaging. Nevertheless, the quality of seismic imaging degrades with depth and with intervening salt and other irregular structures. In addition, key concerns such as the extent of fluid flow or "connectivity" throughout the reservoir cannot be addressed adequately by remote sensing. Consequently, wellbore data is critical to understand formation properties in addition to fluid properties. Figure 3 shows the upper and lower surfaces or horizons of a reservoir, in excess of 20,000 ft below sea level and with a water depth in excess of 4000 ft (Mullins et al. 2007a). Several wells are depicted that were used to perform reservoir evaluation. After favorable results, development wells were subsequently drilled. This reservoir is tilted from basin distortion; faults in the reservoir are evident in seismic imaging. It is imperative to determine the extent of flow connectivity across the reservoir.

Even if faults are identified in seismic imaging, it can remain unclear whether the fault is sealing or transmissive. In addition to flow connectivity or its inverse "compartmentalization" in a reservoir, there are many other concerns associated with the rock formations. The existence of "baffles" which act to divert fluid flow but which are not pressure seals can have a significant impact on production rates and aquifer sweep in production. Baffles are particularly hard to identify prior to production. The extent of aquifer pressure support and sweep are of major production concerns. There are a large variety of concerns regarding the petrophysical properties of the rock formations. Porosity, permeability, oil saturation, and their variations throughout the reservoir must be understood; well log data is essential to address the extensive range of fluid and rock properties of interest.

Downhole Fluid Analysis, Asphaltene Thermodynamics, and Reservoir Fluid Geodynamics. In order to understand fluid distributions in reservoirs, it is critically important to measure fluid gradients accurately. Downhole fluid

analysis (DFA) within oil wells affords this opportunity (Mullins 2008). Thermodynamic modeling enables projection of fluid properties away from wellbores over large reservoirs. In particular, the recent development of the thermodynamics of asphaltene gradients has led to powerful new methods of reservoir evaluation. Moreover, this asphaltene modeling applied to DFA gradient data can be used to identify fluid equilibrium throughout reservoirs. Equilibrated fluids indicate connected reservoirs giving new ways to address a major industry concern. When the fluid columns are not equilibrated, there are often processes taking place in geologic time that preclude equilibrium. This has led to the recent formalization of a new technical discipline, "reservoir fluid geodynamics," accounting for fluid redistributions and phase transitions in reservoirs after trap filling. To understand present-day realizations of reservoirs, it is important to understand their evolution in geologic time. This chapter treats many aspects of these new developments.

2   Wireline Well Logging and Downhole Fluid Analysis (DFA)

In a typical scenario, a tool package is lowered into the well on a cable or "wireline" shortly after a well is drilled. In this circumstance, the wellbore is "openhole" meaning uncased in the well section of interest which intersects the reservoir. That is, there is no steel casing placed in this section of the well. In openhole conditions, the well is filled with drilling mud which maintains higher borehole pressure than that in penetrated permeable formations, thereby preventing blowouts, the uncontrolled entry of formation fluids into the well. The drilling fluid contains clay, thus its name drilling mud, which acts to form a mud cake on permeable formations thereby preventing excessive fluid loss from the wellbore into the formations. Nevertheless, some drilling fluid "filtrate" leaks into permeable zones thereby contaminating formation crude oils. Oil-based muds (OBM) are commonly used, and OBM filtrates are miscible with reservoir crude oils making OBM filtrate contamination a significant concern (Mullins 2008). Water-based muds are much less problematic for acquisition of crude oil samples.

Oil wells provide the opportunity to measure all manners of properties; whole cores can be acquired addressing key geological concerns. Petrophysical logs can be run and include high-resolution imaging, nuclear magnetic resonance, neutron scattering, gamma ray scattering, phase and amplitude response at many electromagnetic frequencies, and phase and amplitude response at many acoustic frequencies (Ellis and Singer 2008), all providing a wealth of information about the formation primarily near the wellbore. Fluid samples can be acquired using an openhole sampling tool such as the Schlumberger MDT (Zimmerman et al. 1989, 1990).

To acquire a sample of formation fluid in openhole, a tool is lowered into the well on a wireline cable; Fig. 4 depicts the Schlumberger MDT (Modular Formation Dynamics Tester) used for this purpose (Mullins 2008; Zimmerman et al. 1989, 1990; Kazakevich et al. 2016). DFA tools are mounted onboard the
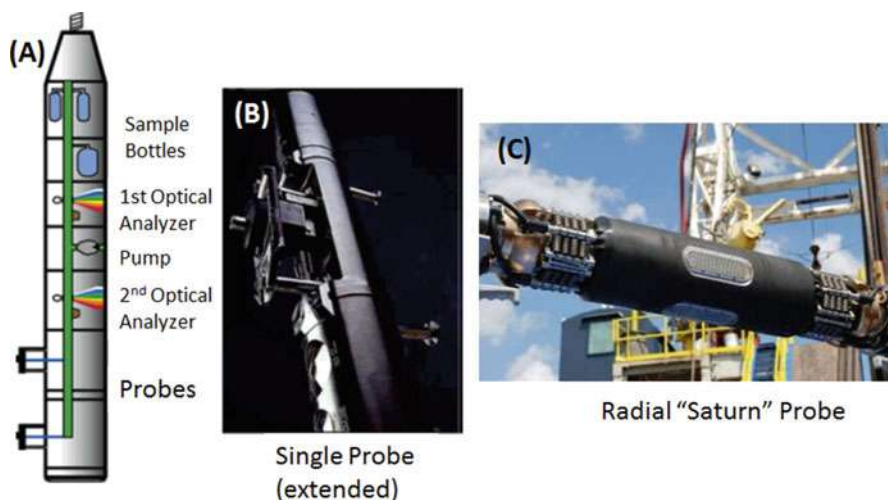
Fig. 4 (**a**) Schematic of the MDT fluid acquisition and analysis tool for openhole wells (Zimmerman et al. 1989, 1990). Probes interface with permeable formations for fluid extraction using pumps. Optical fluid analyzers (DFA tools) and high-pressure sample bottles are depicted (Mullins 2008; Zimmerman et al. 1989, 1990). (**b**) A photograph of a "probe" with a protruding steel tube that is pressed firmly against the borehole wall to make hydraulic communication with permeable zones in the oil well to extract formation fluids. (**c**) Photograph of the Schlumberger radial Saturn probe where fluid entry occurs at four symmetrically placed ports in the middle of a rubber packer, which contact the borehole wall when deployed (Kazakevich et al. 2016)

MDT and enable accurate measurement of fluid properties and their variations within the oilfield (Mullins 2008). First, DFA identifies the presence or absence of drilling fluid filtrate that can leak into permeable zones and contaminate reservoir crude oil (Mullins 2008). Pumping fluid out of the formation for extended times by the MDT, for example, a few hours, is frequently required to greatly reduce or eliminate contamination of the crude oil by OBM filtrate. DFA is used to monitor the extent of OBM contamination; crude oils generally have solution gas and/or asphaltenes, while OBM filtrate generally has neither (Mullins 2008; Mullins et al. 2000). This distinction along with knowledge of the typical time rate of change of contamination allows real-time determination of contamination during the pumping process (Mullins 2008; Mullins et al. 2000). Many different probe types are available that make contact with the borehole wall. This enables sample acquisition and DFA to be performed in a large variety of formations including unconsolidated sands, low-permeability formations, and formations with extensive drilling filtrate invasion (Kazakevich et al. 2016).

Asphaltene content is determined in DFA by measuring the electronic absorption of the oil in the visible and near-infrared spectral ranges (Mullins et al. 1992; Mullins and Zhu 1992; Kharrat et al. 2013). Figure 5 shows the optical absorption spectra of many dead crude oils. Plotting optical density (OD) on a log plot versus

photon energy shows that all crude oils exhibit the same slope. This behavior is reminiscent of the Urbach tail from solid-state physics and relates to the constant ratio of small to large bandgap components in all crude oils (Mullins et al. 1992; Mullins and Zhu 1992).

The two-stretch overtone region provides the opportunity to estimate the gas-oil ratio (GOR) of live crude oils (Mullins et al. 2001; Dong et al. 2002; Agthoven et al. 2002; Fujisawa et al. 2002). Figure 6a shows linearity of the spectra of methane, heptane, and a mixture of the two. The NIR peaks of methane, methyl, and methylene groups are readily resolvable using optical hardware compatible with the downhole environment (Mullins et al. 2001; Dong et al. 2002). Figure 6b shows the CH two-stretch overtone region of methane, a series of live crude oils, and a dead crude oil. Large spectral differences are seen in both the methane peak and "oil" peak for the crude oils. Figure 6c shows the correlation of GOR with OD ratios for many crude oils. A variety of interpretation algorithms are available to obtain the GOR from a live crude oil using NIR spectroscopy (Mullins et al. 2001; Dong et al. 2002). In addition, other analytes such as CO_2 can be measured using NIR spectroscopy in situ in oil wells (Agthoven et al. 2002).

DFA measurements also include density and viscosity measurements using the response to some form of mechanical vibrating object that interacts with the MDT flowline fluid. For example, the ringdown (loss) of a mechanical oscillator in fluid is greatly impacted by fluid viscosity (Sullivan et al. 2009; Etchart et al. 2008), and the reduction of resonant frequency of a mechanical oscillator is greatly impacted by the surrounding fluid density (Sullivan et al. 2009; Etchart et al. 2008.) Figure 7 provides a schematic of the Schlumberger in situ fluid analyzer (IFA) which is a module of the MDT that performs advanced DFA (O'Keefe et al. 2007).

DFA is performed on a fluid flow stream in the MDT flowline (cf. Fig. 4); visible and near-infrared spectroscopy are the methods of choice for downhole measurement of crude oil composition (Mullins 2008). Reservoir and MDT flowline pressures can exceed 30,000 psi; but this pressure is readily handled in high-pressure optics cells (with sapphire windows). Reservoir temperatures can be 175 °C or more; the primary impact of this temperature on DFA measurements is on optical detectors. The use of InGaAs detectors with bandgaps corresponding to a wavelength of 1.7 μ or slightly longer limits the thermal noise to acceptable levels. Aside from affecting fluid density, reservoir pressure and temperature do not impact the fluid spectra significantly (Fujisawa et al. 2002). To keep optical density values within a good range for crude oils, a pathlength of 2 mm is typically used in DFA tools (Mullins 2008).

Laboratory Two-Dimensional Gas Chromatography (GC×GC). DFA provides the concentrations of dissolved gases, liquids, and dissolved solids of reservoir crude oils. The corresponding gradients are very amenable to a thermodynamic treatment. It is desirable to perform a detailed compositional analysis of the crude oils, especially with a geochemical interpretation, to compare with the thermodynamic analysis.
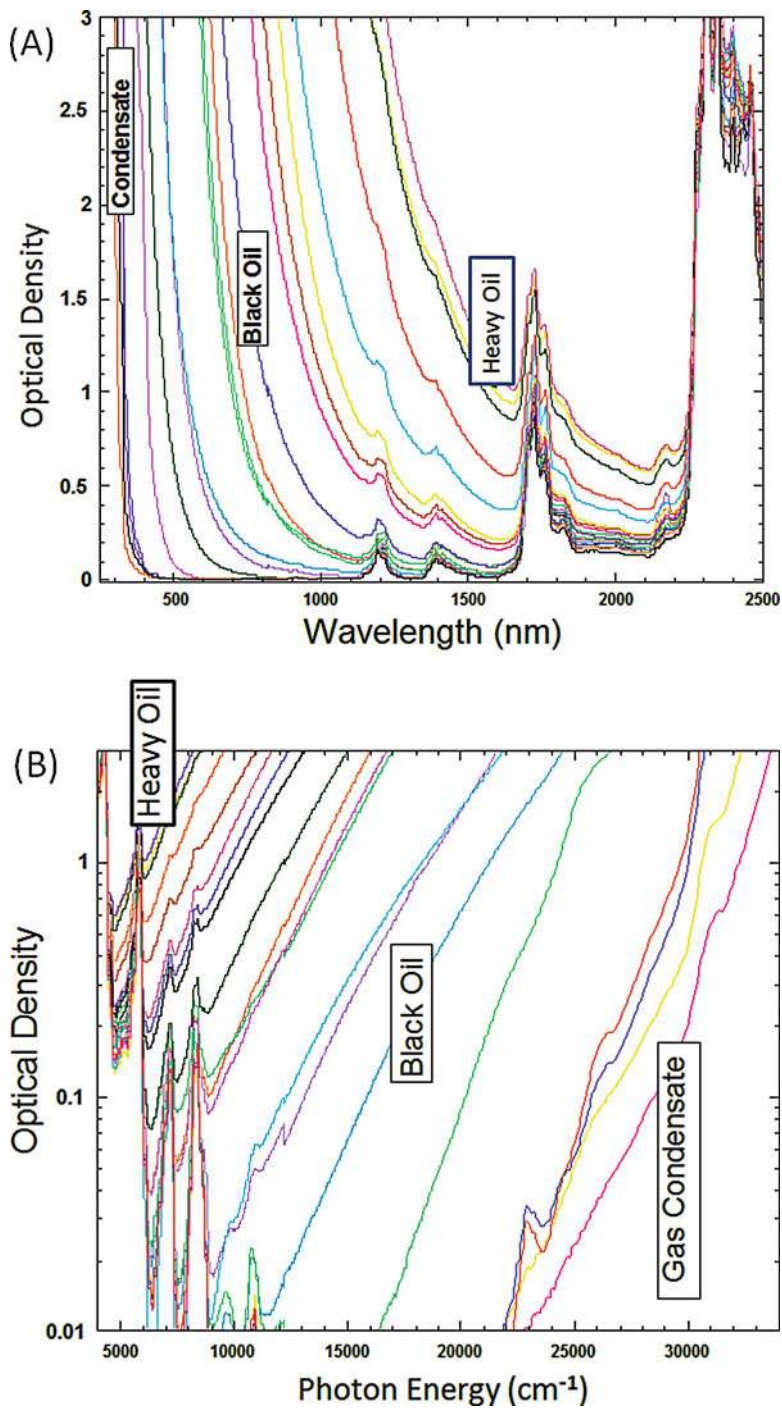
Fig. 5 (continued)

Gas chromatography (GC) is a powerful tool to separate complex molecular mixtures for compound identification and quantification. Conventional GC uses a single capillary column coated with a stationary phase that influences how compounds are separated by a combination of properties such as molecular weight, volatility, polarity, or polarizability. GC has been very useful for analyzing complex mixtures, such as crude oil, and for resolving many individual compounds of interest. However, conventional GC has limited utility for unresolved complex mixtures (UCMs) in weathered and biodegraded oils and overlapping or interfering peaks that impede compound identification and analysis. In contrast, comprehensive two-dimensional gas chromatography (GC×GC) adds a second column after the primary column with a different stationary phase to separate compounds according to two combinations of physicochemical attributes. GC×GC offers advantages over conventional GC in separating highly complex mixtures, such as crude oil, into thousands of fully resolved individual compound peaks and with low variability in response factors between compounds (Nelson et al. 2006, 2016; Frysinger and Gaines 2001). Compound peaks in GC×GC are less affected by coeluting components due to higher peak capacity and increased resolving power (Frysinger and Gaines 2001). In a common configuration of columns, GC×GC separated components according to their molecular weight (first axis) and polarizability (second axis). Those properties are particularly diagnostic of RFG processes such as biodegradation, water washing, and thermal maturity, making GC×GC a natural complement to DFA for RFG analysis (Nelson et al. 2006).

GC and GC×GC with flame ionization detection (FID) offer advantages of quantitative detection of peak abundances, reproducibility, and increased sensitivity. GC-FID is used to examine the overall *n*-alkane distribution followed by GC×GC-FID for simultaneous detection of saturated and aromatic compounds plus establishing chemical retention maps. Compounds in GC and GC×GC-FID are identified through the use of standard materials, retention times, and comparison with literature. To fully understand the identity of unknown compounds in chromatograms, time-of-flight mass spectrometry (TOFMS) is coupled with GC×GC permitting identification of compounds with fragmentation patterns matching with known components in various compound libraries. GC×GC-TOFMS is used to assign retention times to compounds that are then used in GC×GC-FID measurements for quantitative analysis of individual peaks. Figure 8 is an example of GC×GC analysis shown below for a crude oil acquired in the Llanos basin, Colombia (Bartha et al. 2015).

Fig. 5 (**a**) The optical absorption spectra of dead crude oils (at 1 atmosphere). There is a huge and systematic variation of crude oil color among different crude oils. (**b**) Optical density on a logarithm plot versus photon energy yields parallel, straight electronic absorption edges, which is reminiscent of the Urbach tail (Mullins and Zhu 1992; Kharrat et al. 2013). The coloration of an oil is a linear function of asphaltene content (Mullins et al. 2007a; Kharrat et al. 2013)
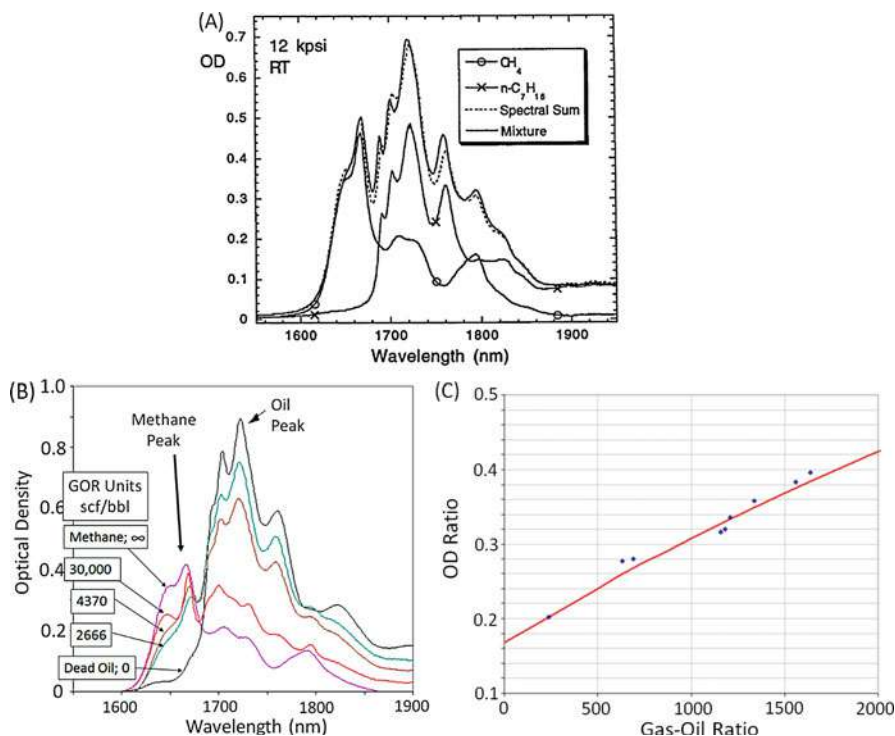
Fig. 6 (**a**) Two-stretch overtone for selected hydrocarbons showing largely separable peaks dominated by methane (1650 nm), methyl (1700 nm), and methylene (1725 nm) groups (Mullins et al. 2001). These peaks add linearly in compositional mixtures (Mullins et al. 2001; Dong et al. 2002). (**b**) Spectra of methane, live crude oils (with dissolved gases), and dead oil (Dong et al. 2002). (**c**) The optical density (OD) ratio of the oil peak to methane peak correlates with GOR (Mullins et al. 2007a). This spectral analysis is used to estimate GOR downhole using various algorithms (Mullins et al. 2001; Dong et al. 2002)

3   Thermodynamic Modeling of Reservoir Fluids

With these chemical evaluation tools of DFA and complementary laboratory methods, the chemical properties of oil samples can be resolved with great accuracy. Nevertheless, the ability to interpret fluid properties between wells and extrapolate fluid properties where no wells exist is critical to oilfield development planning. Naturally, thermodynamic equations of state are the first step in this process.

Crude oils in reservoirs can exhibit significant vertical and lateral compositional gradients. In general, the most important gradients relate to GOR and asphaltene content. GOR gradients are important because many production parameters relate to this variable. Economic value depends critically on GOR; in certain oilfield settings, gas cannot be transported so is not economic. In production, handling high-pressure gas is a major concern, while liquids are relatively easy to handle; again, GOR is key.
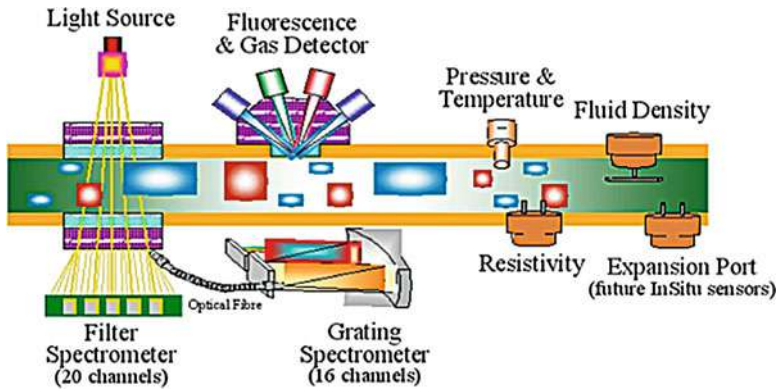
Fig. 7 The in situ fluid analyzer (IFA), an advanced DFA tool that is a module of the MDT (O'Keefe et al. 2007). This tool performs near-infrared and visible spectroscopy with a filter spectrometer and a grating spectrometer. Gas detection and fluorescence detection are also performed. Fluid density, viscosity (not shown), pressure, and temperature are also measured

Frequently, it is desirable to maintain reservoir pressure in excess of the saturation pressure of the oil because the appearance of newly formed gas bubbles in porous media results in a decline of the formation oil relative permeability. In addition, because gas has much lower viscosity than liquid, it can flow preferentially, leaving valuable liquids behind in the formation. The saturation pressure of an oil is closely related to its GOR. For these and other reasons, it is important to understand and model GOR gradients. Various so-called cubic equations of state (EoSs) such as the Peng-Robinson equation are very effective at modeling GOR gradients as well as the gas-liquid-phase behavior of crude oils and have been in use for 40 years (Peng and Robinson 1976). The various cubic EoSs that are used for live crude oils are variants of the familiar van der Waals EoS. There are many excellent treatments of the cubic EoS applied to crude oil (Peng and Robinson 1976; Høier 1997; Høier and Whitson 2001). Of course, equation of state modeling applies only for crude oil in thermodynamic equilibrium, and many reservoirs are not equilibrated. Nevertheless, the cubic EoS provides a good starting point.

For equilibrated crude oils, large GOR gradients are expected only for high GOR crude oils because they are compressible (Peng and Robinson 1976; Høier 1997; Høier and Whitson 2001). Hydrostatic head pressure of the oil column creates higher pressure toward the base of the column. For compressible crude oils, this higher pressure causes higher density toward the base of the column. The density gradient causes lower-density components such as methane to accumulate toward the top of the column. Thus, the density gradient creates a compositional gradient. In contrast, low GOR crude oils are not very compressible; the hydrostatic head pressure of the oil column does not induce a density gradient. Consequently, equilibrated low GOR crude oils have fairy homogeneous GORs (Høier 1997; Høier and Whitson 2001). For so-called black oils with their limited GORs, and for heavy oils with low GORs, the cubic EoS has limited application. Indeed,
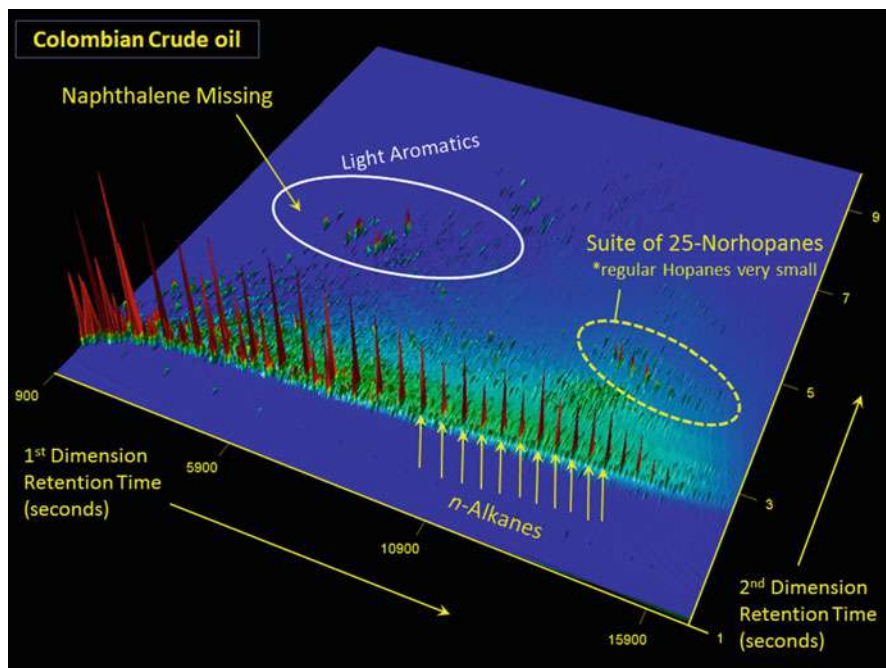
Fig. 8 GC×GC chromatogram of a Colombian oil from the Llanos basin (Bartha et al. 2015). The excellent resolving power allows chemical specificity of many peaks. Here, *n*-alkanes are dominant indicating no biodegradation. In addition, 25-norhopanes are much more prominent than the corresponding hopanes, indicating severe biodegradation. Two crude oil charges entered this reservoir; the first was severely biodegraded after emplacement. The second lighter charge entered the reservoir after subsidence and heating of the reservoir beyond 80 °C, thus deactivating and killing the microbes. Petroleum system modeling gives concurring results (Bartha et al. 2015). The loss of naphthalene is due to water washing of the crude oil; the aquifers in the Llanos basin are extremely active (Bartha et al. 2015)

the properties of black oils and heavy oils are defined in great measure by their asphaltene content, not their GOR. It is important to employ an EoS for asphaltene gradients to understand reservoirs and contained fluids.

Asphaltene Nanostructures and the Yen-Mullins Model. The development of an EoS for asphaltene gradients requires knowledge of the size of asphaltene structure in reservoir crude oil. The force F of gravity g must be understood to model reservoir fluids. Newton's second law of motion, $F = mg$ in a gravitational field, requires the mass m of the object. For asphaltenes, this was previously unknown, thus precluding a first-principles EoS for asphaltene gradients in reservoirs. Indeed, even the molecular weight of asphaltenes was unresolved over several orders of magnitude (Mullins et al. 2007b, 2008a). The first measurement of molecular diffusion of asphaltenes utilized time-resolved fluorescence depolarization (TRFD) and showed that asphaltenes are small molecules, with predominantly one polycyclic aromatic hydrocarbon (PAH) per molecule (Groenzin and Mullins

1999, 2000). The TRFD measurements only involve dissolution of the asphaltenes under very mild conditions. Laser desorption, laser ionization mass spectrometry (L^{2}MS) compared 23 "island" model compounds (one PAH per molecule), and "traditional archipelago" model compounds (multiple PAHs per molecule, linked by alkane chains) with asphaltenes and confirmed asphaltenes are dominated by island molecular structure (Sabbah et al. 2011). More recently, a group at IBM Zurich obtained ultrahigh-resolution images of asphaltenes (Schuler et al. 2015, 2017a). The analysis of hundreds of molecular images in ten diverse samples with all kinds of asphaltenes showed the lack of even one traditional archipelago molecule (Schuler et al. 2015, 2017a). Figure 9 shows an example of imaging an asphaltene molecule. In addition, this group proved the ability to image a variety of traditional archipelago molecules (Schuler et al. 2017b). Evidently, traditional archipelagos are not observed in asphaltenes because they are not present. In particular, the nondestructive transfer from a rapidly heated wire to a cold surface of individual asphaltenes without fragmentation was validated. Most importantly, experiments were performed to validate that the imaging methods could resolve traditional archipelago molecules (Schuler et al. 2017b). Excellent images were obtained of 1,2-di(pyren-1-yl)ethane with no fragmentation at all despite the presence in this molecule of one of the weakest C–C bonds, the ethane linkage with bond strength 65.2 kcal/mol (Schuler et al. 2017b).

At low concentrations, the asphaltenes form a true molecular solution in toluene. With increasing concentration, nanoaggregates form. The critical nanoaggregate concentration (CNAC) has been observed by a variety of methods including high-Q ultrasonics (Andreatta et al. 2005a, b), NMR diffusion (Freed et al. 2009), DC-conductivity (Zeng et al. 2009; Goual et al. 2011), centrifugation (Goual et al. 2011; Mostowfi et al. 2009), and AC-conductivity (Sheu et al. 2007). The CNAC is about 100 mg/liter in toluene. In addition, several of these methods that are sensitive to the change in Stokes drag upon nanoaggregate formation (NMR, DC-conductivity, centrifugation) verify that the aggregation number is quite small. Surface-assisted laser desorption ionization mass spectrometry (SALDI-MS) has been used to measure the aggregate mass, and knowing the molecular weight from L^{2}MS, the aggregation number is obtained. The resulting aggregation number for asphaltenes is approximately 7 (Wu et al. 2014a, b; Pomerantz et al. 2015). Combined small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) confirm the small size of the nanoaggregate (Eyssautier et al. 2011, 2012a, b). In addition, the divergence of the SANS and SAXS cross sections showed that the core of the nanoaggregate is enriched in aromatic carbon (high Z), while the periphery is enriched in hydrogen (lowest Z), thus alkane (Eyssautier et al. 2011, 2012a, b). The SAXS and SANS data is consistent only with the island molecular architecture; only island structures can exhibit an aromatic core and peripheral alkane. Archipelago molecular structures would place alkane carbon in the core and aromatic carbon in the periphery in contrast to measurement of asphaltenes.

With a further increase in concentration, nanoaggregates form clusters. The critical concentration of cluster formation (CCC) was shown from the aggregation kinetics (Anisimov et al. 1995; Yudin and Anisimov 2007), DC-conductivity (Goual
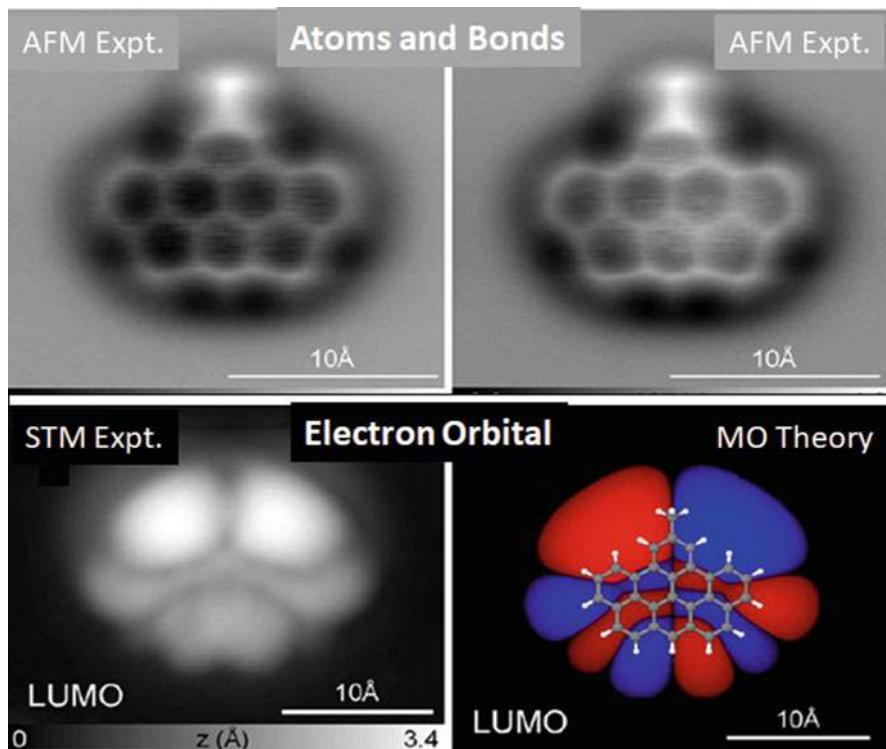
Fig. 9 Extraordinary experimental determination of the structure and a particular molecular orbital of an asphaltene molecule (Schuler et al. 2015). Top: atomic force microscopy (AFM) images of the asphaltene molecule with slightly different experimental parameters. Bottom: Left, scanning tunneling microscopy (STM) imaging of the lowest unoccupied molecular orbital (LUMO) of the asphaltene molecule. Right, molecular orbital theory of the same LUMO, excellent agreement is obtained (Schuler et al. 2015). Images of hundreds of molecules have been reported using these methods (Schuler et al. 2015, 2017a) for diverse asphaltenes. "Island" molecular architecture dominates, and the few archipelagos found are only with a direct aryl linkage between PAHs (Schuler et al. 2015, 2017a). In ten diverse asphaltene samples, not one traditional archipelago molecule is found with PAHs linked with alkane chains (Schuler et al. 2015, 2017a) even though such molecules can be readily transferred and imaged (Schuler et al. 2017b)

et al. 2014), and centrifugation (Goual et al. 2014); the CCC is roughly 3 g per liter in toluene. Small cluster aggregation numbers are shown in several NMR studies (Korb et al. 2013; Dutta Majumdar et al. 2013, 2015, 2017). In addition, the combined SAXS and SANS study showed the existence of a fractal cluster (Eyssautier et al. 2011, 2012a, b). All these techniques are consistent with small aggregation numbers of roughly eight nanoaggregates for the cluster. Detailed accounts of asphaltenes and several reviews of asphaltene science are available (Mullins 2011; Mullins et al. 2014).

The molecular structures and hierarchical nanocolloidal structures, the nanoaggregate and cluster, are represented in Fig. 10 known as the Yen-Mullins model
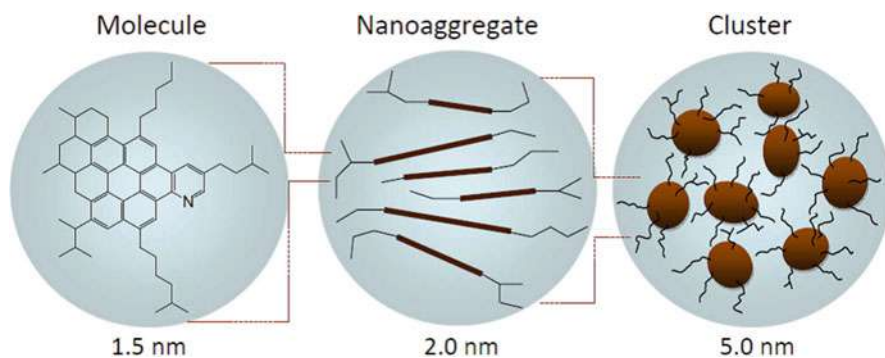
Fig. 10 The Yen-Mullins model of asphaltenes (Mullins 2010; Mullins et al. 2012). The dominant molecular structure with a single PAH ("island" architecture) is shown on the left. In dilute solution, asphaltenes are dispersed as a true molecular solution. At higher concentrations, asphaltenes form nanoaggregates with aggregation number of about 6. At yet higher concentrations, clusters of nanoaggregates form with an aggregation number of about 8

(Mullins 2010; Mullins et al. 2012). With the size of asphaltene structures resolved, the gravity and other terms can be developed in an equation of state for asphaltene gradients.

Hildebrand Solubility Parameter and the Hansen Solubility Parameters. The Hildebrand solubility parameter δ is a numerical value that indicates the relative solvency behavior of a specific solvent or solute. It is derived from the cohesive energy density of the solvent or solute, which in turn is obtained from the heat of vaporization ΔE per unit molar volume V.

$$\delta = \sqrt{\frac{\Delta E}{V}} \tag{1}$$

The Hildebrand solubility parameter can be decomposed into three constituent solubility parameters accounting for the three dominant intermolecular interactions of organic compounds: polarizability (known as the dispersion force) δ_D, polarity δ_P, and hydrogen bonding δ_H (Hansen 2005).

The PAH core of asphaltenes is fluffy and polarizable. Alkanes have lower polarizability. Teflon is even lower polarizability, thus is not at all chemically sticky. Helium atoms have almost no polarizability which is why helium does not liquefy until the temperature is 4 °K. Polarizability of helium atoms would incorporate the $n = 2$ electronic shell which has an excitation energy of 26 eV, thus negligible polarizability.

Polarity refers to molecules that have a permanent dipole moment. Carbon and hydrogen have nearly equal electronegativity, so compounds made of these two elements are nonpolar. Alkanes and aromatic compounds have no dipole moment with rare exception (such as azulene). In contrast, oxygen and hydrogen have very different electronegativity so bonds made of these two elements are quite polar

(e.g., water). Asphaltenes are dominantly carbon and hydrogen, and the dominant heteroatom is sulfur which is typically in nonpolar groups. Asphaltenes might contain a few percent nitrogen and oxygen, which give asphaltenes a bit of polarity.

H-bonding (or hydrogen-bonding) compounds refer to compounds that are involved with this special, strong bond. Water molecules form strong hydrogen bonds, and consequently water remains liquid at unusually high temperatures for compounds of such small mass. All compounds that form H-bonds are polar, but the converse is much less true.

$$\delta^2 = \sqrt{\delta_D^2 + \delta_P^2 + \delta_H^2} \tag{2}$$

The projection of the Hildebrand solubility parameter into three orthogonal Hansen solubility parameters is shown in Fig. 11 (Hansen 2005). In chemistry "like dissolves like"; the specific condition of two compounds being mutually soluble is that their Hildebrand solubility parameters should be similar. Moreover, each component of their respective Hansen solubility parameters should also be similar.

In Fig. 11, this condition of similar solubility parameters is graphically represented by a sphere centered on the solubility parameter of a given compound. Points inside the sphere represent similar solubility parameters to the given compound, thus are soluble. Points outside the sphere are insoluble in/with the given compound.

Green points Fig. 11 represents compounds soluble in the given compound (inside the sphere) with similar Hansen components. Red points represent compounds with very different Hansen components; thus they are insoluble with the given compound (Hansen 2005).

Fig. 11 Graphical representation of the Hildebrand solubility parameter (S) for a given compound projected into the three Hansen components, dispersion (D), polarity (P), and hydrogen bonding (H). The chemistry axiom "like dissolves like" dictates that only compounds within the sphere of radius R shown centered on S will be soluble with the given compound
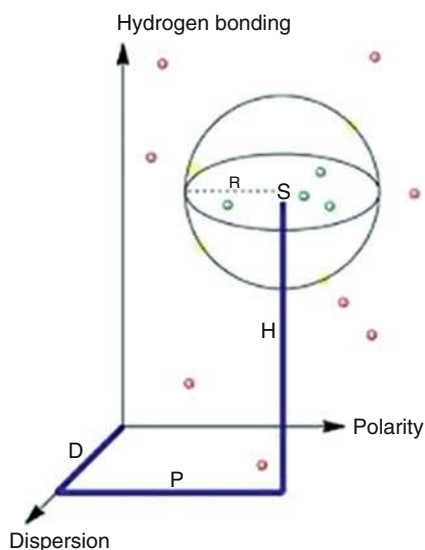
Table 1 Hildebrand and Hansen solubility parameters

Compound	Hildebrand δ (MPa$^{1/2}$)		Hansen parameter	
		$\delta_{Disp.}$	δ_{Pol}	δ_{H-bond}
Alkanes				
n-pentane	14.5	14.5	0	0
n-heptane	15.3	15.3	0	0
n-dodecane	16.0	16.0	0	0
Aromatics				
Benzene	18.6	18.6	0	2.0
Toluene	18.2	18.0	0	2.0
Naphthalene	20.3	19.2	2.0	5.9
Hetero compounds				
Acetone	20.0	15.5	10.4	7.0
Ethyl alcohol	26.5	15.8	8.8	19.4
Phenol	24.1	18.0	5.9	14.9
Acetic acid	21.4	14.5	8.0	13.5
Carbon disulfide	20.5	20.5	0	0.6
Acetonitrile		15.3	18.0	6.1
Teflon	12.4	–	–	–
C8F18	12.6	–	–	–
Water	47.8	15.6	16.0	42.3
Asphaltenes	20.4	19.5	4.7	4.2

A table of Hildebrand and Hansen solubility parameters is instructive. Teflon and other perfluorocarbons are known to have very low intermolecular interactions which is why Teflon is a preferred coating in cookware; nothing sticks to it. A value of 12.4 (MPa)$^{1/2}$ for Teflon in Table 1 is seen to correspond to such a low intermolecular interaction (not shown but almost entirely the polarizability component). In contrast, Table 1 shows that water with its strong intermolecular interaction has a Hildebrand solubility parameter of 47.8 (MPa)$^{1/2}$. Teflon and water establish the range of Hildebrand parameters from very weak 12 (MPa)$^{1/2}$ to very strong 48 (MPa)$^{1/2}$ for condensed phase. Moreover, water is dominated by H-bonding, a strong and spatially directional bond. Alkanes, the dominant component of crude oil, are seen to have Hildebrand parameters 16 (MPa)$^{1/2}$ somewhat above Teflon, and those of aromatics are somewhat above the alkanes. Moreover, Table 1 shows that the alkanes and the aromatics are dominated by the simple, weak polarizability forces.

The solubility parameters of the asphaltenes are very similar to those of aromatics. Indeed, asphaltenes are defined to be toluene soluble and thus must have solubility parameters similar to that of toluene. Note that the difference between the solubility parameters of asphaltenes and alkanes are sufficient to dictate no solubility. The solubility parameters of toluene and asphaltenes are dominated by the weak and simple polarizability component. The asphaltenes do have a component of

polarity and H-bonding, but these contributions are small especially considering it is the square of each component that contributes to the Hildebrand solubility parameter (Eq. 2).

Table 1 shows that the solubility parameter of asphaltenes is dominated by the polarizability term (Redelius 2005). This fact is often overlooked when elaborate supramolecular assemblies are proposed for asphaltenes with binding associated with hydrogen bonding, organometallic complexes, and the like. In fact, polarizability dominates consistent with a stack of PAHs (Eyssautier et al. 2011, 2012a, b; Mullins 2010; Mullins et al. 2012).

Methane has low intermolecular interaction but is not listed in Table 1 because the molar volume is very dependent on temperature and pressure making the numeric value of solubility parameter somewhat arbitrary (cf. Eq. 1). Nevertheless, addition of methane to crude oil results is a decrease in solubility parameter because both the density decreases and methane has low intermolecular interaction.

The Flory-Huggins-Zuo Equation of State. With this understanding, we can now examine the FHZ EoS. The FHZ is generally written as a ratio of ODs, optical densities, as measured by DFA tools. The ODs correspond to the strength of oil color at a specific wavelength. The color OD measurement is linear in asphaltene content. However, we do not assume a constant of proportionality, so the exact asphaltene content is not determined. For the OD ratio, the proportionality constant cancels; thus the FHZ EoS is written in terms of the OD ratio. The FHZ provides a thermodynamic expression for the DFA measured asphaltene gradients.

The FHZ EoS (below) takes into account three forces that result in asphaltene gradients: gravity, entropy, and solubility. The variables k, ϕ, v, δ, T, g, ρ, and h are Boltzmann's constant, volume fraction, molecular or particle volume, solubility parameter, temperature, earth's gravitational acceleration, density, and depth, respectively. Subscript a denotes the properties of asphaltenes; subscripts h_1 and h_2 stand for the properties at depths h_1 and h_2, respectively.

Gravitational gradients result from density difference or negative buoyancy between asphaltenes and the remainder of the oil, with relatively dense asphaltenes accumulating toward the bottom of the reservoir. This term contains Archimedes buoyancy $V_a \Delta \rho g$ where V_a is the volume of the asphaltene species from the Yen-Mullins model, $\Delta \rho$ is the density contrast between asphaltene and the oil, and g is earth's gravitational acceleration. Archimedes buoyancy multiplied by height h is the gravitational potential energy. The FHZ EoS includes the Boltzmann distribution for gravity, exp(-mgh/kT), where mgh (or more accurately $V_a \Delta \rho g h$) is the energy required to lift the asphaltene particle to height h from the bottom of the reservoir, and thermal energy kT is doing the lifting, k is Boltzmann's constant and T is temperature.

$$\frac{\phi_a(h_2)}{\phi_a(h_1)} = \exp\left\{ \frac{v_a g\,(\rho - \rho_a)\,(h_2 - h_1)}{kT} + \frac{v_a}{kT}\left[(\delta_a - \delta)_{h_1}^2 - (\delta_a - \delta)_{h_2}^2 \right] \right.$$
$$\left. + \left[\left(\frac{v_a}{v}\right)_{h_2} - \left(\frac{v_a}{v}\right)_{h_1} \right] \right\}$$

$$(3)$$

The solubility term utilizes the solubility parameter δ discussed above and depends on the difference in solubility parameters of the oil and asphaltene at different heights in the column. GOR can change for crude oils at different heights in the column, and this yields a difference in liquid-phase solubility parameter which can produce large changes in asphaltene concentration. Thus, for oils with high solution gas, asphaltene concentration will be low. For oils with large GOR gradients, such as equilibrated condensates and oil columns undergoing gas entry by diffusion, there will also be a large gradient in asphaltenes (or heavy ends).

The entropy term employs ratios of the molar volumes of crude oil and asphaltene at different heights. This term tends to be small so is often neglected. Essentially the entropy term is related to the statistical permutations of arranging solute and solvent.

The Flory-Huggins-Zuo equation of state (FHZ EoS) has been used in conjunction with the Yen-Mullins model in approximately 40 oilfield studies very successfully. This methodology will be reviewed in the next section. In addition, the Yen-Mullins model coupled with the (simple) Langmuir EoS has found significant utility associated with oil-water interfacial characterization (Rane et al. 2013, 2015). This capability in the more demanding arena of interfacial science lends credence to the related formalisms applied to bulk crude oil.

4 Reservoir Evaluation and Reservoir Fluid Geodynamics

Reservoir fluid geodynamics (RFG) is the name given to account for processes which hydrocarbon reservoir fluids undergo after (and during) reservoir trap filling (Mullins et al. 2017). ("Geodynamics" has been used by geologists to describe evolution of the formations (Miall 2000), and "fluid geodynamics" has been used to account for magma flows in the mantle within geodynamic treatments.) RFG includes processes long studied by geochemists including biodegradation and water washing (Peters et al. 2005). Nevertheless, even for these processes, use of equation of state modeling greatly clarifies a detailed spatial accounting of fluid properties and their variations within the reservoir. In particular, incorporation of the FHZ EoS and its reliance on the Yen-Mullins model have been very revealing of all manners of complexities in reservoir as established in many reservoir case studies. Figure 12 shows a schematic of a timeline with different disciplines indicated. Petroleum system modeling provides the timing, type, and volume of fluids entering the reservoir. Reservoir fluid geodynamics accounts for the many processes that can redistribute and alter the fluids and cause phase transitions such as tar formation. Simulation modeling then accounts for production in a present-day setting.

Compartmentalization and Connectivity. One of the most important concerns is the extent of reservoir "connectivity" or its inverse, compartmentalization (Mullins 2008). A compartment must be intersected by a well for drainage. A reservoir might consist of one or a few large compartments. For reservoirs that exhibit good connectivity, few wells are required. For reservoirs consisting of many small compartments, many wells are required for drainage. When well costs are high, compartmentalized reservoirs can be uneconomic. An industry study
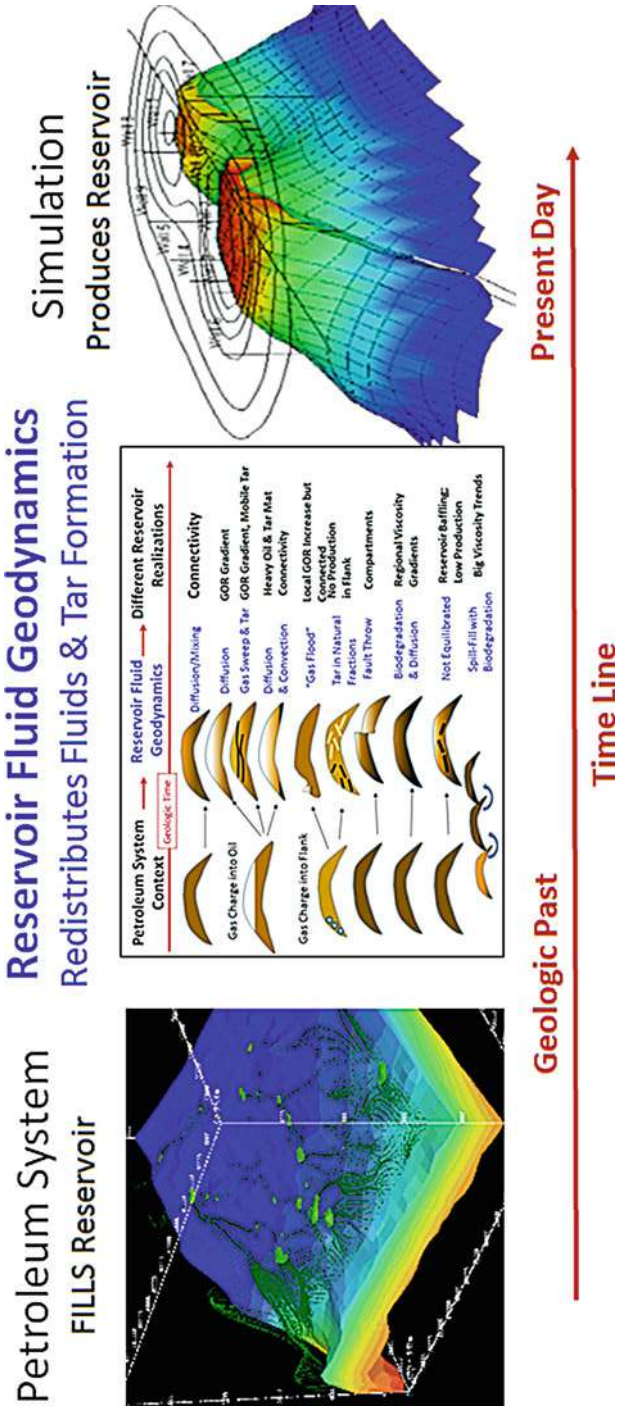
Fig. 12 A timeline showing the different disciplines. Petroleum system modeling accounts for the timing, type, and volume of fluids that enter the reservoir. Reservoir fluid geodynamics, a newly formalized discipline, accounts for the many processes that redistribute fluids within the reservoir and that account for phase transitions such as tar formation (Mullins et al. 2017). Simulation then accounts for reservoir production in present day

concluded that 75% of deepwater reservoirs in the Gulf of Mexico underperformed in both rate of production and ultimate oil recovery, primarily due to unrecognized compartmentalization (Mishra et al. 2014). A large variety of sealing barriers separating compartments can be present in oilfield reservoirs. Some barriers correspond to (sealing) faults that are observed in seismic imaging. Seismic imaging is inherently low resolution so that layers or objects that are much smaller than 10 m thick are difficult to resolve. Other sealing barriers are not only invisible to seismic imaging, they can even be invisible to wireline petrophysical logging measurements performed in oil wells which often have a resolution of centimeters (Andrews et al. 2001). Indeed, such "invisible" barriers can even hold off 2000 psi of depletion pressure differential (Andrews et al. 2001).

Compartmentalization. One method to reveal compartmentalization is to perform pressure surveys. In an oil column, the pressure increases with depth according to the fluid density, so a pressure gradient is always obtained. In a new, unproduced field, if two formations are not in pressure communication, then they are not in flow communication and represent two separate compartments. Similarly, if a well intersects two oil-bearing formations, and the oil in the upper formation has a higher concentration of asphaltenes, then the formations are likely not connected as asphaltenes are negatively buoyant in crude oil.

Figure 13 shows an example of a highly compartmentalized series of oil-bearing sandstone layers (sands) in a single well. Pressure measurements show Sand A is isolated from Sand B, Sand C is isolated from Sand D, and Sand E is isolated from Sand F. DFA color measurements show the oil in Sand A has three times the asphaltene content as the oil in Sand B below it; clearly these are not connected. The oil in Sand B has more asphaltene than the oil in Sand C. The asphaltene difference is large between Sand C and Sand D, so lack of connectivity is plausible. Sand D has more asphaltene than Sand E below it; thus they are isolated. Combining the pressure and DFA color data shows the sands are all isolated from each other; the oil company ceased all activities in the field due to this conclusion (Mullins et al. 2007c, 2008b).

Dead crude oils were obtained from most of these zones to look for detailed compositional differences. No significant differences could be detected in the heavy ends using ultrahigh-resolution mass spectroscopy (Mullins et al. 2007c), nor in the liquid phases using comprehensive two-dimensional gas chromatography (GC×GC) (Mullins et al. 2008b). The crude oils have very similar chemical components but differ mostly in their asphaltene fractions (Mullins et al. 2007c, 2008b). That is, the chemical constituents in each phase are very similar as shown by high-resolution analytical chemistry. What differs is the overall concentrations of the different gas, liquid, and solid constituents; DFA measurements are well suited to analyze these fractions. With DFA measurements, there is minimal sample handling reducing sample transfer problems that can arise, especially in remote, offshore locations. In addition, the DFA tools within a well have the same time, operating engineer, temperature, calibration, and baselines, such that many systematic errors cancel in establishing fluid gradients.
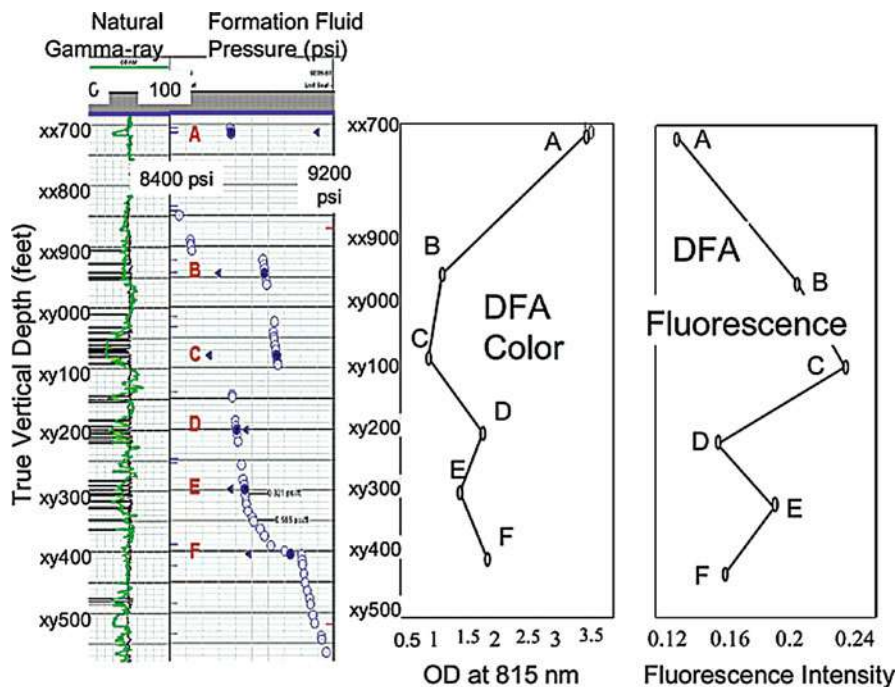
Fig. 13 Well log data from deepwater Gulf of Mexico (Mullins et al. 2007c, 2008b). The exact depth is confidential, thus the leading xy in depth. Six stacked sandstone formations are intersected (A,B,C,D,E and F) with intervening shale zones with slightly higher gamma ray signal (green curve). Formation pressures with gradients are shown. DFA-measured coloration at 815 nm for six sampling points is shown. Fluorescence is also shown and is higher for crude oils of lower color (asphaltene content). Both pressure and DFA color measurements show significant compartmentalization; the oil company ceased any activity in this reservoir due to this conclusion

Nevertheless, the more important issue is to identify the good news of connectivity (better economic value), not just the bad news of compartmentalization (lower economic value). While pressure measurement can conclusively demonstrate the bad news of compartmentalization, it is not definitive to establish connectivity. Pressure connectivity is a necessary but insufficient condition to establish flow connectivity in production time frames (years to 10 years) (Mullins 2008). Indeed, in the industry-wide study showing the huge problem with unrecognized compartmentalization in deepwater Gulf of Mexico, all of the fields had extensive pressure surveys (Mishra et al. 2014); pressure measurement is grossly inadequate to establish reservoir connectivity. The cause of the shortcoming is simple: very little mass transfer between two compartments is needed to equilibrate pressure. A leaky seal between two compartments can act as a conduit for a little mass transfer equilibrating pressure in geologic time, but leaky seals are not adequate to handle significant oil flow in production time.

Connectivity and Equilibrated Asphaltenes. The equilibration of asphaltenes in a reservoir requires extensive mass flow, likely both during and post trap filling. Consequently, if asphaltenes are equilibrated across a field, the field is likely connected. Many oilfield case studies have established that equilibrated asphaltenes imply connectivity in a production time frame (Mullins 2008; Zuo et al. 2013). Figure 14 shows examples of equilibrated asphaltenes involving all three structures in the Yen-Mullins model as well as heavy resins for a light condensate (Zuo et al. 2013).

Figure 14a shows a heavy-end gradient in a light condensate (Zuo et al. 2013). In this case, the DFA color measurement was close to zero throughout the visible spectral range; thus, the DFA fluorescence measurement was used (Zuo et al. 2013). For low concentrations of fluorophores and chromophores, the fluorescence intensity is linear in concentration of fluorophores, which in this case are the heavy ends of a resin. For another light condensate, the predominant fluorophore was identified as perylene (Juyal et al. 2011), and the corresponding molecular dimension of perylene is used in the FHZ EoS analysis in Fig. 14a. Indeed, perylene fluorescence gave this oil a blue appearance under normal illumination; a photograph of this condensate under visible illumination is shown in Fig. 14a as an inset.

Fig. 14b is a heavier condensate with a molecular dispersion of asphaltenes. The asphaltenes across the field were shown to be equilibrated; however, the dissolved gas was *not* equilibrated across the field. The reservoir had two separate anticlinal structures, each with a gas cap and with different GOCs. The field was put into production and was proven to be connected; equilibrated asphaltenes again are associated with reservoir connectivity. The GOR was not equilibrated due to a gas charge (gas and oil charge into reservoir that has previously) filling different amounts of gas into the two anticlines. Equilibration of the two GOCs by diffusive methane transport across the field is a very slow process. In contrast, the asphaltene content of saturated crude oils is not impacted by further gas-phase addition to the reservoir. (Saturated crude oils are at their bubble point and cannot hold more solution gas.) Figure 14c, d show two oilfields with black oils; black oils have moderate GORs and moderate asphaltene content. Black oils almost always have asphaltenes dispersed as nanoaggregates (Zuo et al. 2013). Fig. 14e shows a heavy oil; heavy oils have asphaltenes dispersed as clusters (Zuo et al. 2013).

Asphaltene 2 nm Nanoaggregate Equilibration Across a Reservoir. A recent reservoir case study, deepwater Gulf of Mexico, allows a close look at asphaltene equilibration (Chen et al. 2017). Two wells were drilled in a middle Pliocene oilfield, and DFA data was acquired in two stacked sandstone formations (cf Fig. 15). Samples were acquired and extensive laboratory measurements were made (Chen et al. 2017).

Figure 15 shows that the asphaltene gradients in each sand are consistent with an equilibrium distribution of asphaltene nanoaggregates; the asphaltene gradients match the FHZ EoS with asphaltene nanoaggregates. Moreover, the two asphaltene gradients are offset from each other except at the top of the reservoir. The implication of the offset is that the intervening shale layer is laterally extensive but
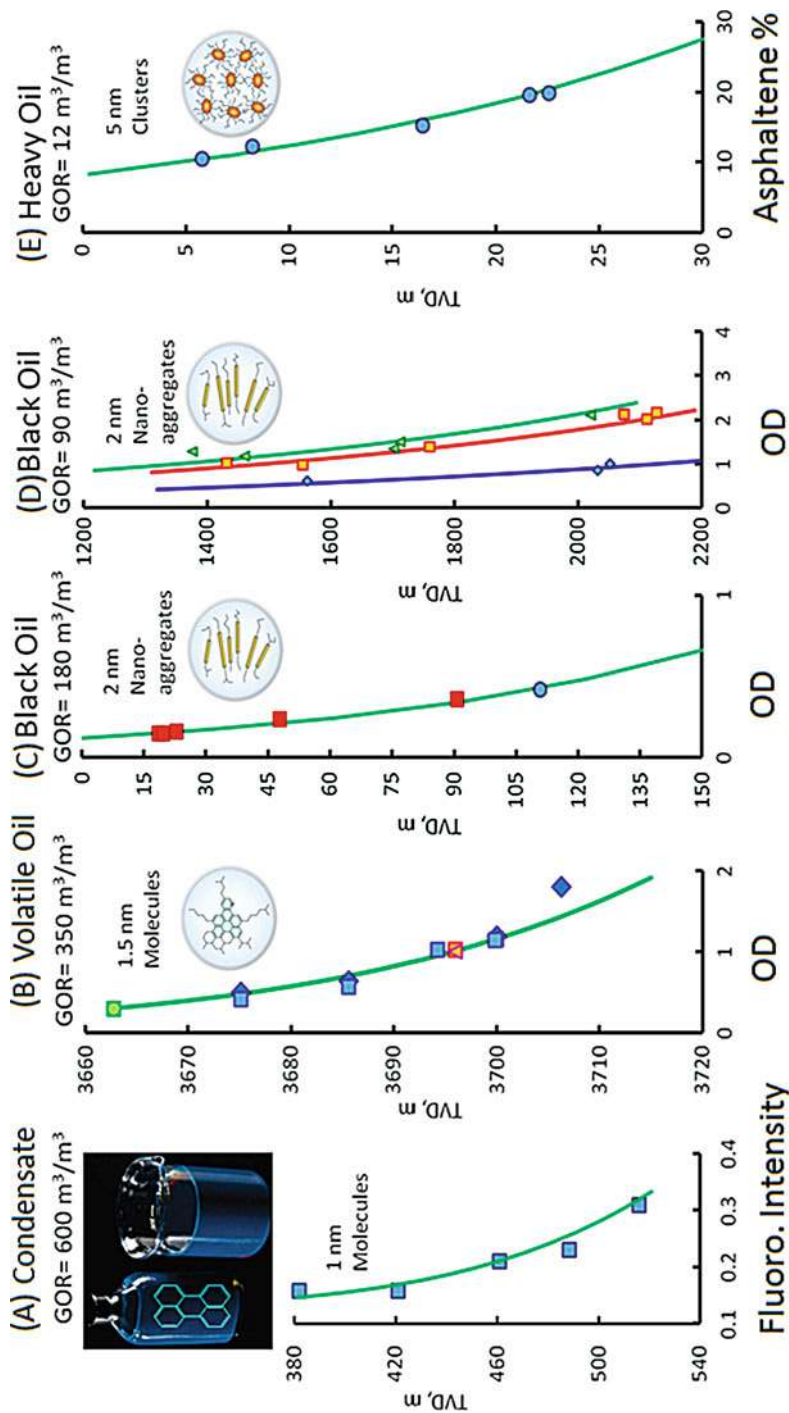
Fig. 14 Equilibrated asphaltenes imply reservoir connectivity as shown in five oilfields examples; in all cases, production proved the connectivity prediction to be correct (Zuo et al. 2013). These five examples also involve the three species of the Yen-Mullins model as well as heavy resins for a light condensate
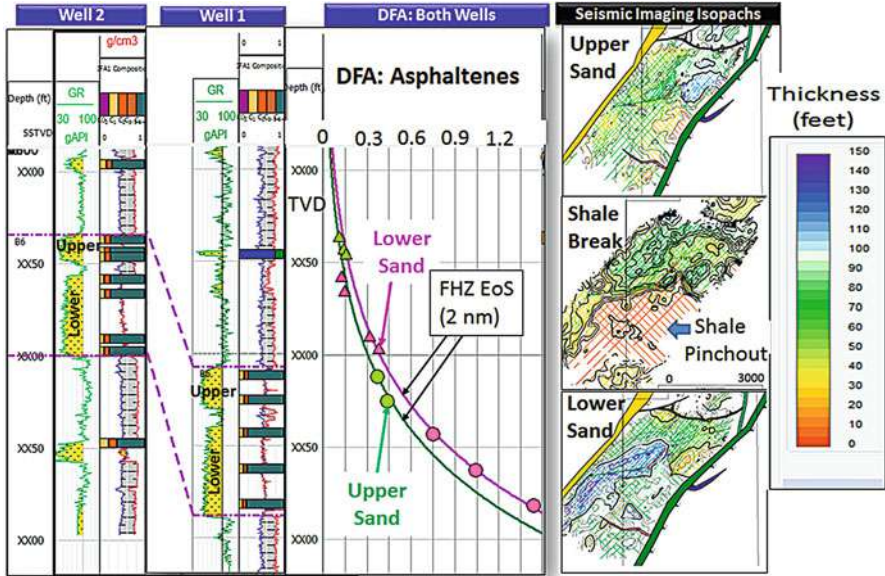
Fig. 15 A deepwater reservoir with two stack sands, for the most part separated by an laterally extensive shale layer (Chen et al. 2017). Two wells were drilled in this reservoir separated by roughly 1 km. Track 7 (DFA asphaltenes) shows each sand has equilibrated asphaltenes implying flow connectivity across the field. The two asphaltene curves are offset indicating the intervening shale layer is laterally extensive. Track 1, depth (the leading xx's are to conceal the absolute reservoir depth). Track 2, the natural gamma ray log (GR) of Well 2 which identifies sand formations (yellow). Track 3, horizontal bars with locations of DFA and sampling depths in Well 2. Green is oil, blue is water, yellow is methane, orange is other hydrocarbon gases. Track 4, the natural gamma ray log (GR) of Well 1 which identifies two sand formations (yellow). Track 5, horizontal bars with locations of DFA and sampling depths in Well 1. Also shown are the seismic "isopach" images showing the thickness over an extended lateral area of the upper sand, intervening shale layer, and the lower sand (Chen et al. 2017)

possibly pinches out (disappears) near the crest of the reservoir. This is consistent with the thickness "isopach" models of the shale shown in Fig. 15 (Chen et al. 2017).

Production history matching and reservoir simulations of production for 12 months of production are consistent with all conclusions from the asphaltene gradient analysis (Chen et al. 2017). Production is considered ultimate arbiter determining the connectivity profile of reservoirs (Chen et al. 2017). The agreement in this reservoir between predictions based on thermodynamics and 1 year of production is a strong validation for the FHZ EoS and the Yen-Mullins model.

Figure 16 shows that the fluid gradients of GOR, fluid density, and viscosity are consistent with the interpretation of equilibrated asphaltene gradients. However, these other fluid gradients are not as clearly identified as being equilibrated as the asphaltene gradients; often the asphaltene gradients convey the most useful information. The thermal maturity biomarker ratio Ts/(Ts + Tm) shows no gradient
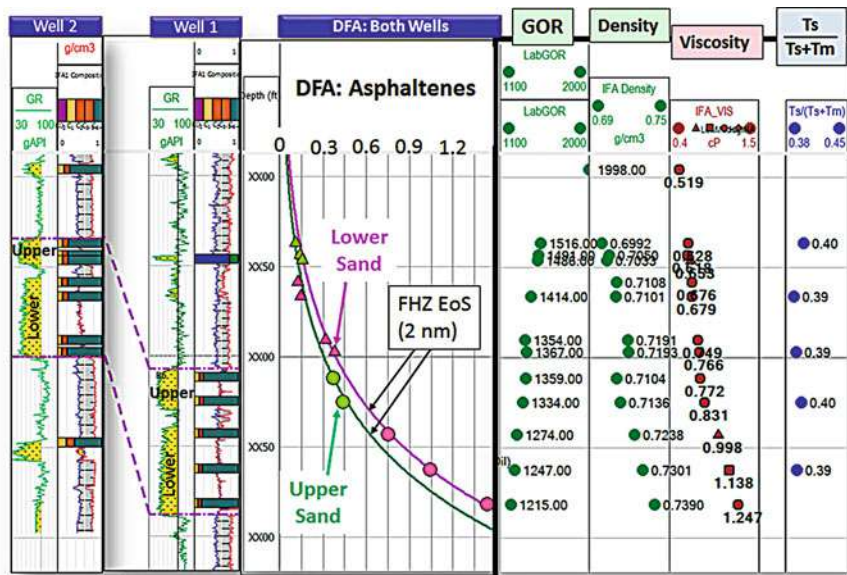
Fig. 16 The same reservoir as depicted in Fig. 15. Tracks 1–6 as described in figure caption for Fig. 15 (Chen et al. 2017). Fluid measurements of Track 7, GOR (from the lab); Track 8, IFA-measured density; Track 9, IFA-measured viscosity. Track 10, thermal maturity biomarker ratio Ts/(Ts + Tm) (Chen et al. 2017)

(Chen et al. 2017). This oil reservoir appears to have had a single charge of oil, thus, lacking various fluid complexities and providing a somewhat simplified case of testing asphaltene thermodynamic modeling (Chen et al. 2017).

Asphaltene 5 nm Cluster Equilibration Across an Oil Reservoir. When asphaltene concentrations are high, asphaltenes are dispersed as clusters, not nanoaggregates. An example of this is observed in a giant reservoir in Saudi Arabia. The reservoir is a four-way sealing anticline – a turtle shell structure (see inset in Fig. 17 with eight well locations). There is a black oil of a few percent asphaltene content in most of the field. At the bottom of the oil column, there is a 60 m vertical heavy oil rim around the entire 100 km periphery of the oilfield. This heavy oil is underlain by a 10 m thick tar mat; this represents a massive accumulation of asphaltenes. Figure 17a shows the asphaltene content as a function of depth in this oil column.

Figure 17 shows in dramatic fashion consistency of the heavy oil gradient matching the FHZ EoS with asphaltene clusters over a 100 km perimeter of the oilfield with a factor of ten variation of asphaltene concentration. Similar heavy oil gradients have been observed elsewhere (Mullins et al. 2013). Moreover, the implication of asphaltene equilibration is reservoir connectivity over this large field, and that is known from production (Mullins et al. 2013). As expected from successful application of the Yen-Mullins model and the FHZ EoS, there is no
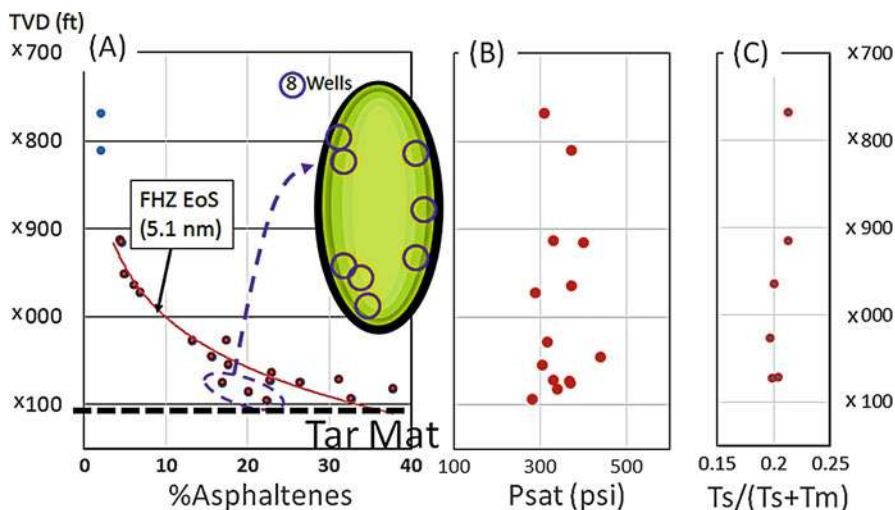
Fig. 17 (**a**) Asphaltene concentration versus true vertical depth (TVD) in the heavy oil rim of a large oilfield. The asphaltene gradient matches the FHZ EoS with 5.1 nm clusters (vs. 5.0 nm nominal) with no other adjustable parameters around the 100 km perimeter of the oilfield. The equilibrated asphaltene cluster gradient implies connectivity which has been confirmed in production (Mullins et al. 2013). (**b**) There is no record of this huge asphaltene gradient in the saturation pressure (Psat). The saturation pressure is low and invariant. (**c**) The thermal maturity biomarker ratio Ts/(Ts + Tm) varies only slightly; the asphaltene gradient is not due to a thermal maturity gradient

variation in the asphaltene chemistry with regard to molecular weight (Pomerantz et al. 2013), nanoaggregate aggregation number (Pomerantz et al. 2013), or sulfur chemical speciation throughout the heavy oil column (Wu et al. 2014b). The only way to model this giant asphaltene gradient with an invariant liquid phase is to use a gravitational gradient of asphaltene clusters as given by the FHZ EoS (Mullins et al. 2013; Pomerantz et al. 2013; Wu et al. 2014b; Forsythe et al. 2015).

The heavy oil in the rim in Fig. 17 did not arise as a thermal maturity variation; the ratio of the thermal maturity biomarkers Ts/(Ts + Tm) is almost invariant, and there is no biodegradation (Forsythe et al. 2015). This massive accumulation of asphaltenes giving rise to the 60 m vertical heavy oil column underlain by a 10 m tar mat arose from instability of incompatible charges into the reservoir followed by asphaltene migration to the base of the reservoir. With such a large accumulation of asphaltenes, it is fairly easy to detect this occurrence in reservoir evaluation. Nevertheless, the asphaltene instability did not remove all asphaltenes from the oil.

Validation of Reservoir Fluid Geodynamics. Another oilfield study showed a variety of complexities associated with a late gas charge into oil reservoirs. Indeed, this study proves the validity of RFG because three reservoirs exhibit totally different reservoir realizations even though they all had the same petroleum system process of a gas charge into oil (Achourov et al. 2015; Pfeiffer et al. 2016, 2017).
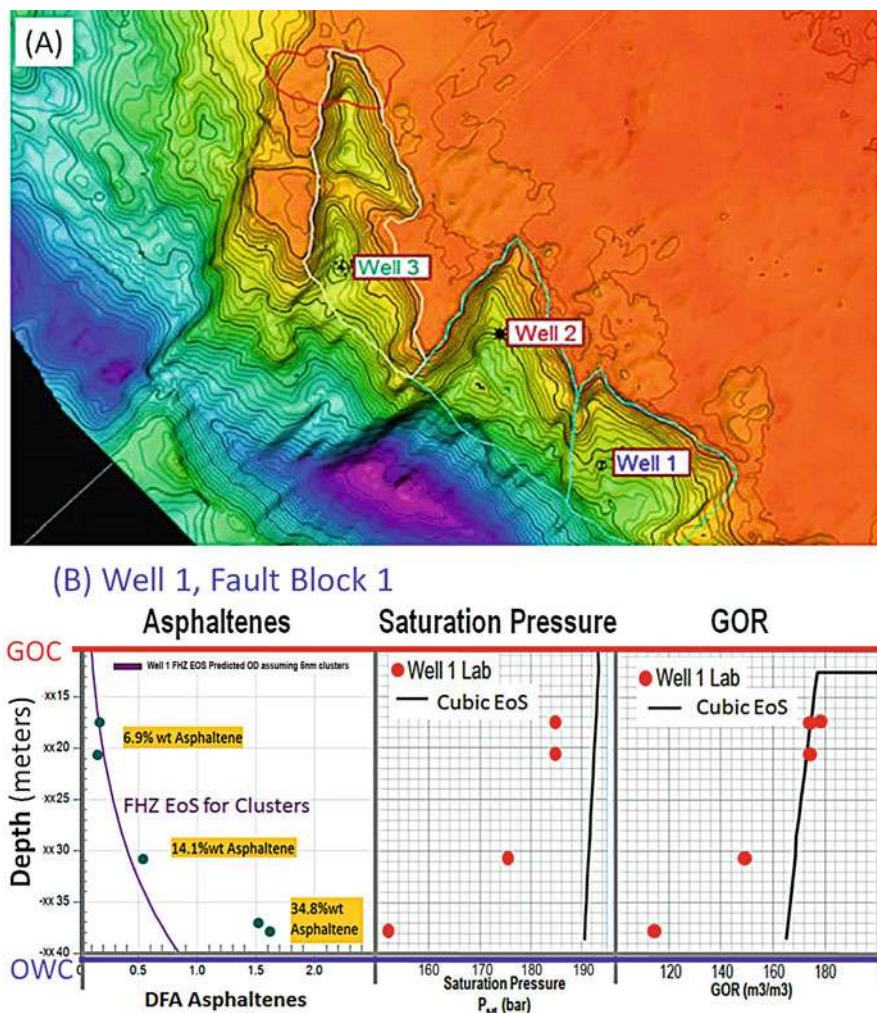
Fig. 18 (**a**) Areal contour map of the Norwegian continental shelf showing three adjacent fault blocks, each with one well (Achourov et al. 2015; Pfeiffer et al. 2016). (**b**) The vertical well in Fault Block 1 has both a gas-oil contact (GOC) and an oil-water contact (OWC); thus, all fluid migration is vertical and measurable within this well. The asphaltene gradient is very large and in disequilibrium; the data points do not match the FHZ EoS. In addition, the asphaltene content in the oil at the base of the column is 35%; this is a heavy oil. The saturation pressure (Psat) and GOR exhibit large, disequilibrium gradients; they do not match the cubic EoS

Fig. 18a shows the areal contour map on the continental shelf of Norway of three adjacent fault blocks, each with one well. Fault blocks are very large blocks of rock, often many kilometers in extent, created by tectonic and localized stresses in the Earth's crust. Fault blocks are often seen in seismic imaging.
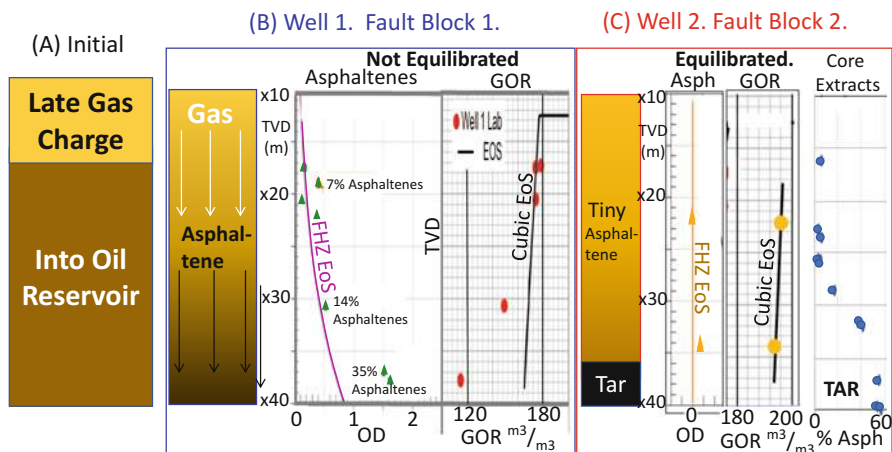
Fig. 19 (**a**) The initial condition after charge of a density-stacked gas on top of an undersaturated oil (Achourov et al. 2015; Pfeiffer et al. 2016). (**b**) Fault Block 1, schematic of gas diffusing into oil increasing GOR at the top of the oil column; this causes asphaltenes migration to the base of the column. (**c**) Schematic of the completion of the RFG processes of gas diffusion and asphaltene migration down. The solution gas is high and equilibrated in Well 2. The asphaltene content is very small and equilibrated in Well 2. The asphaltenes migrated to the base of the oil column and underwent phase separation forming a tar mats seen in analysis of core extracts (Achourov et al. 2015; Pfeiffer et al. 2016, 2017)

Figure 18b shows that the oil column is in disequilibrium in Well 1 (Achourov et al. 2015; Pfeiffer et al. 2016). The asphaltene gradient is very large and does not match the FHZ EoS with clusters; there are no larger stable colloidal particles in crude oil. In addition, the saturation pressure and GOR also exhibit large, disequilibrium gradients. The crude oil at the bottom of the column has a high asphaltene concentration of 35%; thus is heavy. Well 1 has both a GOC and an OWC; thus, all fluid migrations are vertical and observable within this vertical well. This contrasts the frequent occurrence of tilted reservoirs, where the GOC is not only vertically displaced but also laterally displaced from the OWC so vertical wells often insect only one of these contacts. The issue arises as to why this column is in disequilibrium considering the very small vertical distances that are involved, just tens of meters. In all three fault blocks, that late gas charge is thought to be Pleistocene, limiting the time duration for establishing equilibrium. To address this question further, it is best to also consider the adjacent Fault Block 2.

"Movie" of Tar Mat Formation. Figure 19 shows that Well 1 is undergoing the process of gas diffusion into the oil column from the top, while the asphaltenes are migrating down away from high solution gas oil (Achourov et al. 2015; Pfeiffer et al. 2016). Well 2 shows the completion of this process in geologic time. The asphaltenes migrated to the base of the column increasing their concentration. Gas diffused to the base of the column increasing solution gas. Both fault blocks had this process initiate at about the same time; the question arises what slowed
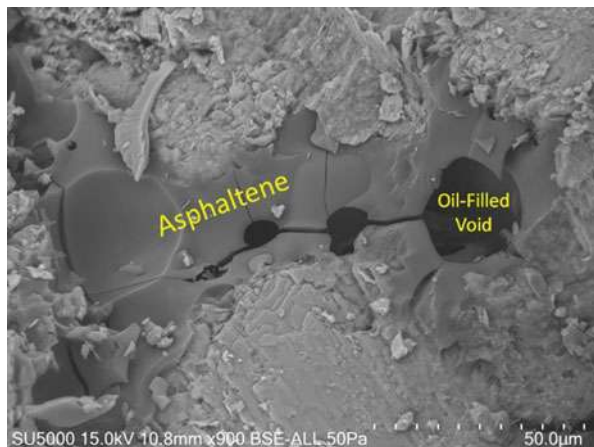
Fig. 20 Scanning electron micrograph obtained from the tar zone in Well 2 (Pfeiffer et al. 2017). The asphaltene-enriched phase is a smooth, cracked material that is conformally attached to rock surfaces. Void space that is oil filled in the reservoir is also evident. The tar mat consists of two organic phases; once the asphaltene phase seals sufficient numbers of pore throats, permeability becomes negligible (Pfeiffer et al. 2017)

equilibration in Fault Block 1. The answer is obtained in measurement of both vertical permeability and oil production rates (Achourov et al. 2015; Pfeiffer et al. 2016). Well 1 was shown to have much lower vertical permeability by pressure interference testing (Pfeiffer et al. 2016). In addition the production rates in Well 1 are ten times lower than Well 2 after accounting for viscosity differences of the oils (Pfeiffer et al. 2016).

Most of the asphaltenes at the base of the oil column in Well 2 have undergone phase separation. Figure 20 shows an SEM from this tar zone showing an asphaltene-enriched phase and a void space that is filled by oil in the reservoir; the tar mat consists of two organic phases (Pfeiffer et al. 2017). The asphaltene phase adheres conformally to the rock surface which is consistent with heterogeneous nucleation. However, asphaltenes are nanocolloidal with huge surface area; thus, there is no standard homogeneous nucleation difficulty associated with formation of new surface energy in single-phase true molecular solutions. Instead, a plausible explanation for the heterogeneous nucleation observed here is due to the slow kinetics of reaction-limited aggregation known for floc formation from asphaltene clusters versus the relatively rapid rate of collisions with pore walls.

Figure 21 shows the application of a single RFG model to Wells 1 and 2 (Zuo et al. 2017). The single model accounts for the disequilibrium gradients of both GOR and asphaltenes in Well 1. The same RFG model also accounts for the equilibrated GOR and asphaltene gradients as well as the thick tar mat in Well 2. The major difference between the wells (and fault blocks) is the vertical permeability in the model. Well 1 is known to have low vertical permeability, and Well 2 is known to have higher vertical permeability in accordance with vertical
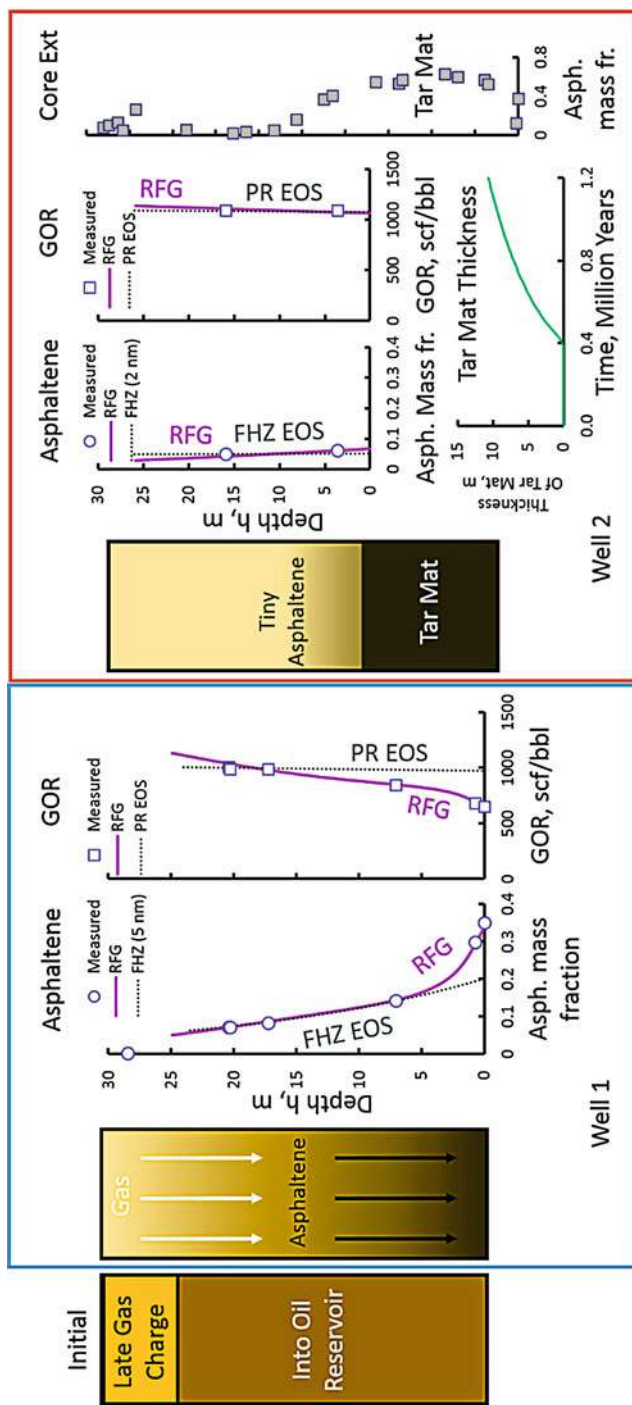
Fig. 21 A single RFG model accounts for present-day fluid and tar distributions in two adjacent fault blocks in an oilfield (cf. Fig. 19) (Zuo et al. 2017). The initial condition of a late gas charge into an undersaturated oil applies to both fault blocks. The RFG model incorporates gas diffusion into the oil, asphaltene migration away from high solution gas downward in the oil column, and asphaltene phase change. Well 1 is in the middle of these processes in geologic time, thus not equilibrated. Well 2 shows the final stage of this process (Zuo et al. 2017)

pressure interference testing (Achourov et al. 2015; Pfeiffer et al. 2016). Moreover, much higher production rates were obtained from Well 2 than Well 1 in accord with the higher permeability of Well 2 (Achourov et al. 2015). From a production standpoint, this RFG model is addressing many important concerns including the spatial distribution of asphaltenes and solution gas, thus oil viscosity and the presence or absence of a tar mat which has a huge impact on pressure support and aquifer sweep. Understanding the evolution of the fluid column greatly improves the understanding of present-day fluid and tar distributions (Zuo et al. 2017). This understanding is essential for optimal field development planning.

Well 3 in Fault Block 3 in this oilfield exhibits a very complex behavior of deposited asphaltenes. At this point, modeling can only account for the observations above the shale break as shown in Fig. 22.

Figure 22 shows a complex distribution of deposited asphaltenes for Well 3 in Fault Block 3 (Pfeiffer et al. 2016). The fluids and pressures are equilibrated above and below the shale layer; thus, it is likely the shale is not a sealing layer but is a baffle and pinches out away from the wellbore. Above the shale layer, the late charge of gas density-stacked above the liquid and diffused down into the oil column. This
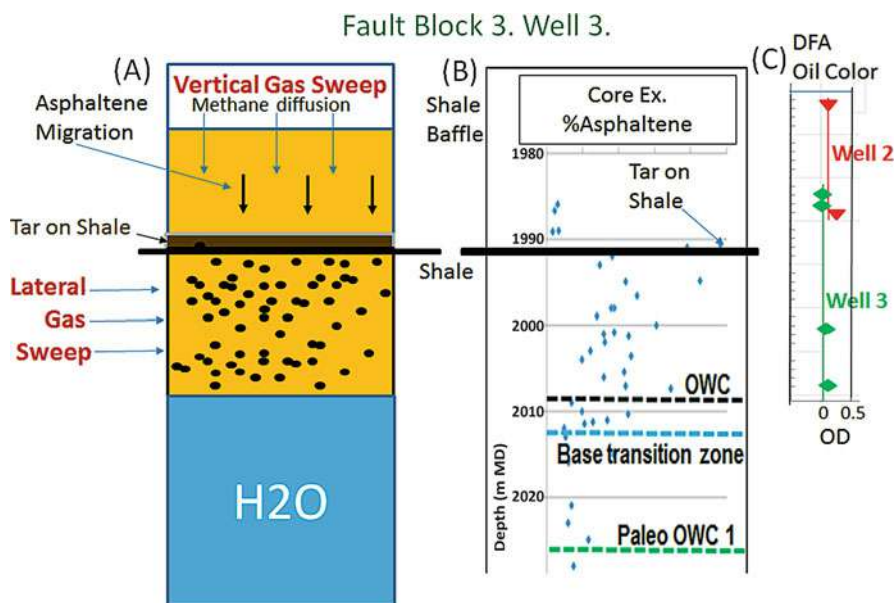
Fig. 22 Well 3 log data (Pfeiffer et al. 2016). (**a**) Schematic of the processes that led to complex asphaltene deposition observed in core extracts. Above the shale baffle, the same process occurs as in Well 2. Methane diffused down driving the asphaltene onto the shale baffle as a phase-separated material. Below the shale baffle, methane diffusion from above is impeded, and lateral gas sweep causes local deposition of asphaltene throughout the core with no tar mat at the base (Pfeiffer et al. 2016). (**c**) Little asphaltene remains in the oil due to the high GOR (~1200 scf/bbl). The asphaltenes, GOR (not shown), and pressure (not shown) are equilibrated from top to bottom in the oil column; the shale layer is a baffle, not a seal (Pfeiffer et al. 2016)

caused the asphaltenes to migrate onto the shale break where they became trapped (ponded as in a shale bowl) and underwent phase separation to form a thin tar mat. The tar mat is thin because there is only a small oil column above this shale break; thus little asphaltene was available to accumulate (Pfeiffer et al. 2016).

Below the shale break, gas could not diffuse from above. In this section of the well, the asphaltene is deposited throughout the column; moreover, there is no tar mat because there was no vertical migration of asphaltenes. This has been attributed to lateral gas sweep into the reservoir (Pfeiffer et al. 2016). Lateral gas sweep means that a significant lateral variation of fluid properties exists within a permeable formation. The origin of such variations is often associated with recent entry of new hydrocarbons, potentially gas, into the reservoir from the flank of a field (Achourov et al. 2015; Pfeiffer et al. 2016; Uchytil et al. 2016). With lateral migration, the asphaltenes cannot migrate downward away from the gas front; they become trapped locally and undergo phase transition. Lateral sweep of charge fluids has been incontrovertibly established by a black oilfield with a late gas charge (Uchytil et al. 2016). Lateral sweep of charge fluids is particularly likely if the well is located near the charge point of the reservoir, as is the case here for Well 3 (Pfeiffer et al. 2016; Uchytil et al. 2016).

These three adjacent fault blocks all experienced a late gas charge into oil reservoirs. However, the post charge evolution of these fault blocks led to three entirely different reservoir realizations in present day. Major production concerns of oil production rate, aquifer sweep, and type of produced fluid all depend on these different post charge processes and which are all described within a reservoir fluid geodynamics framework.

5   Conclusions

In this chapter, we have described the combination of new technology (downhole fluid analysis) coupled with new science (the first equation of state for asphaltene gradients) which have enabled the recent formalization of a new technical discipline "reservoir fluid geodynamics." Several oilfield case studies are presented elucidating major production concerns within a universal RFG workflow. Einstein stated "the most incomprehensible thing about the universe is that it is comprehensible." In our own microcosm of oilfield reservoirs, the same sentiment applies. We have shown here that measurements of asphaltene nanostructures apply to reservoirs at a length scale of 100 km. Laboratory measurements of asphaltene molecular diffusion at the nanosecond time scale apply to asphaltene diffusion in reservoirs in the tens of million years' time frame. The applicability of this extrapolation over roughly 14 orders of magnitude in both space and time dimensions is unexpected and remarkable. Reservoir fluids are not stochastic in nature; they are well behaved. The Flory-Huggins-Zuo equation of state is basically simple in that it employs a single "chemical" parameter, the Hildebrand solubility parameter each for the solute (asphaltenes) and solvent (live crude oil). The FHZ EoS is used with the specification of the asphaltene particle size from the Yen-Mullins model. The straightforward

approach enables broad understanding of many reservoir concerns over great space and time. The combination of this thermodynamic approach with exacting analytical chemistry methods and geochemical interpretation produces robust results. When new technology is combined with new science, great accomplishments come into view.

References

Achourov V, Pfeiffer T, Kollien T, Betancourt SS, Zuo JY, di Primio R, Mullins OC (2015) Gas diffusion into oil, reservoir baffling and tar mats analyzed by downhole fluid analysis, pressure transients, core extracts and DSTs. Petrophysics 56(4):346–357

Andreatta G, Bostrom N, Mullins OC (2005a) High-Q ultrasonic determination of the critical nanoaggregate concentration of asphaltenes and the critical micelle concentration of standard surfactants. Langmuir 21:2728

Andreatta G, Goncalves CC, Buffin G, Bostrom N, Quintella CM, Arteaga-Larios F, Perez E, Mullins OC (2005b) Nanoaggregates and structure-function relations in asphaltenes. Energy Fuel 19:1282–1289

Andrews RJ, Beck G, Castelijns K, Chen A, Cribbs ME, Fadness FH, Irvine-Fortescue J, Williams S, Hashem M, Jamaluddin A, Kurkjian A, Sass B, Mullins OC, Rylander E, Van Dusen A (2001) Quantifying contamination using color of crude and condensate. Oilfield Rev 13:24–43

Anisimov MA, Yudin IK, Nikitin V, Nikolaenko G, Chernoutsan A, Toulhoat H, Frot D, Briolant Y (1995) Asphaltene aggregation in hydrocarbon solutions studied by photon correlation spectroscopy. J Phys Chem 99(23):9576–9580

Bartha A, De Nicolais N, Sharma V, Roy SK, Srivastava R, Pomerantz AE, Sanclemente M, Perez W, Nelson RK, Reddy CM, Gros J, Arey JS, Lelijveld J, Dubey S, Tortella D, Hantschel T, Peters KE, Mullins OC (2015) Combined petroleum system modeling and comprehensive two dimensional gas chromatography to improve understanding of the crude oil chemistry in the Llanos Basin, Colombia. Energy Fuel 29(8):4755–4767

Chen L, Forsythe JC, Wilkinson TW, Winkelman B, Meyer JA, Canas JA, Xu W, Zuo JY, Betancourt SS, Lake P, Mullins OC (2017) A study of connectivity and baffles in a deepwater Gulf of Mexico reservoir linking downhole fluid analysis and geophysics. In: SPE ATCE-MS 187231, San Antonio. https://doi.org/10.2118/187231-MS

Dong C, Mullins OC, Hegeman PS, Teague R, Kurkjian A, Elshahawi H (2002) In-situ contamination monitoring and GOR measurement of formation samples. In: SPE 77899, SPE Asia Pacific Meeting, Melbourne. https://doi.org/10.2118/77899-MS

Dutta Majumdar R, Gerken M, Mikula R, Hazendonk P (2013) Validation of the Yen-Mullins model of athabasca oil-sands asphaltenes using solution-state ^{1}H NMR relaxation and 2D HSQC spectroscopy. Energy Fuel 27:6528–6537

Dutta Majumdar R, Gerken M, Hazendonk P (2015) Solid-state ^{1}H and ^{13}C nuclear magnetic resonance spectroscopy of athabasca oil sands asphaltenes: evidence for interlocking π-stacked nanoaggregates with intercalated alkyl side chains. Energy Fuel 29:2790–2800

Dutta Majumdar R, Montina T, Mullins OC, Gerken M, Hazendonk P (2017) Insights into asphaltene aggregate structure using ultrafast magic angle spinning solid-state ^{1}H NMR spectroscopy. Fuel 193:359–368

Ellis DV, Singer JM (2008) Well logging for earth scientists. Springer, Dordrecht

Elshahawi H, Mullins OC, Hows M, Colacelli S, Flannery M, Zuo JY, Dong C (2009) Reservoir fluid analysis as a proxy for connectivity in deepwater reservoirs. SPWLA, The Woodlands

Etchart I, Chen H, Dryden P, Jundt J, Harrison C, Hsu K, Marty F, Mercier B (2008) MEMS sensors for density-viscosity sensing in a low-flow and high pressure microfluidic environment. Sensors Actuators 141:266–275

Eyssautier J, Levitz P, Espinat D, Jestin J, Gummel J, Grillo I, Barré L (2011) Insight into asphaltene nanoaggregate structure inferred by small angle neutron and X-ray scattering. J Phys Chem B 115:6827–6837

Eyssautier J, Henaut I, Levitz P, Espinat D, Barré L (2012a) Organization of asphaltenes in a vacuum residue: a small-angle X-ray scattering (SAXS)−viscosity approach at high temperatures. Energy Fuel 26(5):2696–2704

Eyssautier J, Espinat D, Gummel J, Levitz P, Becerra M, Shaw S, Barré L (2012b) Mesoscale organization in a physically separated vacuum residue: comparison to asphaltenes in a simple solvent. Energy Fuel 26(5):2680–2687

Forsythe JC, Pomerantz AE, Seifert DJ, Wang K, Chen Y, Zyo JY, Nelson RK, Reddy CM, Schimmelmann A, Sauer P, Peters KE, Mullins OC (2015) A geological model for the origin of fluid compositional gradients in a large Saudi Arabian oilfield: an investigation by two-dimensional gas chromatography and asphaltene chemistry. Energy Fuel 29(9):5666–5680

Freed DE, Lisitza NV, Sen PN, Song YQ (2009) A study of asphaltene nanoaggregation by NMR. Energy Fuel 23:1189–1193

Frysinger GS, Gaines RB (2001) Separation and identification of petroleum biomarkers by comprehensive two-dimensional gas chromatography. J Sep Sci 24(2):87–96

Fujisawa G, Van Agthoven MA, Rabbito P, Mullins OC (2002) Near-infrared compositional analysis of gas and condensate reservoir fluids at elevated pressures and temperatures. Appl Spectrosc 56:1615

Goual L, Sedghi M, Zeng H, Mostowfi F, McFarlane R, Mullins OC (2011) On the formation and properties of asphaltene nanoaggregates and cluster by DC-conductivity and centrifugation. Fuel 90:2480–2490

Goual L, Sedghi M, Mostowfi F, McFarlane R, Pomerantz AE, Saraji S, Mullins OC (2014) Cluster size and critical clustering concentration by centrifugation and DC-conductivity. Energy Fuel 28(8):5002–5013

Groenzin H, Mullins OC (1999) Asphaltene molecular size and structure. J Phys Chem A 103:11237–11245

Groenzin H, Mullins OC (2000) Molecular sizes of asphaltenes from different origin. Energy Fuel 14:677

Hansen CM (2005) Hansen solubility parameters: a user's handbook, 2nd edn. CRC Press/Taylor & Francis Group, New York

Hantschel T, Kauerauf A (2009) Fundamentals of basin and petroleum systems modeling. Springer, Berlin

Høier L (1997) Miscibility variations in compositionally grading petroleum reservoirs. PhD thesis, Norwegian University of Science and Technology, Trondheim

Høier L, Whitson CH (2001) Compositional grading – theory and practice. In: SPE reservoir evaluation & engineering, 4, 6, 525–535; also presented as paper SPE 63085 at the SPE annual technical conference and exhibition, Dallas

Juyal P, McKenna AM, Yen A, Rodgers RP, Reddy CM, Nelson RL, Andrews AB, Atolia E, Allenson SJ, Mullins OC, Marshall AG (2011) Analysis and identification of biomarkers and origin of blue color in an unusually blue crude oil. Energy Fuel 25:172–182

Kazakevich E, Abram P, Wichers W, Weinheber P, Makhmotov A, Kazuho S, El-Battawy A (2016) The first application of the wireline 3D radial probe for determination of permeability and mobile fluid type in hishly water saturated rocks in the Piltun-Astokhskoye field. In: SPE 182559, Annual Casp. technical conference and exhibition, Kazalhstan

Kharrat AM, Indo K, Mostowfi F (2013) Asphaltene content measurement using an optical spectroscopy technique. Energy Fuel 27(5):2452–2457

Korb JP, Louis-Joseph A, Benamsili L (2013) Probing structure and dynamics of bulk and confined crude oils by multiscale NMR spectroscopy, diffusometry, and relaxometry. J Phys Chem B 117:7002–7014

Miall AD (2000) Principle of sedimentary basin analysis. Springer, Berlin

Mishra V., Canas J, Dumont H, Chen L, De Santo, I, Pfeiffer T, Achourov V, Zuo JY, Mullins OC (2014) DFA connectivity advisor, a new workflow to use measured and modeled fluid gradients for analysis of reservoir connectivity. In: SPE OTC 25173

Mostowfi F, Indo K, Mullins OC, McFarlane R (2009) Asphaltene nanoaggregates and the critical nanoaggregate concentration from centrifugation. Energy Fuel 23:1194–1200

Mullins OC (2008) The physics of reservoir fluids; discovery through downhole fluid analysis. Schlumberger Press, Houston

Mullins OC (2010) The modified yen model. Energy Fuel 24:2179–2207

Mullins OC (2011) The asphaltenes. Annu Rev Anal Chem 4:393–418

Mullins OC, Zhu Y (1992) First observation of the Urbach tail in a multicomponent organic system. Appl Spectros 46:354

Mullins OC, Mitra-Kirtley S, Zhu Y (1992) Electronic absorption edge of petroleum. Appl Spectros 46:1405

Mullins OC, Schroer J, Beck G (2000) Real-time Quantification of OBM Filtrate Contamination during openhole wireline sampling by optical spectroscopy. In: SPWLA 41st annual symposium, Dallas, Paper SS

Mullins OC, Daigle T, Crowell C, Groenzin H, Joshi NB (2001) Gas-oil ratio of live crude oils determined by near-infrared spectroscopy. Applied Spectros 55:197

Mullins OC, Betancourt SS, Cribbs ME, Creek JL, Andrews AB, Dubost F, Venkataramanan L (2007a) The colloidal structure of crude oil and the structure of reservoirs. Energy Fuel 21:2785–2794

Mullins OC, Sheu EY, Hammami A, Marshall AG (eds) (2007b) Asphaltenes, heavy oil and petroleomics. Springer, New York

Mullins OC, Rodgers RP, Weinheber P, Klein GC, Venkatramanan L, Andrews AB, Marshall AG (2007c) Oil reservoir characterization via crude oil analysis by downhole fluid analysis in oil wells with visible-near-infrared spectroscopy and by laboratory analysis with electrospray ionization-Fourier transform ion cyclotron resonance mass spectroscopy. Energy Fuel 21:256

Mullins OC, Martinez-Haya B, Marshall AG (2008a) Contrasting perspective on asphaltene molecular weight; this comment vs. the overview of A.A. Herod, K.D. Bartle, R. Kandiyoti. Energy Fuel 22:1765–1773

Mullins OC, Ventura GT, Nelson RL, Betancourt SS, Raghuraman B, Reddy CM (2008b) Oil reservoir characterization by coupling downhole fluid analysis with laboratory 2D-GC analysis of crude oils. Energy Fuel 22:496–503

Mullins OC, Sabbah H, Eyssautier J, Pomerantz AE, Barré L, Andrews AB, Ruiz-Morales Y, Mostowfi F, McFarlane R, Goual L, Lepkowicz R, Cooper T, Orbulescu J, Leblanc JM, Edwards J, Zare RN (2012) Advances in asphaltene science and the Yen-Mullins model. Energy Fuel 26:3986–4003

Mullins OC, Zuo JY, Seifert D, Zeybek M (2013) Clusters of asphaltene nanoaggregates observed in oilfield reservoirs. Energy Fuel 27:1752–1761

Mullins OC, Pomerantz AE, Zuo JY, Dong C (2014) Downhole fluid analysis and asphaltene science for petroleum reservoir evaluation. Ann Rev Chem Biomolec Eng 5:325–345

Mullins OC, Zuo JY, Pomerantz AE, Forsythe JC, Peters KE (2017) Reservoir fluid geodynamics: the chemistry and physics of oilfield reservoir fluids after trap filling. Energy Fuel 31(12):13088–13119

Nelson RK, Kile BM, Plata DL, Sylva SP, Xu L, Reddy CM, Gaines RB, Frysinger GS, Reichenbach SE (2006) Tracking the weathering of an oil spill with comprehensive two-dimensional gas chromatography. Environ Forensic 7(1):33–44

Nelson RK, Aeppli C, Arey JS, Chen H, de Oliveira AHB, Eiserbeck C, Frysinger GS, Gaines RB, Grice K, Gros J, Hall GJ, Koolen HHF, Lemkau KL, McKenna AM, Reddy CM, Rodgers RP, Swarthout RF, Valentine DL, White HK (2016) Applications of comprehensive two-dimensional gas chromatography (GC $\times$ GC) in studying the source, transport, and fate of petroleum hydrocarbons in the environment. In: Stout SA, Wang Z (eds) Standard handbook oil spill environmental forensics, 2nd edn. Academic, Boston, pp 399–448

O'Keefe M, Godefroy S, Vasques R, Agenes A, Weinheber P, Jackson R, Ardila M, Wichers W, Daungkaew S, De Santo I (2007) In-situ density and viscosity measured by wireline formation testers. In: SPE 110364, Asia Pacific Oil & Gas conference and exhibition. https://doi.org/10.2118/110364-MS

Peng DY, Robinson DB (1976) A new two-constant equation of state. Ind Eng Chem Fundam 15:59–64

Peters KE, Walters CC, Moldowan JM (2005) The biomarker guide, 2nd edn. Cambridge University Press, Cambridge, UK

Pfeiffer T, DiPrimio R, Achourov V, Mullins OC (2016) Tar mat on baffles in the middle of an oil column. In: SPWLA annual symposium, Iceland

Pfeiffer T, Di Primio R, Vladislav Achourov V, Mullins OC (2017) Scanning electron micrographs of tar mat intervals formed by asphaltene phase transition. Petrophysics 58:141–152

Pomerantz AE, Bake KD, Craddock PR, Qureshi A, Zeybek M, Mullins OC, Kodalen BG, Mitra-Kirtley S, Bolin TB, Seifert DJ (2013) Sulfur speciation in asphaltenes from a highly compositionally graded oil column. Energy Fuel 27:4604–4608

Pomerantz AE, Wu Q, Mullins OC, Zare RN (2015) Laser-based mass spectroscopic assessment of asphaltene molecular weight, molecular architecture and nanoaggregate number. Energy Fuel 29:2833–2842

Rane JP, Pauchard V, Couzis A, Banerjee S (2013) Interfacial rheology of asphaltenes at oil-water interfaces and interpretation of the equation of state. Langmuir 29:4750–4759

Rane JP, Zarkar S, Pauchard V, Mullins OC, Christie D, Andrews AB, Pomerantz AE (2015) Banerjee, applicability of the langmuir equation of state for asphaltene adsorption at the oil–water interface: coal-derived, petroleum, and synthetic asphaltenes. Energy Fuel 29(6):3584–3590

Redelius P (2005) Hansen solubility parameters of asphalt, bitumen, and crude oils. In: Hansen CM (ed) Hansen solubility parameters: a user's handbook, 2nd edn. CRC Press/Taylor & Francis Group, New York, Chapter 9, pp 151–175

Sabbah H, Morrow AL, Pomerantz AE, Zare RN (2011) Evidence for island structures as the dominant architecture of asphaltenes. Energy Fuel 25:1597–1604

Schuler B, Meyer G, Pena D, Mullins OC, Gross L (2015) Unraveling the molecular structures of asphaltenes by atomic force microscopy. J Amer Chem Soc 137(31):9870–9876

Schuler B, Fatayer S, Meyer G, Rogel E, Moir M, Zhang Y, Harper MR, Pomerantz AE, Bake K, Witt M, Pena D, Kushnerick JD, Mullins OC, Ovalles C, van den Berg FGA, Gross L (2017a) Heavy oil mixtures of different origins and treatments studied by AFM. Energy Fuel 31:6856–6861

Schuler B, Zhang Y, Collazos S, Fatayer S, Meyer G, Perez D, Guitián E, Harper MR, Kushnerick JD, Peña D, Gross L (2017b) Chem Sci 8:2315–2320

Sheu EY, Long Y, Hamza H (2007) Asphaltene self-association and precipitation in solvents, AC-conductivity measurements. In: Mullins OC, Sheu EY, Hammami A, Marshall AG (eds) Asphaltene, heavy oils and petroleomics. Springer, New York, Chapter 10, pp 259–277

Sullivan M, Harrison C, Goodwin A, Hsu K, Godefroy S (2009) On the operation of a vibrating viscometer at large amplitude. Fluid Phase Equilib 276:99–107

Tissot BP, Welte DH (1984) Petroleum formation and occurrence, 2nd edn. Springer, Berlin

Uchytil S, Mishra VK, Betancourt SS, Guthrie J, Huang J, Teerman S, Nguyen A, Stan Evans S, Nagarajan N, Mullins OC (2016) Impact of a secondary condensate charge into an oil reservoir evaluated by downhole fluid analysis, core analysis, and production. In: OTC 27240, Houston

Van Agthoven MA, Fujisawa G, Rabbito P, Mullins OC (2002) Near-infrared spectral analysis of gas mixtures. Appl Spectrosc 56:593

Wu Q, Pomerantz AE, Mullins OC, Zare RN (2014a) Laser-based mass spectrometric determination of aggregation numbers for petroleum- and coal-derived asphaltenes. Energy Fuel 28:475–482

Wu Q, Seifert DJ, Pomerantz AE, Mullins OC, Zare RN (2014b) Constant asphaltene molecular and nanoaggregate mass in a gravitationally segregated reservoir. Energy Fuel 28:3010–3015

Yudin IK, Anisimov MA (2007) Dynamic light scattering monitoring of asphaltene aggregation in crude oils and hydrocarbon solutions. In: Mullins OC, Sheu EY, Hammami A,

Marshall AG (eds) Asphaltenes, heavy oils and petroleomics. Springer, New York, Chapter 17, pp 439–468

Zeng H, Song YQ, Johnson DL, Mullins OC (2009) Critical nanoaggregate concentration of asphaltenes by low frequency conductivity. Energy Fuel 23:1201–1208

Zimmerman TH, Pop JJ, Perkins JL (1989) Down hole tool for determination of formation properties. US Patent No. 4,860,581

Zimmerman TH, Pop JJ, Perkins JL (1990) Down hole method for determination of formation properties. US Patent No. 4,936,139

Zuo JY, Mullins OC, Freed DE, Dong C, Elshahawi H, Seifert DJ (2013) Advances in the Flory-Huggins-Zuo equation of state for asphaltene gradients and formation evaluation. Energy Fuel 27:1722–1735

Zuo JY, Pan S, Wang K, Mullins OC, Harfoushian JH, Elshahawi E (2016) Delineation of gravitational instability induced by gas charges into oil reservoirs using diffusion and Flory-Huggins-Zuo equations. In: SPE-APOGC 182380, Perth. https://doi.org/10.2118/182380-MS

Zuo JY, Mullins OC, Achourov V, Pfeiffer T, Pan S, Wang K, Kollien T, Di Primio R (2017) Fluid distributions during light hydrocarbon charges into oil reservoirs using multicomponent Maxwell-Stefan diffusivity in gravitational field. Fuel 209:211–223

Multiscale Modeling of Cohesive-Frictional Strength Properties in Cementitious Materials

68

Steven D. Palkovic and Oral Büyüköztürk

Contents

Abstract

Computational modeling has been a valuable tool for studying the role of molecular structure and chemical composition in cementitious materials. A significant challenge is linking these molecular details to macroscale properties that define the durability of the material. This objective requires a holistic multiscale approach that translates the essential physics across atomistic (~ 1 nm), mesoscale (~ 100 nm), and microstructural (~ 100 μm) length scales. Recent studies have identified the importance of cohesive-frictional interactions on the material strength development. Cohesive-frictional materials exhibit a higher strength under compression than tension loading and a shear strength that increases when an external confinement is applied. These concepts are well

S. D. Palkovic · O. Büyüköztürk (✉)
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: spalkovi@mit.edu; obuyuk@mit.edu

established for cementitious materials at the macroscale where they are routinely implemented in plasticity and strength models for structural design. However, macroscale phenomenological models are unable to identify the underlying structures and interactions that are the driving force behind this fundamental system response. In this chapter, observations from molecular dynamics (MD), discrete element method (DEM), and finite element method (FEM) modeling at multiple length scales are described to further understand the role of cohesive-frictional interactions in the strength development of cementitious materials.

1   Introduction

Cementitious materials like cement paste, mortar, and concrete are widely used because of their low cost and adaptability to a wide range of mix ingredients. This sensitivity to raw material inputs provides an opportunity for engineers to produce sustainable and durable materials for resilient infrastructure. Sustainable building materials require the incorporation of locally available additives such as fly ash, blast furnace slag, and volcanic ash in lieu of ordinary Portland cement (OPC) (Lothenbach et al. 2011). OPC is the component of cementitious building materials with the highest embodied energy demand, and its production currently accounts for up to 6% of global carbon emissions (Marland et al. 2017). The influence of additives on cementitious materials is difficult to predict and currently requires a trial-and-error approach to evaluate different combinations of constituents. Furthering our understanding of the role of additives at multiple length scales will reduce uncertainties and enable the design of sustainable materials that meet the performance requirements of a building or structure.

Computational modeling provides a powerful tool for establishing structure-property relationships within hierarchical materials. Nanoscale models provide significant insights for the principal hydrated phase of OPC-based materials, calcium-silicate-hydrate (C-S-H), that form more than 60% of a cement paste's volume (Double et al. 1978) and have a strong influence on the overall elasticity, strength, and long-term performance (Bernard et al. 2003; Pichler et al. 2009; Vandamme and Ulm 2009). C-S-H is a nanostructured material that has characteristic features spanning orders of magnitude in length scales which evolve dynamically with time. This material complexity provides a significant challenge for researchers attempting to use a bottom-up multiscale approach to establish connections between nanoscale behavior and engineering properties at the macroscale.

In this chapter, recent studies on the mechanical strength development of cementitious materials spanning from 10^{-10} to 10^{-2} m are discussed, with a particular focus on changes in cohesive-frictional behavior. A cohesive-frictional material exhibits a strength asymmetry in compression and tension and has a shear strength that is dependent on the confinement applied through an external normal stress or hydrostatic pressure. This behavior has been known for decades at the macroscale, where Mohr-Coulomb or Drucker-Prager models are used to predict the material strength envelope and plastic deformation behavior (Chen 2007). However, an

understanding of the features and length scales that control the origin and evolution of cohesive-frictional properties requires a holistic multiscale approach (Palkovic et al. 2016). This is accomplished through computational modeling that employs molecular dynamics (MD), discrete element methods (DEM), or continuum finite element methods (FEM) depending on the model size and features included. Information from lower-scale models are used to determine the appropriate model assumptions and inputs at larger scales. By understanding the critical aspects of cohesive-frictional strength development in cementitious building materials, more sustainable and durable structures can be obtained.

2   Hierarchies in Cement Paste

Cement paste is the matrix phase in larger-scale composites like mortar and concrete. Therefore, it is an appropriate starting point for developing a multiscale framework for cementitious building materials. Cement paste is a heterogeneous mixture of hydrating phases, mineral inclusions, and pores. The essential model features and length scales are summarized in Fig. 1. Models smaller than 1 μm focus on the nanostructured C-S-H phase. C-S-H is composed of colloidal nanoparticles with an internal layered structure of calcium-silicate layers separated by an interface containing water molecules, hydroxyl groups, and counter ions (Richardson 2008). This layered configuration provides a building block for studying the influence of atomistic structure and chemical composition on nanoscale interactions (Hou et al. 2015; Qomi et al. 2014). Forces within the layers and from the surrounding electrostatically charged pore solution influence the assembly, pore structure, and mechanical properties of colloidal nanoparticles (Ioannidou et al. 2014, 2016;
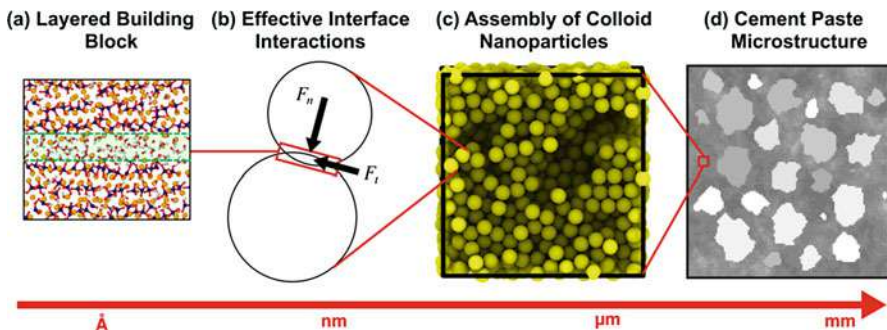
Fig. 1 Multiscale description of cement paste. (**a**) A layered building block describes the molecular structure which changes with varying chemical composition. (**b**) The effective behavior of this interface can be used to describe interparticle forces between C-S-H colloids. These forces include a normal component (F_n) for cohesive interactions and tangential component (F_t) for the shear resistance between particle pairs. (**c**) These particles assemble to form mesoscale solid-pore structures of colloids interacting through cohesive-frictional forces. (**d**) At larger scales, C-S-H behaves as a continuous material within a cement paste microstructure (Palkovic 2018)

Masoero et al. 2012). These colloids have a size ranging from 3 to 15 nm (Chiang et al. 2012; Jennings 2008; Skinner et al. 2010) and assemble to form a fibrillar pore network (Ioannidou et al. 2014). The arrangement and interactions of colloidal assemblies control the mesoscale behavior of C-S-H that bridges between discrete models and continuum descriptions. Continuum behavior is input to cement paste microstructure models as constitutive relationships that homogenize the discrete features of the material. Microstructure models consider the local distribution in strength and stiffness of hydrated phases that has been measured experimentally through statistical indentation (Constantinides and Ulm 2007). In the following sections, computational modeling is applied to study the development of cohesive-frictional properties within these structures.

3 Molecular C-S-H Interfaces

At the smallest length scales, all matter is composed of atoms which interact through intermolecular forces that are dependent on chemical composition and molecular structure. Nanoscale models must account for the discrete nature of materials at submicron length scales where classical continuum theories can no longer be applied.

For cement paste, atomistic modeling has been used to investigate the internal layered structure of C-S-H colloids. These models are based on the mineral phase tobermorite, which is systematically modified by introducing defects to the crystalline structure to match experimental chemical composition, silica chain connectivity, degree of hydration, and density measurements (Pellenq et al. 2009). These models frequently incorporate the reactive force field ReaxFF (Van Duin et al. 2001), which was parameterized through density functional theory (DFT) simulations (Manzano et al. 2012b). Reactive force fields are capable of capturing the formation and breaking of bonds, water dissociation, and atomic charge fluctuations during a simulation (Manzano et al. 2012a, 2013).

A C-S-H structure with a calcium/silicon (Ca/Si) ratio of 1.70 is studied because this composition is typical of what is produced during the hydration of OPC mixtures (Hou et al. 2015). The structure consists of relatively crystalline layers separated by a disordered interface (Bauchy et al. 2014). Under shear loading, the interface allows for strain localization while the more ordered layers experience little deformation. This is shown in Fig. 2 by calculating the local atomic shear strain as the overall structure is subjected to a linearly varying affine shear strain. The heterogeneous response to homogeneous deformation indicates that the interface is the limiting region for shear loading. This is further confirmed by the agreement in strength properties obtained for direct and affine shear loading, where direct shear loading is used to constrain localization within the interface (Palkovic et al. 2015). Other researchers have attributed this to a "lubricating" effect of water molecules which screen the interactions of adjacent calcium-silicate layers (Pellenq et al. 2009).
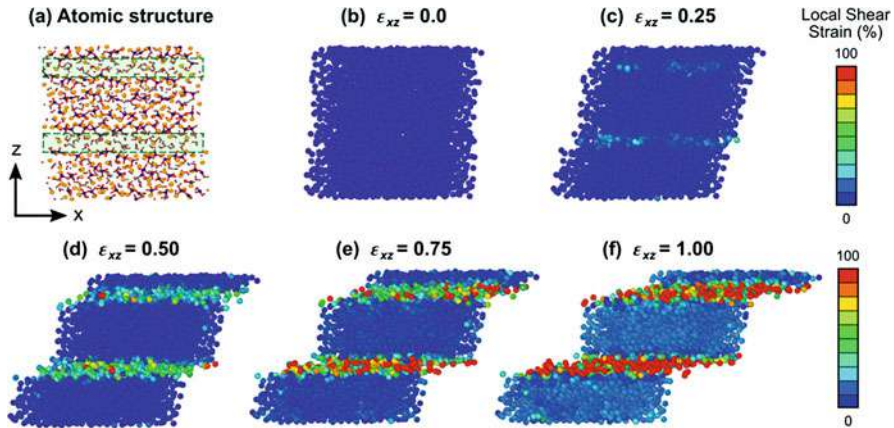
Fig. 2 (**a**) Two-dimensional view of a representative C-S-H atomic structure with Ca/Si of 1.65, with the interlayer region shown by the highlighted region. Atoms are colored by element type, with orange for Ca, purple for Si, red for O, and gray for H. (**b–f**) System response to applied affine shear (ε_{xz}), with atoms colored by the magnitude of local atomic shear strain. Atomic shear strain localizes within the water-filled interface region as the calcium-silicate layers remain relatively undeformed (Büyüköztürk and Palkovic 2017)

The response of the molecular C-S-H structure to external confinement is measured by applying a compressive strain normal to the plane of the interface (Palkovic et al. 2017b). Figure 3 shows that the interfacial shear strength increases with a greater magnitude of applied compressive strain. This response is characteristic of cohesive-frictional materials and can be quantified using a Mohr-Coulomb strength model given by

$$\tau_n = \mu\sigma_n + c \tag{1}$$

where τ_n is the shear strength and σ_n is the normal stress along the maximum shearing plane. The Mohr-Coulomb model is defined by two constants that vary for different material systems, the internal friction coefficient μ and cohesion c. The shear strength can be determined by averaging the fluctuations of the system response during the relatively perfectly plastic flow regime that occurs for all models after approximately 0.25 strain.

The strength envelope and resulting Mohr-Coulomb model for the average of five molecular C-S-H structures with a Ca/Si ratio of 1.7 is shown in Fig. 4. The Mohr-Coulomb fit provides a cohesion of $c = 2.6 \pm 0.2$ GPa and $\mu = 0.14 \pm 0.01$. The internal friction coefficient is consistent with other amorphous materials such as 0.12 for a metallic glass (Lund and Schuh 2003) and 0.07–0.16 for a polymer glass (Rottler and Robbins 2001). This suggests that the magnitude of μ is related to the disordered structure within the interface region and is less dependent on the atomic bonding which greatly differ for a metal, polymer, and C-S-H system. Amorphous materials exhibiting strength asymmetry and a dilatant response under
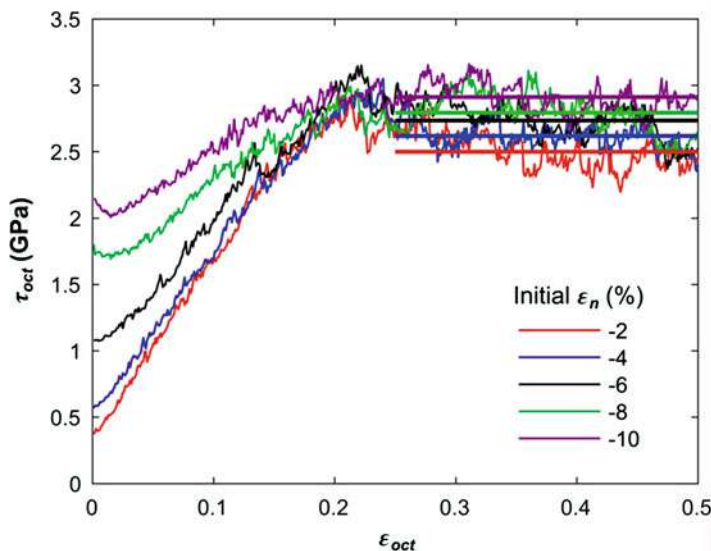
Fig. 3 Octahedral stress-strain for molecular C-S-H structures with an initial strain applied normal to the plane of the interface (ε_n). The horizontal lines show the average octahedral flow stress for strain higher than 0.25 that is used to determine the cohesive-frictional strength envelope

shear loading are believed to undergo plastic deformation through non-affine shear transformations (Argon 1979). A shear transformation is the rearrangement of groups of atoms that occurs under loading to relieve the buildup of internal stress. Atoms surrounding the shear transformation apply a restraining force that increases the shear load necessary to initiate the shear transformation and results in a shear strength dependency of the material confinement. The repeated stress drops observed under shear loading (Manzano et al. 2012a; Pellenq et al. 2009), and the measurable internal friction coefficient in Fig. 4, are consistent with shear transformations occurring within a disordered water-filled interface of the molecular structure as the origin of plastic deformation and cohesive-frictional behavior in C-S-H. The influence of cohesive-frictional properties at larger scales is investigated in Sects. 4 and 5.

4 Colloidal C-S-H

The computational complexity of a reactive force field limits the system size to approximately 10^4 atoms. To reach larger length scales, coarse-graining must be used to determine effective interactions that reduce the model degrees of freedom and allow for modeling of larger-scale structures. For colloidal C-S-H, the interface region that limits the strength and deformation behavior of the molecular structure is a natural choice to parameterize larger-scale interactions. One study developed a
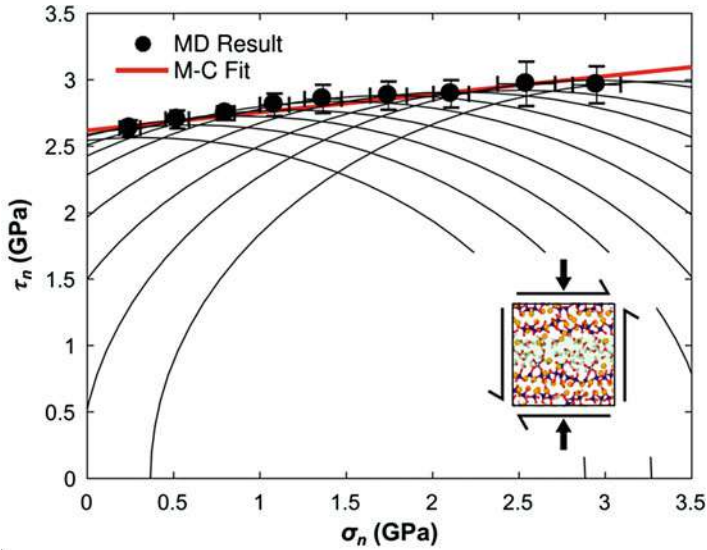
Fig. 4 Shear strength as a function of normal stress along the maximum shear plane for molecular C-S-H. A Mohr-Coulomb model is fit to the strength envelope that is tangent to the Mohr's circle for different magnitudes of confinement using least-squares linear regression

Lennard-Jones pair potential that was modified to account for particle of dissimilar size while also incorporating the shear rupture strain and elastic stiffness of the interface region (Masoero et al. 2012). However, a molecular dynamics pair potential like the Lennard-Jones function cannot sustain shear forces for interacting particle pairs, in disagreement with atomistic simulations of the layered C-S-H structure.

4.1   Cohesive-Frictional Force Field (CFFF)

To provide more realistic interface interactions, Palkovic et al. (2017a) proposed a cohesive-frictional force field (CFFF) that incorporates aspects of adhesive contact mechanics (Derjaguin et al. 1975) which are implemented through numerical techniques developed for discrete element modeling (DEM) (Cundall and Strack 1979). In this force field, frictional forces between contacting particles are calculated to represent the cohesive-frictional behavior of the interface. This tangential force F_t is based on the shear stiffness of the interface up to a maximum value given by a Mohr-Coulomb limit as

$$F_t(r) = \min\left[8\Delta_t G_e a, \;\; \mu F_n + \pi c a^2\right] \tag{2}$$

where r is the center-to-center spacing of interacting colloids, $G_e = G/(2(2 - v))$ is the effective shear stiffness from the interface shear modulus G and Poisson's ratio v, Δ_t is an increment of tangential displacement, F_n is the normal force between colloids, μ is the internal friction coefficient, and c is the interface cohesion. The tangential force varies with the contact area between colloids $a = a_0 + \sqrt{D_e (D_0 - r)/4}$, where $D_e = 2D_i D_j/(D_i + D_j)$ is the harmonic mean of particles D_i and D_j and $D_0 = (D_i + D_j)/2$ is the average particle size defining the equilibrium spacing. The adhesive component of the contact radius $a_0 = \sqrt[3]{3\pi D_e^2/(32 E_e)}$ is given by the Derjaguin-Muller-Toporov (DMT) theory and is a function of w the work of adhesion and $E_e = E/(2(1 - v^2))$ the effective normal stiffness given by the Young's modulus E and Poisson's ratio v. DMT is a continuum theory that considers interactions between smooth spherical particles. As a result, the contact area is an upper bound for atomically rough C-S-H colloids.

The normal force-displacement function was derived by fitting a three-parameter exponential function to the tension and compression behavior of the atomistic C-S-H structure (Palkovic et al. 2017a). This pair potential was then extended to spherical colloids of varying diameters by considering interacting harmonic springs in series. The final pair potential normal force relationship is given by

$$F_n(r) = \frac{\pi E D_e (r - D_0) \left(1 + e^{-\alpha}\right)}{4 \left(1 + \exp\left(\frac{\alpha(r-D_0)}{D_0 \varepsilon_u} - 1\right)\right)} \tag{3}$$

where E is the Young's modulus that controls the compressive and initial stiffness of the pair potential, ε_u is a strain parameter that offsets the position of the maximum cohesive force, and α is a dimensionless parameter that controls how brittle or ductile the function behaves. Figure 5 compares Eq. 3 with the Lennard-Jones

Fig. 5 Comparison of interparticle force-separation curves for average particle sizes of 3, 6, and 9 nm for a polydisperse Lennard Jones (LJ) (Masoero et al. 2014), Eq. 3 with a constant strain parameter of 4.3%, and a size-dependent strain parameter calculated with Eq. 4 (Palkovic et al. 2017a)
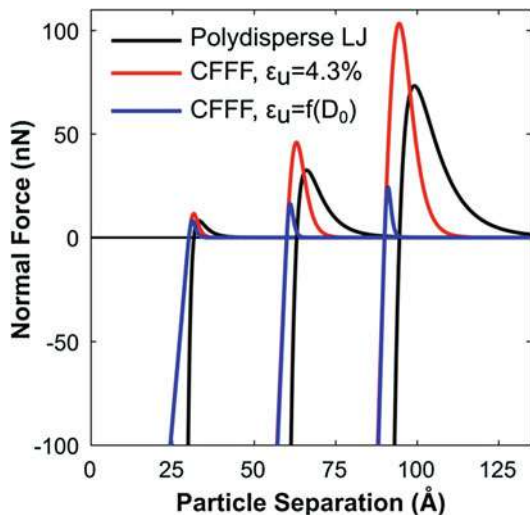
function that was modified to account for polydisperse particles of varying diameter. Two forms of Eq. 3 are shown, (i) with a constant strain parameter $\varepsilon_u = 4.3\%$ and (ii) with maximum adhesive force predicted by DMT theory $F_n^{\max} = \pi E w D_e / 2$. A maximum adhesive force can be enforced with a size-dependent strain parameter, which for $\alpha = 2$ is given by

$$\varepsilon_u = \frac{2w\alpha}{E D_0 \left(1 + e^{-\alpha}\right)}. \tag{4}$$

The reduced adhesive force obtained when applying the size-dependent strain parameter defined in Eq. 4 shows better agreement with atomic force microscopy (AFM) measurements (Lesko et al. 2001; Plassard et al. 2004), and the lower tensile strain is consistent with dynamic rheology experiments (Nachbaur et al. 2001). The pair potential shown in Fig. 5 uses a Young's modulus $E = 61.3$ GPa, Poisson's ratio $v = 0.28$, $\alpha = 2.0$, and adhesion energy $w = 1.72$ N/m^2 (Bauchy et al. 2015) that were measured directly from the molecular C-S-H structure (Palkovic et al. 2015, 2017a). These inputs are used to investigate the mechanical response of colloid assemblies in Sect. 4.2.

4.2   Mechanics of Colloidal C-S-H

The CFFF allows for the investigation of the role of molecular behavior within colloidal assemblies of C-S-H particles through the fitting of force field inputs. On the order of hundreds of nm, small-angle scattering experiments indicate that models should consider polydispersity through varying colloid sizes and long-range particle correlations (Allen et al. 2007). Recent studies have assumed that particle sizes are given by a uniform random distribution with minimum diameter D_m and maximum diameter D_M, which can be defined by a normalized polydispersity parameter $\delta = (D_M - D_m) / \left(\sqrt{3}\,(D_M + D_m)\right)$. Polydispersity influences the system packing density (Masoero et al. 2012), which nanoindentation has shown exists across a broad range of densities due to varying gel and capillary porosity (Constantinides and Ulm 2007). The model packing density for N spherical particles is given by $\eta = \left(\sum \pi D_i^3/6\right) / V$, where D_i is the diameter of particle i and V is the total volume of the simulation cell. Dense C-S-H configurations are generated by randomly inserting colloidal particles with a Monte-Carlo algorithm to progressively fill an initially empty simulation cell. These colloidal configurations represent dense regions within the open C-S-H network that have been studied with a Yukawa potential that captures long-range repulsion in the charged solution (Ioannidou et al. 2014). Table 1 provides an overview of the model parameters considered for subsequent analysis. When cohesive-frictional interactions are considered, a shear modulus $G = 25.6$ GPa and Mohr-Coulomb strength envelope given by $\mu = 0.15$ and $c = 2.6$ GPa of the molecular interface are utilized.

δ	D_m	D_M	L	η	N
	(nm)	(nm)	(nm)		(particles)
0	5.0	5.0	43–65	0.50–0.65	583–2769
0.29	3.0	9.0	51–77	0.55–0.70	831–6296
0.38	3.0	15.0	51–78	0.60–0.74	576–4364

Table 1 Mesoscale C-S-H systems considered. The number of particles N is given for the smallest and largest packing density η
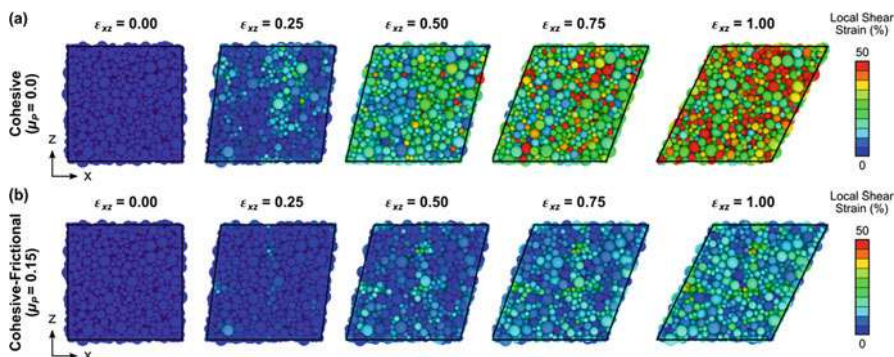
Fig. 6 Colloidal C-S-H response to shear deformation for systems with the same packing density and polydispersity but varying coefficient of friction (μ_P). Particles are colored by the local atomic shear strain (Palkovic 2018)

The influence of cohesive-frictional interactions is significant under large deformations. The local particle shear strain at varying magnitudes of overall shear deformation is shown in Fig. 6 for a polydisperse model with $\delta = 0.29$ and $\eta = 0.70$. Models with purely cohesive interactions ($F_t = 0$) exhibit localized deformations throughout the C-S-H structure. When cohesive-frictional interactions are included, a more distributed strain pattern and lower magnitude of local strain are observed.

The local strain patterns are consistent with the calculated stress-strain response of C-S-H structures under external confinement shown in Fig. 7. Colloidal models are first compressed to an initial pressure P_0 using an NPT ensemble at 300 K and pressures ranging from 0.25 to 1.50 GPa. Pure shear loading is then applied in an NVT ensemble with strain increments of 0.1% up to a maximum value of 50%. After each strain increment, a MD relaxation of 50,000 timesteps of 2.5 fs is applied with averaging of the Cauchy stress tensor performed over the final 10% of the relaxation time. Cohesive-frictional interactions result in a significant increase in octahedral stress, an extended elastic regime, and a more dilatant response compared to purely cohesive models. The stress-strain response also shows a smoother behavior, with less frequent stress drops that are likely associated with shear transformations of colloidal particles. These simulations suggest that shear resistance between interacting colloids results in a more continuous response as frictional forces resist the applied deformation and increase the resistance to local particle rearrangements that facilitate plastic deformation.

Fig. 7 Octahedral stress (τ_{oct}) and hydrostatic pressure (P) as a function of applied octahedral shear strain (ε_{oct}) and initial hydrostatic pressure P_0, for a colloidal C-S-H model with polydispersity $\delta = 0.29$ and packing density $\eta = 0.70$ but with varying internal coefficient of friction (μ_P). Models use a size-dependent strain parameter calculated with Eq. 4, but identical loading conditions that use CFFF interactions with a constant strain parameter are provided in Palkovic et al. (2017a)
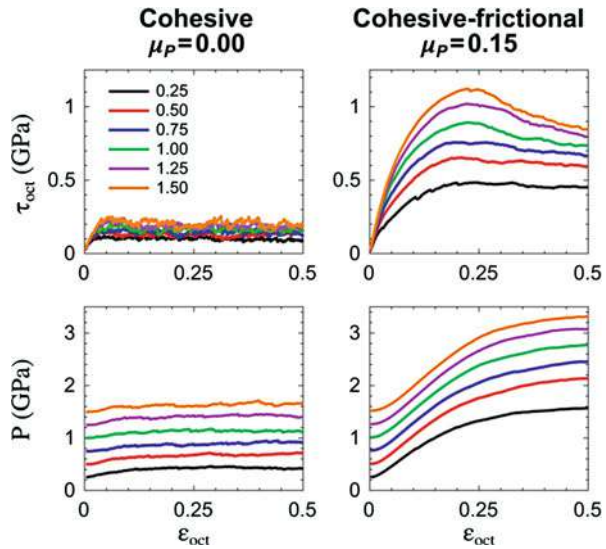
The bulk Mohr-Coulomb strength envelope for the colloidal C-S-H structures can be determined from C-S-H models subjected to varying hydrostatic pressure. These bulk properties are appropriate for material volumes on the order of hundreds of nm and can be compared with the molecular interfaces measured previously. The bulk cohesion c_B and internal friction coefficient μ_B are calculated as a function of polydispersity, packing density, and different particle normal interactions in Fig. 8. Polydispersity has a limited influence on the Mohr-Coulomb strength properties, and results are averaged over all polydispersity values $\langle \delta \rangle$ for a given packing density to make trends more apparent. Stronger adhesive forces between particles as a result of a constant strain parameter ($\varepsilon_u = 4.3\%$) lead to a significant increase in c_B. These models also reach greater packing densities because the significant tensile residual stresses that accumulate during model generation lead to densification during subsequent NPT relaxation that is performed prior to mechanical loading. Comparing models with the same size-dependent strain parameter given by Eq. 4, cohesive-frictional interactions lead to significant increases in both c_B and μ_B. Models that neglect shear forces between particles exhibit a similar bulk μ_B, regardless of the strength of adhesive forces, and the value ranges from 0.10 to 0.15 which is a similar magnitude to the layered molecular structure. This is attributed to the disordered structure that is exhibited by both the interfacial region of molecular C-S-H and overall colloidal assembly. However, when frictional forces are included, the shear resistance between particles greatly amplifies the overall shear strength and dilatant response of C-S-H. Prior simulations of cohesive-frictional models with a constant strain parameter showed even higher values of μ_B approaching 0.6 (Palkovic et al. 2017a). Experimental measurements of c_B and μ_B are not available in the literature, but we can investigate their influence within microstructural models described in Sect. 5 to compare with macroscale experiments.
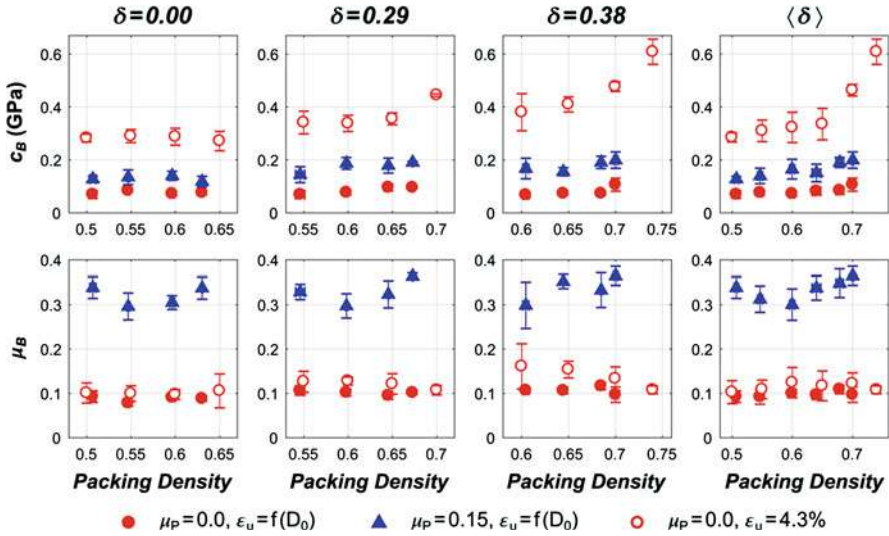
Fig. 8 Mohr-Coulomb bulk strength properties of colloidal C-S-H models with varying polydispersity (δ), packing density, radial strain parameter (ε_u), and internal coefficient of friction (μ_P). Data points and error bars represent the average and standard deviation of four or more samples for each combination of δ and packing density. Results are also averaged for the same packing density but varying polydispersity ($\langle\delta\rangle$). Results with stronger adhesive forces obtained from (Palkovic et al. 2017a)

5   Cement Paste Microstructure

At larger scales ($>1\,\mu$m), the discreteness of underlying features such as the layered molecular structure or assemblies of colloidal nanoparticles is homogenized. As a result, the material behaves according to continuum mechanics which can be modeled using finite element methods (FEM). In the FEM, the mechanical response of a material domain under prescribed boundaries and loads is approximated across local finite elements using polynomial interpolation functions (Bathe 2006). The discretization of the material domain with regions of varying material properties allows for the modeling of a discrete microstructure. The challenges for developing finite element models of cement paste are the random and heterogeneous nature of the material which complicates the determination of a realistic microstructure geometry and interactions between phases. A new approach is to combine statistical nanoindentation concepts (Ulm et al. 2007) with random field modeling techniques (Torquato 2013) to produce random field finite element method (RF-FEM) models of the cement paste microstructure (Palkovic et al. 2018). This methodology allows for the generation of three-dimensional realizations that match experimental measurements on hardened cement paste samples with varying mix composition and hydration degree. Cohesive-frictional interactions can be input

within the constitutive models describing plasticity of hydrated phases like C-S-H. An overview of the method and application on an OPC material are described in Sects. 5.1 and 5.2, respectively.

5.1   Random Field Finite Element Method (RF-FEM) Models

Statistical indentation allows for the measurement of the distribution of mechanical properties in a heterogeneous composite. The overall procedure is summarized in Fig. 9. An indentation test uses an axisymmetric stylus to deform an elastic-plastic substrate. The normal force P acting on the stylus is monitored as a function of the indentation depth h below the undeformed surface (Oliver and Pharr 1992). During loading, the indented material responds through predominantly plastic deformation as a permanent impression is formed on the sample surface which is characterized by the indentation hardness given by
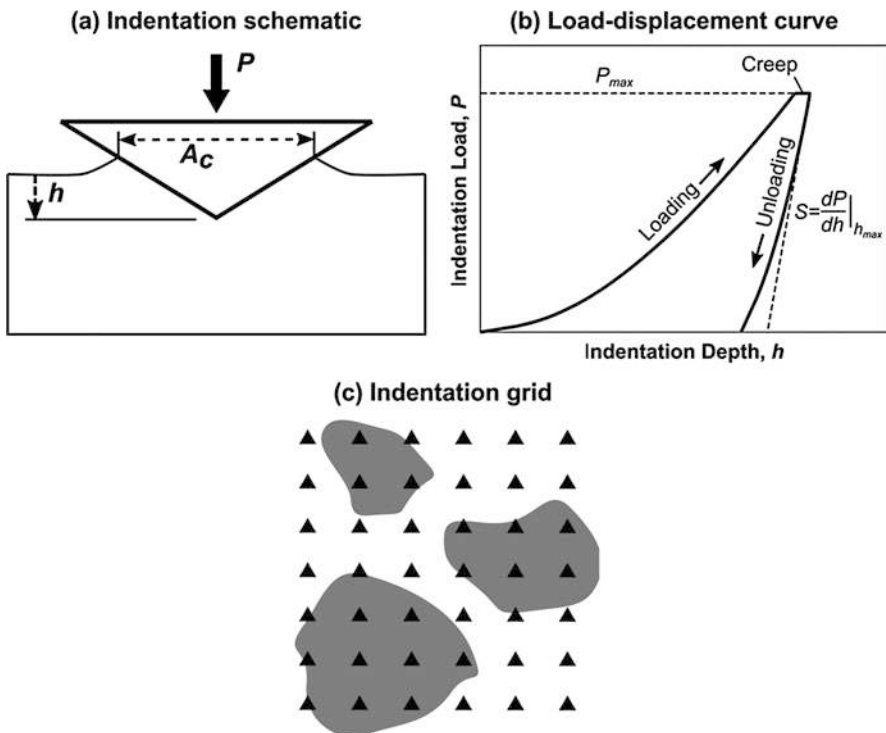
Fig. 9 (**a**) Indentation of an elastic-plastic material with an axisymmetric stylus. (**b**) Load-displacement curve for a cementitious material obtained from each indentation. (**c**) Statistics are obtained by randomly sampling a composite with an indentation grid (Palkovic et al. 2018)

$$H = \frac{P_{\max}}{A_c} \tag{5}$$

where $P_{\max}$ is the maximum applied force and A_c is the projected contact area. The material is assumed to respond elastically when the load is removed, so that the elastic stiffness can be measured with the indentation modulus:

$$M = \frac{\sqrt{\pi} S}{2\sqrt{A_c}} \tag{6}$$

where $S = \frac{dP}{dh}\big|_{h_{\max}}$ is the unloading stiffness measured at the maximum indentation depth $h_{\max}$.

Statistics on mechanical properties are obtained by randomly sampling a microstructure using a grid of indentation tests. Local mechanical properties can be obtained if the volume of material probed during the indentation satisfies scale separability laws given by $0.01 < h_{\max}/D < 2 \cot \theta$, where D is a characteristic size of the microstructure and θ is the half-angle of the stylus ($\theta = 70.3°$ for a Berkovitch tip) (Constantinides et al. 2006). For cement paste, researchers have proposed a maximum indentation depth of 200 to 500 nm based on approximately 5 nm colloids as the smallest feature size and homogeneous regions of hydration product extending 5 or more μm as the largest microstructural feature (Ulm et al. 2010).

Statistical indentation measurements are presented in Fig. 10a for a hardened OPC paste with a water-to-cement (w/c) ratio of 0.35 that was aged for 28 days before being submerged in acetone to retard hydration. The average indentation depth was approximately 315 nm, within the proposed scale separability criteria. The results are analyzed to separate the porous hydration products that form a continuous matrix and inclusion phases that are suspended within the matrix. Hydration products include C-S-H, calcium hydroxide, and ettringite that exhibit a lower strength and stiffness than more crystalline inclusion phases like Portland cement grains. A maximum mechanical threshold of $H \le 3$ GPa and $M \le 75$ GPa is used to separate hydration products from inclusions. This threshold is based on statistical indentation testing on cementitious materials of varying (w/c) ratio that were analyzed with Gaussian mixture models to determine mechanical properties of local phases (Vandamme et al. 2010) and atomistic models of the molecular C-S-H structure excluding gel and capillary pores (Qomi et al. 2014). The resulting indentation modulus and indentation hardness distributions are shown in Fig. 10b–c. The inset of these figures shows the separated hydration products, along with a lognormal probability density function (PDF):

$$p^{HP}(\boldsymbol{x}) = \frac{1}{s_N \sqrt{2\pi \boldsymbol{x}}} \exp \frac{(\ln \boldsymbol{x} - m_N)^2}{2s_N^2} \tag{7}$$
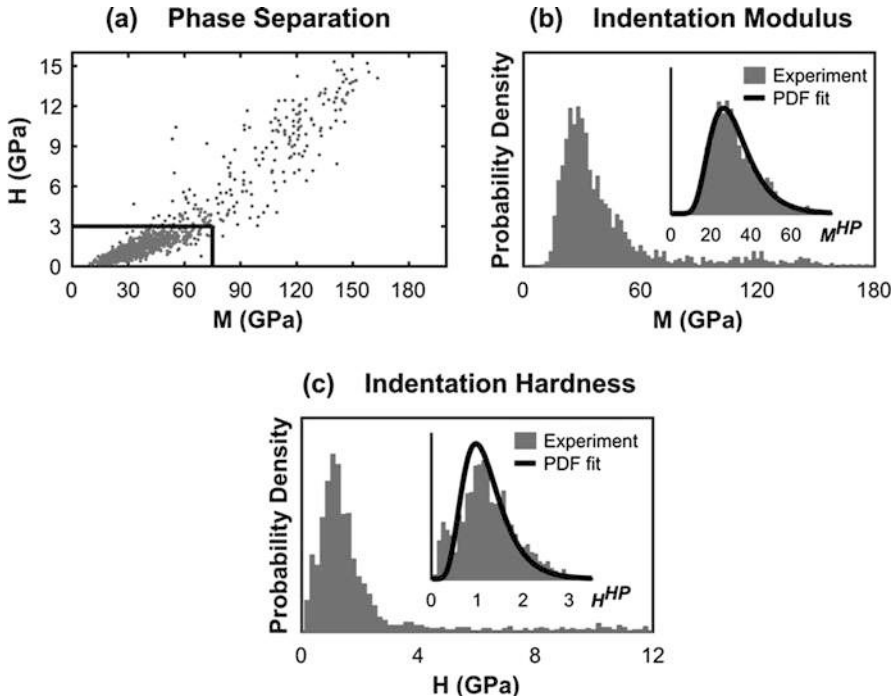
Fig. 10 Analysis of nanoindentation results for random field modeling. (**a**) Porous hydration products are separated based on maximum indentation properties of M ≤ 75 GPa and H ≤ 3 GPa. (**b**) Indentation modulus distribution and (**c**) indentation hardness distribution from 1600 tests on an OPC sample. The insets show the separated hydration products and a lognormal probability density function (PDF) calculated for this subset of data (Palkovic et al. 2018)

calculated for the indentation properties using the mean m_N and variance s_N^2 of the associated normal distribution given by $\ln p^{HP}(x)$. The mean and variance of the lognormal distribution are obtained using $m_{LN} = \exp\left(m_N + s_N^2/2\right)$ and $s_{LN}^2 = \exp\left(2m_N + s_N^2\right)\left(\exp\left(s_N^2\right) - 1\right)$. For the OPC material tested, the moments of the indentation modulus are $m_{LN}^M = 31.8$ GPa and $s_{LN}^M = 11.7$ GPa and for the indentation hardness are $m_{LN}^H = 1.22$ GPa and $s_{LN}^H = 0.52$ GPa.

The measured mechanical property distributions are input within existing random field techniques to generate spatially correlated random fields that have an overall distribution of mechanical properties matching the values measured with nanoindentation. Overbars are used to differentiate the local material properties input within RF-FEM models and the effective microstructure properties obtained from simulations (e.g., $\widetilde{E}$ for local Young's modulus and E for effective microstructure Young's modulus). To be suitable with FEM, indentation properties are converted to an isotropic material model. The indentation modulus can be converted to the Young's modulus E using (Johnson 1987)

$$\frac{1}{M} = \left[\frac{1-v^2}{E} + \frac{1-v_s^2}{E_s}\right]^{-1} \tag{8}$$

where $v_s = 0.07$ and $E_s = 1140$ GPa are the Poisson's ratio and Young's modulus of a diamond-tipped Berkovitch stylus. The indented material Poisson's ratio is taken as 0.28 for hydration products and 0.30 for inclusion phases (Bernard et al. 2003). A Mohr-Coulomb plasticity model given by Eq. 1 can be calculated from the measured indentation hardness using

$$H = \frac{c}{\mu} \sum_{k=1}^{6} (a_k \mu)^k \tag{9}$$

where a_k are fitting coefficients that depend on the geometry of the stylus (Ganneau et al. 2006). Colloidal modeling shows that Mohr-Coulomb properties are dependent on the presence of cohesive-frictional interactions between particles. To study this effect at the microstructural scale, local internal friction coefficients $\widetilde{\mu}_B$ ranging from 0.09 to 0.47 are investigated. With this approach, a material with the same indentation hardness distribution can have varying local cohesive strength $\widetilde{c}_B$ based on the assumed internal friction coefficient. Inclusion phases are assumed to remain elastic because of their higher strength, so no yield criterion is applied.

RF-FEM geometry is generated according to a covariance function that defines the correlation of the random field based on spatial separation $r = \|x_i - x_j\|$ between position x_i and x_j in the three-dimensional grid used to define the random field. The covariance function is given by

$$B(r) = s_N^2 \exp\left(2\,|r|\,/L_c\right) \tag{10}$$

which is defined by the variance s_N^2 of the measured mechanical property PDF and the correlation length L_c. An increase in correlation length results in larger correlated regions within the microstructure. A periodic Gaussian random field $y_N(x)$ is generated with a fast Fourier transform (FFT) algorithm (Roberts and Teubner 1995) and then transformed to the desired lognormal random field using $y_{LN}(x) = \exp\left(m_N + y_N(x)\right)$, where m_N is the mean of the measured mechanical property PDF. This procedure results in a correlated random field representing the continuous matrix of cement paste with an overall mechanical property distribution that is based on nanoindentation measurements of the hydration products. A composite model representing the complete cement paste microstructure can be obtained by inserting inclusion phases of random size, shape, and position within the periodic material domain. The Young's modulus of inclusions is taken from a uniform distribution spanning from 60 to 160 GPa. The influence of the covariance function on the random field geometry is shown along with examples of a representative matrix and composite model in Fig. 11. Additional details regarding the implementation of random field models are described elsewhere (Palkovic et al. 2018).
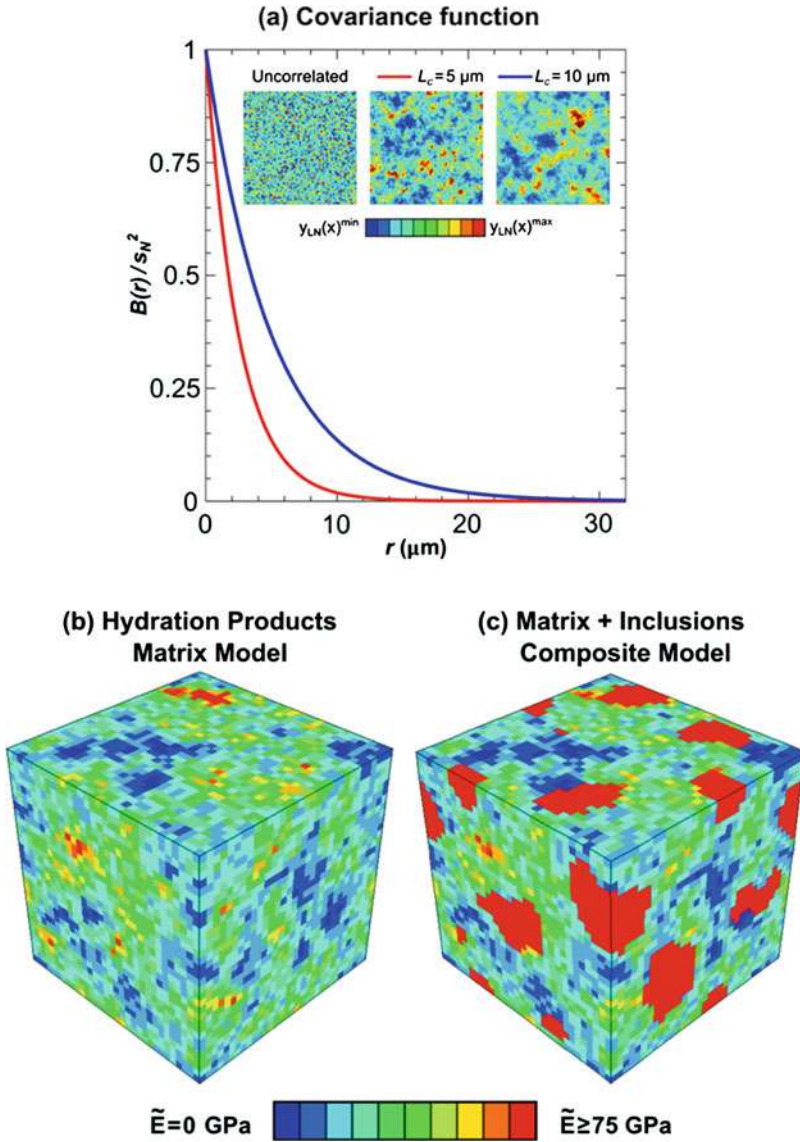
Fig. 11 RF-FEM model development. (**a**) The spatial geometry of the continuous matrix is set by the covariance function. The local Young's modulus distribution is shown for a representative (**b**) matrix and (**c**) composite model (Palkovic et al. 2018)

Fig. 12 Response of 25 realizations of matrix and composite RF-FEM models to uniaxial compression with varying confinement and local internal friction coefficient. (**a**) Uniaxial stress-strain response along the loading axis. (**b**) Relationship between volumetric strain and applied axial compressive strain (Palkovic et al. 2018)
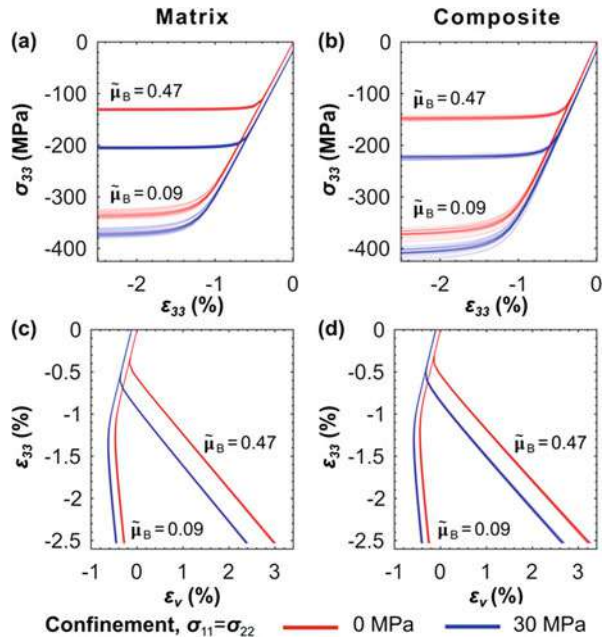
5.2 Mechanical Response of Cohesive-Frictional Microstructural Models

RF-FEM models are subjected to uniaxial compression with varying magnitudes of lateral confinement. Confinement is applied through a constant normal stress applied to opposing faces of the cubic model domain. Results for 25 realizations of matrix and composite models with varying local internal friction coefficient $\tilde{\mu}_B$ are shown in Fig. 12. The uniaxial stress-strain ($\sigma_{33} - \varepsilon_{33}$) relationship shows an initial elastic regime followed by perfectly plastic behavior after yielding. Models with a higher $\tilde{\mu}_B$ show a decreased compressive strength and less sensitivity to random field geometry and are more strongly influenced by confinement. The material volumetric response is also highly dependent on $\tilde{\mu}_B$. All models show an initial decrease in volumetric strain ε_v associated with compaction, followed by a dilatant response after yielding. More dilation is observed for models with higher $\tilde{\mu}_B$. When inclusions are added to the heterogeneous matrix, a small increase in the uniaxial compression strength is observed with little change in volumetric behavior.

The compressive strength as a function of confinement can be used to evaluate the effective cohesive-frictional properties of the microstructure. The Mohr-Coulomb fit for a representative matrix and composite material at low and high $\tilde{\mu}_B$ is shown in Fig. 13. Comparing matrix and composite models, the macroscale internal friction coefficient of the homogenized microstructure is not influenced by the presence of inclusions. The macroscale values are also almost unchanged from the local

Fig. 13 Mohr-Coulomb strength properties of a representative matrix and composite microstructure model with varying local internal friction coefficient $\tilde{\mu}_B$ (Palkovic et al. 2018)
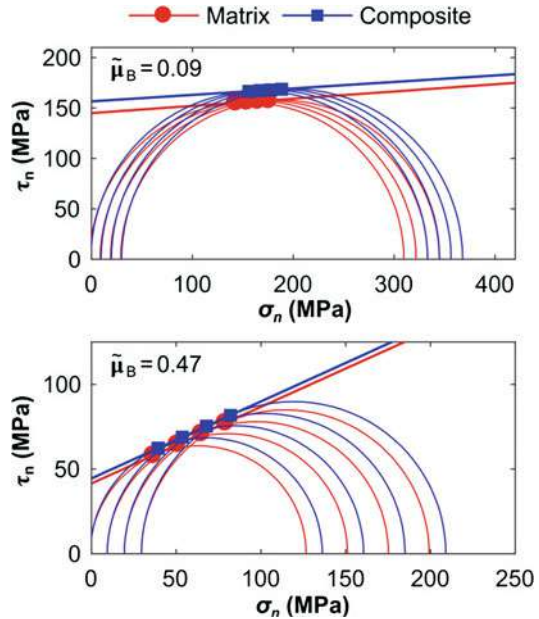
internal friction coefficient input for the yield criterion of hydration products. This observation is true for models with varying correlation length, model size, and measured lognormal PDFs (Palkovic et al. 2018). These findings suggest that microstructural features do not strongly influence the effective internal friction coefficient. Therefore, it is the layered molecular structure and colloidal interfaces that control the inelastic dilatancy and plasticity of cementitious materials.

Microstructure models are compared with uniaxial compression experiments by calculating the porosity intrinsic to hydration product measurements obtained with nanoindentation. Nanoindentation tests satisfying scale separability for cement paste probe a material volume on the order of 1 to 2 μm, which is large enough to include gel and smaller capillary pores (Chen et al. 2010). The internal porosity can be estimated using a self-consistent micromechanical model which is commonly given by (Constantinides and Ulm 2007)

$$P^{SC} = \frac{1}{2}\left(1 - \frac{\langle M^{HP}\rangle}{M^{HP}_{\max}}\right) \tag{11}$$

where $\langle M^{HP}\rangle$ is the expected value of the measured indentation modulus distribution and $M^{HP}_{\max} = 75$ GPa is the threshold value used to separate phases. Applying Eq. 11, the porosity of matrix models is 29%, while the porosity of composite models is 25% if we assume no porosity within inclusion phases. The porosity of 5-cm cement paste cubes is calculated using empirical relationships for OPC materials that are dependent on the sample w/c ratio and hydration degree (Nielsen

Fig. 14 Comparison of uniaxial compression experiments on 5-cm cement paste cubes with microstructure model simulations for matrix and composite models
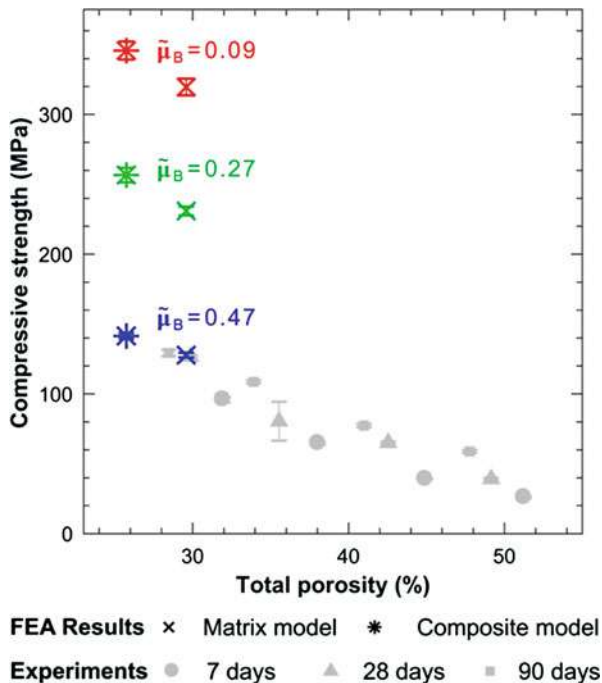
1993). The comparison of models and experiments is shown in Fig. 14 for varying local internal friction coefficients. Models with higher $\widetilde{\mu}_B$ show better agreement with macroscale experiments as a result of the decreased local cohesive strength of hydration products in the cement paste matrix. Triaxial loading experiments on cement paste and mortar materials that are converted from a Drucker-Prager to Mohr-Coulomb cohesive-frictional strength model find a macroscopic friction coefficient ranging from 0.42 to 0.60, providing further confirmation of model results (Heukamp et al. 2001, 2003).

6   Summary and Future Directions

This chapter tracks the development of cohesive-frictional interactions within cementitious materials from molecular to microstructural scales. The Mohr-Coulomb strength properties that were calculated at each length scale are summarized in Table 2. These values are specific to the parameter of the models shown and will vary depending on the force fields, interactions, and structures considered. However, the main observations and trends are expected to be applicable to a broad range of cementitious materials.

   At molecular scales, the disordered water-filled interface within the atomistic structure of C-S-H provides a region for strain localization that is the origin of

Table 2 Cohesive-frictional properties across length scales of cement paste

Model	Scale	μ	c (MPa)	Comments
Molecular	~1 nm	0.05–0.15	2500–4000	Cohesive-frictional behavior originates within the interface between calcium-silicate layers, likely due to shear transformations of disordered region
Colloidal	~100 nm	0.30–0.40	5–200	Amplification of μ due to shear resistance of particle interactions. Reduced c from gel porosity, capillary porosity, and adhesion energy of interactions
Microstructure	~100 μm	0.42–0.60	5–50	Properties are dominated by the cohesive-frictional inputs of hydration products forming a matrix. Inclusions and randomness of microstructure affect c, but not μ

Notes: Range of molecular values from unpublished models with Ca/Si ratio varies from 1.1 to 1.9, whereas results in Sect. 3 use 1.70. Colloidal results are based on models with adhesive strength limited by adhesion energy. Microstructure values are based on experimental and simulation observations, which suggest that a higher friction coefficient of cementitious materials is appropriate

cohesive-frictional behavior. This system response is attributed to shear transformations, a mechanism of plastic deformation for amorphous materials at low temperatures and strain rates. When cohesive-frictional properties are included in larger-scale interactions between colloidal C-S-H particles, the additional friction forces resist shear deformations and distribute strain throughout the material. As a result, there is a significant amplification of the bulk friction coefficient describing the inelastic dilatancy and strength asymmetry in compression and tension. The amplified friction coefficient is in better agreement with macroscale triaxial loading experiments of cement paste (Heukamp et al. 2001). Microstructure models suggest that this is because cement paste is insensitive to μm-scale features such as the presence of low volume fraction of inclusions, correlation length of hydration products, and random variations within the material microstructure. Therefore, it is the nm-scale behavior at molecular interfaces and colloidal interactions within hydration products that dominate the cohesive-frictional behavior of cementitious materials.

The development of this multiscale computational description of cementitious materials spanning from 10^{-10} to 10^{-2} m provides new possibilities for future exploration. Promising areas include the parameterization of the CFFF with molecular models of C-S-H with different chemical composition. This is especially important for understanding the role of additives which introduce additional aluminum, magnesium, and sodium ions within the atomistic structure (Lothenbach et al. 2015; Puertas et al. 2011). Using a CFFF model developed for multiple

molecular C-S-H structures allows for mixtures of colloidal C-S-H chemistries to establish chemo-mechanical relationships. Colloid geometries can also be modified to be more representative of experiments by using elliptical particles in lieu of idealized spheres. For this objective, the radial pair potential for calculating normal forces between particles could be replaced with a Gay-Berne potential that considers the necessary shape and orientation parameters (Yu et al. 2016). Finally, future research can extend microstructure models to larger scales to consider mortar and concrete. Cement paste is the matrix phase in these materials; however, the larger-scale fine and coarse aggregates do not participate in hydration reactions, and an interface transition zone (ITZ) may exist with a reduced strength and stiffness that must be considered within FEM models (Buyukozturk et al. 1972; Hillemeier and Hilsdorf 1977; Scrivener et al. 2004). Future developments in numerical modeling will provide further insights on the mechanisms controlling macroscale behavior in cementitious materials to enable the design of more sustainable and durable cementitious building materials.

References

Allen AJ, Thomas JJ, Jennings HM (2007) Composition and density of nanoscale calcium–silicate–hydrate in cement. Nat Mater 6:311–316

Argon A (1979) Plastic deformation in metallic glasses. Acta Metall 27:47–58

Bathe KJ (2006) Finite element procedures. K.J. Bathe, Watertown

Bauchy M, Qomi MA, Ulm F-J, Pellenq R-M (2014) Order and disorder in calcium–silicate–hydrate. J Chem Phys 140:214503

Bauchy M, Laubie H, Qomi MA, Hoover C, Ulm F-J, Pellenq R-M (2015) Fracture toughness of calcium–silicate–hydrate from molecular dynamics simulations. J Non-Cryst Solids 419:58–64

Bernard O, Ulm F-J, Lemarchand E (2003) A multiscale micromechanics-hydration model for the early-age elastic properties of cement-based materials. Cem Concr Res 33:1293–1309

Büyüköztürk O, Palkovic SD (2017) Multiscale modelling for sustainable and durable concrete. In: 1st international conference on construction materials for sustainable future, Zadar (Croatia) pp 28–35

Buyukozturk O, Nilson AH, Slate FO (1972) Deformation and fracture of particulate composite. J Eng Mech Div 98:581–593

Chen W-F (2007) Plasticity in reinforced concrete. J. Ross Publishing, Boca Raton

Chen JJ, Sorelli L, Vandamme M, Ulm FJ, Chanvillard G (2010) A coupled Nanoindentation/SEM-EDS study on low water/cement ratio Portland cement paste: evidence for C–S–H/Ca (OH) 2 nanocomposites. J Am Ceram Soc 93:1484–1493

Chiang W-S, Fratini E, Baglioni P, Liu D, Chen S-H (2012) Microstructure determination of calcium-silicate-hydrate globules by small-angle neutron scattering. J Phys Chem C 116:5055–5061

Constantinides G, Ulm F-J (2007) The nanogranular nature of C–S–H. J Mech Phys Solids 55:64–90

Constantinides G, Chandran KR, Ulm F-J, Van Vliet K (2006) Grid indentation analysis of composite microstructure and mechanics: principles and validation. Mater Sci Eng A 430: 189–202

Cundall PA, Strack OD (1979) A discrete numerical model for granular assemblies. Geotechnique 29:47–65

Derjaguin BV, Muller VM, Toporov YP (1975) Effect of contact deformations on the adhesion of particles. J Colloid Interface Sci 53:314–326

Double D, Hellawell A, Perry S (1978) The hydration of Portland cement. Proc R Soc Lond A: Math Phys Eng Sci 359:435–451

Ganneau F, Constantinides G, Ulm F-J (2006) Dual-indentation technique for the assessment of strength properties of cohesive-frictional materials. Int J Solids Struct 43:1727–1745

Heukamp F, Ulm F-J, Germaine JT (2001) Mechanical properties of calcium-leached cement pastes: triaxial stress states and the influence of the pore pressures. Cem Concr Res 31:767–774

Heukamp FH, Ulm F-J, Germaine JT (2003) Poroplastic properties of calcium-leached cement-based materials. Cem Concr Res 33:1155–1173

Hillemeier B, Hilsdorf H (1977) Fracture mechanics studies on concrete compounds. Cem Concr Res 7:523–535

Hou D, Zhao T, Ma H, Li Z (2015) Reactive molecular simulation on water confined in the nanopores of the calcium silicate hydrate gel: structure, reactivity, and mechanical properties. J Phys Chem C 119:1346–1358

Ioannidou K, Pellenq RJ-M, Del Gado E (2014) Controlling local packing and growth in calcium–silicate–hydrate gels. Soft Matter 10:1121–1133

Ioannidou K, Krakowiak KJ, Bauchy M, Hoover CG, Masoero E, Yip S, Ulm F-J, Levitz P, Pellenq RJ-M, Del Gado E (2016) Mesoscale texture of cement hydrates. Proc Natl Acad Sci 113:2029–2034

Jennings HM (2008) Refinements to colloid model of CSH in cement: CM-II. Cem Concr Res 38:275–289

Johnson KL (1987) Contact mechanics. Cambridge University press, Cambridge

Lesko S, Lesniewska E, Nonat A, Mutin J-C, Goudonnet J-P (2001) Investigation by atomic force microscopy of forces at the origin of cement cohesion. Ultramicroscopy 86:11–21

Lothenbach B, Scrivener K, Hooton R (2011) Supplementary cementitious materials. Cem Concr Res 41:1244–1256

Lothenbach B, Nied D, L'Hôpital E, Achiedo G, Dauzères A (2015) Magnesium and calcium silicate hydrates. Cem Concr Res 77:60–68

Lund AC, Schuh CA (2003) Yield surface of a simulated metallic glass. Acta Mater 51:5399–5411

Manzano H, Moeini S, Marinelli F, Van Duin AC, Ulm F-J, Pellenq RJ-M (2012a) Confined water dissociation in microporous defective silicates: mechanism, dipole distribution, and impact on substrate properties. J Am Chem Soc 134:2208–2215

Manzano H, Pellenq RJ, Ulm F-J, Buehler MJ, van Duin AC (2012b) Hydration of calcium oxide surface predicted by reactive force field molecular dynamics. Langmuir 28:4187–4197

Manzano H, Masoero E, Lopez-Arbeloa I, Jennings HM (2013) Shear deformations in calcium silicate hydrates. Soft Matter 9:7333–7341

Marland G, Boden TA, Andres RJ, (2017) Global, regional, and national fossil fuel CO_2 emissions. In: Trends: a compendium of data on global change, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tn

Masoero E, Del Gado E, Pellenq R-M, Ulm F-J, Yip S (2012) Nanostructure and nanomechanics of cement: polydisperse colloidal packing. Phys Rev Lett 109:155503

Masoero E, Del Gado E, Pellenq RJ-M, Yip S, Ulm F-J (2014) Nano-scale mechanics of colloidal C–S–H gels. Soft Matter 10:491–499

Nachbaur L, Mutin J, Nonat A, Choplin L (2001) Dynamic mode rheology of cement and tricalcium silicate pastes from mixing to setting. Cem Concr Res 31:183–192

Nielsen LF (1993) Strength development in hardened cement paste: examination of some empirical equations. Mater Struct 26:255–260

Oliver WC, Pharr GM (1992) An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. J Mater Res 7:1564–1583

Palkovic SD (2018) A multiscale computational framework for cement paste. PhD Thesis. Massachusetts Institute of Technology

Palkovic SD, Moeini S, Yip S, Büyüköztürk O (2015) Mechanical behavior of a composite interface: calcium-silicate-hydrates. J Appl Phys 118:034305

Palkovic SD, Brommer DB, Kupwade-Patil K, Masic A, Buehler MJ, Büyüköztürk O (2016) Roadmap across the mesoscale for durable and sustainable cement paste–a bioinspired approach. Constr Build Mater 115:13–31

Palkovic SD, Yip S, Büyüköztürk O (2017a) A cohesive-frictional force field (CFFF) for colloidal calcium-silicate-hydrates. J Mech Phys Solids 109:160–177

Palkovic SD, Yip S, Büyüköztürk O (2017b) Constitutive response of calcium-silicate-hydrate layers under combined loading. J Am Ceram Soc 100:713–723

Palkovic SD, Kupwade-Patil K, Yip S, Büyüköztürk O (2018) Random field finite element models with cohesive-frictional interactions of a hardened cement paste microstructure. J Mech Phys Solids 119:349–368

Pellenq RJ-M, Kushima A, Shahsavari R, Van Vliet KJ, Buehler MJ, Yip S, Ulm F-J (2009) A realistic molecular model of cement hydrates. Proc Natl Acad Sci 106:16102–16107

Pichler B, Hellmich C, Eberhardsteiner J (2009) Spherical and acicular representation of hydrates in a micromechanical model for cement paste: prediction of early-age elasticity and strength. Acta Mech 203:137–162

Plassard C, Lesniewska E, Pochard I, Nonat A (2004) Investigation of the surface structure and elastic properties of calcium silicate hydrates at the nanoscale. Ultramicroscopy 100:331–338

Puertas F, Palacios M, Manzano H, Dolado J, Rico A, Rodríguez J (2011) A model for the CASH gel formed in alkali-activated slag cements. J Eur Ceram Soc 31:2043–2056

Qomi MA, Krakowiak K, Bauchy M, Stewart K, Shahsavari R, Jagannathan D, Brommer D, Baronnet A, Buehler M, Yip S (2014) Combinatorial molecular optimization of cement hydrates. Nat Commun 5:4960

Richardson I (2008) The calcium silicate hydrates. Cem Concr Res 38:137–158

Roberts A, Teubner M (1995) Transport properties of heterogeneous materials derived from Gaussian random fields: bounds and simulation. Phys Rev E 51:4141

Rottler J, Robbins MO (2001) Yield conditions for deformation of amorphous polymer glasses. Phys Rev E 64:051801

Scrivener KL, Crumbie AK, Laugesen P (2004) The interfacial transition zone (ITZ) between cement paste and aggregate in concrete. Interface Sci 12:411–421

Skinner L, Chae S, Benmore C, Wenk H, Monteiro P (2010) Nanostructure of calcium silicate hydrates in cements. Phys Rev Lett 104:195502

Torquato S (2013) Random heterogeneous materials: microstructure and macroscopic properties. Springer, New York

Ulm F-J, Vandamme M, Bobko C, Alberto Ortega J, Tai K, Ortiz C (2007) Statistical indentation techniques for hydrated nanocomposites: concrete, bone, and shale. J Am Ceram Soc 90:2677–2692

Ulm F-J, Vandamme M, Jennings HM, Vanzo J, Bentivegna M, Krakowiak KJ, Constantinides G, Bobko CP, Van Vliet KJ (2010) Does microstructure matter for statistical nanoindentation techniques? Cem Concr Compos 32:92–99

Van Duin AC, Dasgupta S, Lorant F, Goddard WA (2001) ReaxFF: a reactive force field for hydrocarbons. Chem A Eur J 105:9396–9409

Vandamme M, Ulm F-J (2009) Nanogranular origin of concrete creep. Proc Natl Acad Sci 106:10552–10557

Vandamme M, Ulm F-J, Fonollosa P (2010) Nanogranular packing of C–S–H at substochiometric conditions. Cem Concr Res 40:14–26

Yu Z, Zhou A, Lau D (2016) Mesoscopic packing of disk-like building blocks in calcium silicate hydrate. Sci Rep 6:36967

Part X
Modeling the Structural Development and Mechanics of Complex Soft Materials

Modeling the Structural Development and the Mechanics of Complex Soft Materials: Overview

69

Emanuela Del Gado and Roland J.-M. Pellenq

Contents

Abstract

The materials discussed in the chapters of this section are apparently very different from each other (ranging from soft gels to cement and clays), but they all share structural and dynamical complexity that cover a range of length scales. Distinct computational approaches are required for the different length scales or timescales, and a statistical analysis of spatiotemporal fluctuations is crucial

E. Del Gado (✉)
Department of Physics, Institute for Soft Matter Synthesis and Metrology, Georgetown University, Washington, DC, USA

Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA, USA
e-mail: emanuela.del.gado@georgetown.edu; ed610@georgetown.edu

R. J.-M. Pellenq (✉)
Concrete Sustainability Hub, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

MSE2, the MIT/CNRS/Aix-Marseille University Joint Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Centre Interdisciplinaire de Nanosciences de Marseille, CINaM, CNRS/Aix-Marseille Université, Marseille, France

UMI MSE, CNRS/MIT/AMU Joint Laboratory "MultiScale Materials Science for Energy and Environment", Marseille, France
e-mail: pellenq@mit.edu

to develop a deeper fundamental understanding of the material properties. Here we provide an overview of the different chapters, to highlight the connections between them and outline the emerging scientific questions.

1   Introduction

Most of the materials of interest in our everyday life and in the technological world have structural and dynamical complexity that extends over a wide range of length scales and has enormous impact on the material behavior (response to environmental changes, evolution over time etc.), posing formidable challenges to rationalizing and predicting it. Modeling and computational approaches need to address such complexity, in order to provide meaningful answers and insight. Soft materials like gels are an example and have emerged as pivotal in technologies ranging from food to personal care, smart electronics, or biotechnologies, because they can be easily processed (they can be easily stretched, squeezed, or made to flow, Fig. 1) and they have tunable mechanical properties (▶ Chap. 70, "Mechanics of Soft Gels: Linear and Nonlinear Response"). Materials such as cements or clays are not usually regarded as soft since they are mainly used in their hardened state, but it is when they are soft, while solidifying or setting, that their final microstructure develops and their distinctive properties emerge (▶ Chaps. 71, "Mesoscale Structure and Mechanics of C-S-H"; ▶ 72, "Nanoscale Composition-Texture-Property-Relation in Calcium-Silicate-Hydrates"; ▶ 73, "From Microscopic Insight to Constitutive Models: Bridging Length Scales in Soft and Hard Materials"). In all cases, disentangling structure development, dynamics, and mechanics over different length scales is needed to achieve an adequate understanding of the material properties and hence to be able to predict them.

The chapters in this section revolve around this theme: how can modeling and computational approaches include enough details of the chemistry and of the microscopic structure to capture the essential and distinctive features of such complex materials? What is the appropriate level of details needed? In which conditions can more abstract and coarse-grained or mesoscale models provide useful

Fig. 1 Complex materials, soft and hard. This section covers modeling and computational approaches for materials soft as food and hard as concrete. Crucial for all those materials is to be able to capture the development of the microstructure when they solidify. As in glasses, the resulting microstructure lacks long-range order and features frozen-in stresses that have impact on the mechanics

insight and how? How can the insight and understanding gained through modeling and computational studies contribute to the technological and societal concerns that these materials carry?

2 Microstructural and Dynamical Complexity: From Food to Cement

Gels typically comprise a soft, porous and prevalently elastic matrix or solid network interspersed in a fluid (which can be a simple Newtonian liquid or a more complex viscoelastic medium). In many cases, the porous matrix is self-assembled from colloidal or larger units that can also be aggregates, fibers, or bundles hierarchically organized. The units that compose the large-scale structure of these materials are, therefore, already complex to start with (e.g., in most gels of interest for food the colloidal units are proteins), and the additional level of organization into larger-scale structures can add further layers of complexity. Modeling and computational approaches that start from an atomistic description of these materials are typically limited to a very narrow range of length scales that are far from covering the structural, dynamical, and mechanical properties emerging from the larger-scale structure and used in the engineering applications. Coarse-grained models and numerical approaches that do not provide an atomistic description but are able to capture the essential ingredients of the self-assembled, larger-scale structure (e.g., size distributions of pores, aggregates, etc.) are therefore extremely important to unravel the relationship between microstructural features and material behaviors (▶ Chaps. 70, "Mechanics of Soft Gels: Linear and Nonlinear Response"; ▶ 71, "Mesoscale Structure and Mechanics of C-S-H"; ▶ 73, "From Microscopic Insight to Constitutive Models: Bridging Length Scales in Soft and Hard Materials"). All materials of interest are structurally amorphous solids, and their microstructures correspond to non-equilibrium, metastable states that have frozen-in stresses that depend on the material history. As a consequence, the way the microstructure develops has a significant impact on the material properties and hence has to be properly incorporated in the modeling. The chapters in this section give several insightful examples of how the structural disorder, the complexity in terms of length scales covered, and the history dependent nature of the microstructure, can be taken into account.

Concrete and cement do not come to our mind when discussing *soft* gels. Nevertheless, they are relatively close to soft gels, while cement is setting and its nanoscale amorphous structure is further developing and densifying (▶ Chap. 71, "Mesoscale Structure and Mechanics of C-S-H"). In fact, as cement dissolves in water the dissolution of the calcium silicates contained in the powder produces, among other hydration products, calcium-silicate-hydrates (C–S–H), the material at the origin of cement cohesion and concrete strength. During cement hydration, nanoscale C–S–H clusters (or particles) aggregate into gels that constitute the basis of the hardened cement paste used for constructions. The C–S–H gels are extremely cohesive and constitute the main binding agent in concrete, therefore crucial

to the material mechanical performances. Although various other compounds are present in concrete, the C–S–H are already chemically complex, since they exist with a range of different stoichiometries and are amorphous (▶ Chap. 72, "Nanoscale Composition-Texture-Property-Relation in Calcium-Silicate-Hydrates"). Their microstructure and properties result from a complex and strongly non-equilibrium aggregation process. Two of the chapters in this section cover the most recent advances in modeling and computational approaches for cement hydrates. In particular they represent a good example of how concepts developed for other materials, such as high-performance glasses or colloidal gels, have helped scientists develop novel fundamental understanding and insights into cement and concrete.

A common feature in the materials discussed in the next chapters is certainly the presence of frozen-in stresses, which develop as the material solidifies (▶ Chap. 73, "From Microscopic Insight to Constitutive Models: Bridging Length Scales in Soft and Hard Materials"). The internal stresses that develop during cement hydration and setting can have manifestations as extreme as swelling and shrinkage of the incipient solid and even lead to cracking. They are known to have an impact on the durability of the final product. In softer gels, the presence of frozen-in, localized stresses can be more subtle, but it is clearly revealed by the progressive aging of their mechanical properties as well as by the strain localization that characterize their mechanical response even at small deformations, in the linear response regime. Understanding how frozen-in stresses develop during solidification of amorphous materials and how they interplay with an imposed deformation to determine the mechanical response is one of the outstanding challenges in material science and technologies. Think, for example, 3D printing technologies, where one needs to finely control the final properties of the printed layers as the material is being squeezed out of a nozzle (e.g., as in direct writing), flows, and re-solidifies (Chandrasekaran et al. 2018). Controlling and suitably designing the rheological response of the ink (from soft gels to cement) is essential. To do so, a deeper understanding of the basic mechanisms that regulate the emergence and the evolution over time of the frozen-in stresses is needed. The chapters in this section analyze the problem of stress localization in amorphous structures at different levels: for example, through the atomistic disorder in C–S–H nanoparticles and through the mesoscale texture of C–S–H gels in ▶ Chaps. 71, "Mesoscale Structure and Mechanics of C-S-H," and in ▶ 72, "Nanoscale Composition-Texture-Property-Relation in Calcium-Silicate-Hydrates," or through the spatiotemporal statistical analysis of density and stress fluctuations in aging soft gels in ▶ Chap. 70, Mechanics of Soft Gels: Linear and Nonlinear Response". Different levels of description of the material at play require different methodologies, and you will find in this section a discussion of relevant techniques and strategies that could be applied to a variety of different materials. The area of computational mechanics and rheology for this type of complex materials is still nascent, but it has a great potential to connect the microstructural description of the material (even if simplified or coarse-grained) to its rheological response in a way that is still mainly inaccessible through experiments. In particular, the possibility to describe *local* stresses and strains in
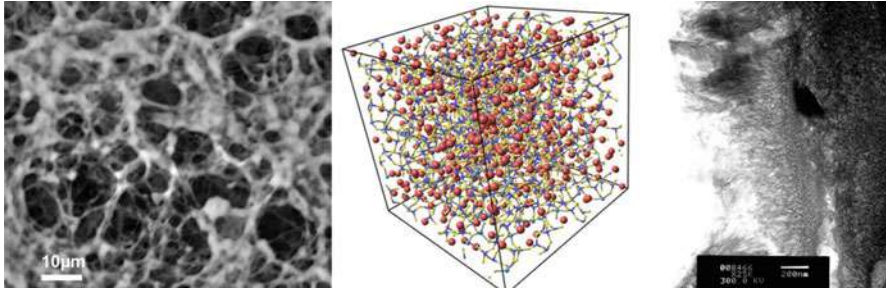
Fig. 2 Microstructural complexity in soft and hard materials. Casein gels are the basis of yogurt products: on the left a SEM image on the left a SEM image of acidified casein gel, prepared as in (Leocmach et al. 2014), courtesy of T. Divoux. Atomistic simulations highlight the structural complexity of high-performance glasses (center) (Bauchy et al. 2017). On the right, a TEM image of the nanoscale structure of cement hydrates. (Reprinted from Del Gado et al. 2014)

response to different types of deformation and deformation rates can radically deepen the current understanding of the mechanics of soft complex materials and in particular of their nonlinear behavior, as discussed in ▶ Chap. 70, "Mechanics of Soft Gels: Linear and Nonlinear Response". While these aspects are here mainly analyzed for soft gels, the next step ahead is to transfer the new concepts and insights gained across materials and technologies (Fig. 2).

Constructing methods and strategies to bring the microscopic information (e.g., chemical composition, atomistic description etc.) into coarse-grained models and, from there, into the continuum level description of the materials is an open, outstanding question. You will see it appear through in all the chapters of this section and discussed more thoroughly in ▶ Chap. 73, "From Microscopic Insight to Constitutive Models: Bridging Length Scales in Soft and Hard Materials," where we sketch a possible strategy. The modeling and computational difficulties to be faced are many and vary depending on the specific material and on the specific engineering context. An important point is that, when materials have the multiscale complexity described in the examples you will see in this section, one needs to develop different approaches at different length scales, and therefore the main question becomes how to connect those approaches.

3   Conclusions

To summarize, the chapters of this section provide examples of how modeling and computational studies can indeed address the structural, dynamical, and mechanical complexity of technologically relevant real materials. They highlight the common traits of non-equilibrium conditions, structural disorder, and frozen-in stresses in a wide range of materials from soft gels to hard cement and unravel fundamental physical mechanisms (microscopic strain and stress localization, stress relaxation, etc.) that ultimately are at the origin of behaviors such as aging, creep, failure, or yielding. The workshop on "Physics of Dense Suspensions" held at the Kavli

Institute of Theoretical Physics of UCSB in 2018 hosted a number of events and several discussions focused on the very same issues. The progress made when addressing such issues is beyond the understanding of a complex material or of a complicated behavior, because the fundamental insights gained can be immediately fed into engineering questions with practical relevance. In the case of cement, the fundamental understanding of the physical chemistry and mechanics is required for new, greener technologies to develop and emerge. Given the volume of construction industry, the role that cement and concrete play in it and the overall environmental impact of cement production for concrete, new strategies to make the material and its production smarter and more sustainable are urgently needed.

References

Bauchy M, Wang M, Yu Y, Wang B, Krishnan NA, Masoero E, Ulm F-J, Pellenq R (2017) Topological control on the structural relaxation of atomic networks under stress. Phys Rev Lett 119(3):035502

Chandrasekaran S, Yao B, Liu T, Xiao W, Song Y, Qian F, Zhu C, Duoss EB, Spadaccini CM, Li Y, Worsley MA (2018) Direct ink writing of organic and carbon aerogels. Materials Horizons 5:1166

Del Gado E, Ioannidou K, Masoero E, Baronnet A, Pellenq RM, Ulm FJ, Yip S (2014) A soft matter in construction – statistical physics approach to formation and mechanics of C–S–H gels in cement. Eur Phys J Spec Top 223(11):2285–2295

Leocmach M, Perge C, Divoux T, Manneville S (2014) Creep and fracture of a protein gel under Stress. Phys Rev Lett 113:038303

Mechanics of Soft Gels: Linear and Nonlinear Response

70

Mehdi Bouzid and Emanuela Del Gado

Contents

Abstract

Soft gels are materials at the core of material technological innovation, and as such, they are constantly evolving to meet different requirements in terms of performance, reliability, durability, and environmental impact. Despite many progresses made in the case of polymer gels, a consistent theoretical framework for the relationship between the microscopic structure and the mechanical properties of a wide range of materials ranging from colloidal gels to protein and biopolymer gels is still lacking. A multitude of different phenomena are

M. Bouzid (✉)
LPTMS, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Orsay, France
e-mail: mehdi.bouzid@u-psud.fr; mb1853@georgetown.edu

E. Del Gado
Department of Physics, Institute for Soft Matter Synthesis and Metrology, Georgetown University, Washington, DC, USA

Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA, USA
e-mail: ed610@georgetown.edu

observed – aging, strain stiffening, creep, banding, and fracture – that are difficult to control and properly tune to design the material properties. Here we discuss how numerical simulations of suitably designed microscopic models can help develop novel insight into the microscopic mechanisms that underlie the complex dynamics of these versatile materials. We provide an overview of the computational approach we have recently developed and of the main outcomes obtained. Finally we discuss outstanding questions and future developments.

1   Introduction

Gels are amorphous materials composed of a liquid phase and an interconnected network like solid phase, which can be formed at very low solid volume fractions. They are widespread, both in nature and in industry – from the pharmaceutical industry (agar gel) to the construction sector (cement hydrate or aluminosilicate gels in cement) – or in everyday life: the bread gets its rubbery texture from the formation of an elastic network of gluten protein, and the formation of protein networks is vital for the successful production of cheese. All soft matter, in fact, from proteins (Lieleg et al. 2011) to colloids (Lu et al. 2008; Gao et al. 2015) and polymers (Chen et al. 2010), easily self-assembles into such weakly elastic solids (see some examples in Fig. 1). They generally present a heterogeneous structure which is the result of the coupling between the kinetics of aggregation, phase separation or demixing, and arrest of the microscopic dynamics (Trappe et al. 2001; Del Gado et al. 2004; Coniglio et al. 2006; Del Gado and Kob 2010; Varrato et al. 2012; Zia et al. 2014).

The solidification process typically induces mechanical heterogeneities and hence internal stresses in the material, which then affect its properties in the long term (Cipelletti and Ramos 2005; Maccarrone et al. 2010a; Guo et al. 2011; Angelini et al. 2014; Bandyopadhyay et al. 2004). Weaker regions, for example, may yield via the sudden and irreversible rupture of bonds between the particles, which can be triggered by thermal fluctuations, while other regions densify. These "micro-collapses" or other aging processes can affect the stability of the material and the reliability of its performance (the mechanical strength of a gelled product can decay, for example, or batteries composed of colloidal gels may fail due to a progressive conductivity loss favored by aging (Youssry et al. 2013; Duduta et al. 2011; Helal et al. 2016; van Doorn et al. 2018)). The mechanical properties of the gel, therefore, depend on both the solidification protocol and the age of the sample. These problems are, in most of the cases, treated empirically because of the lack of fundamental understanding.

Even without the aging, it would be difficult to disentangle the impact of the structure, the dynamics, and the mechanical response in soft gels, because of the topological complexity of the solid network and of the microscopic relaxation mechanisms involved. The microscopic dynamics of gel networks is strongly cooperative and nonlocal (Colombo et al. 2013) due to the coupling among processes occurring over different lengthscales, a fact that has traditionally made theory, experimental, and computational investigations extremely challenging. Capturing
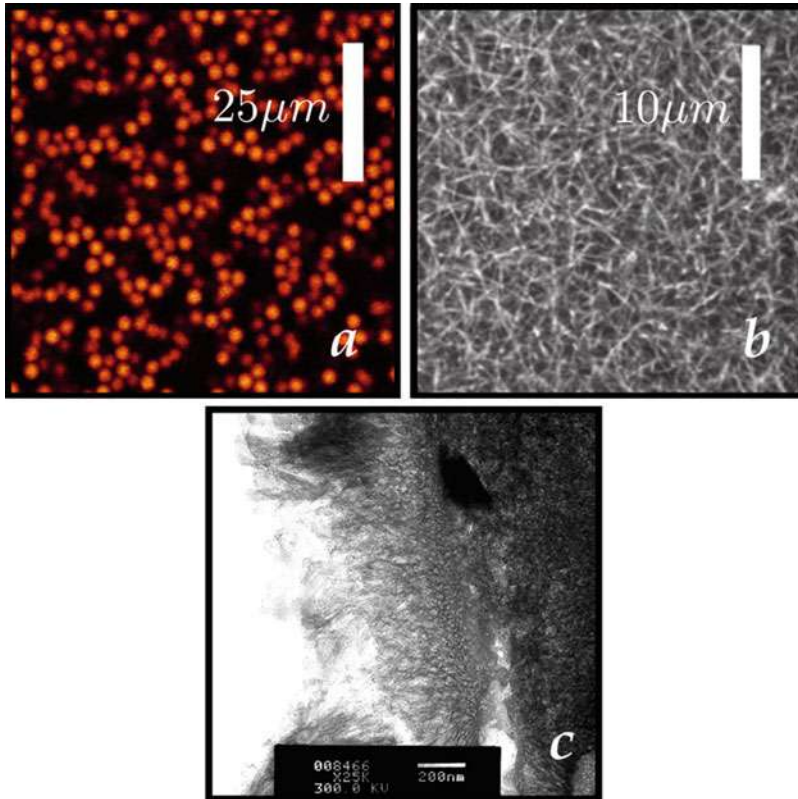
Fig. 1 (**a**) Snapshot of a colloidal particle gel (PMMA). (Reprinted from Tsurusawa et al. 2018). (**b**) A biopolymer gel (Fibrin). (Reprinted from de Cagny et al. 2016), (**c**) bright-field TEM images of calcium-silicate-hydrates C–S–H. (Reprinted from Del Gado et al. 2014)

the mechanics of soft gels and their intrinsically nonequilibrium and nonlinear nature requires bridging the lengthscale of the contacts between the particles to the lengthscale of the particles themselves and to the lengthscales relevant to the mechanics of the networks, which can include much larger structures. In addition, the presence of the fluid that can flow between the deformable pores contributes to the mechanical complexity (de Cagny et al. 2016). Under an imposed load or deformation, a multitude of different phenomena are observed – aging, strain stiffening, creep, banding, and fracture – that are difficult to rationalize, control, and properly tune to design the material properties. A key question to ask, to address this complexity, is whether or not there can be common microscopic underpinnings for the relaxation/deformation mechanisms that lead to aging at rest and to the mechanical response for different types of gel. In the last years, novel technologies have allowed experimentalists to combine rheology with imaging (Cerbino and Trappe 2008; Arevalo et al. 2015), ultrasound velocimetry (Gallot et al. 2013), interferometry (Mao et al. 2017), or spectroscopy (Ruta et al. 2012; Eberle and

Porcar 2012; Maccarrone et al. 2010b; van der Kooij et al. 2018; Aime et al. 2018b). Combining such approaches with computational studies and numerical simulations creates new unique opportunities to fill the gap between the macroscopic rheological behavior of the materials and their micro- and even nanoscale structure/dynamics and provides novel significant insights to develop advanced theories and constitutive models (Nicolas et al. 2018; Fielding 2014; Bouzid et al. 2018a).

In this chapter, we give an overview of new insights that can be gained through judiciously designed computational studies and models and of how the progress made can help understand at a more fundamental level the dynamical and mechanical complexity of soft gels. Building on the results obtained, we discuss how to tackle outstanding questions and how to bridge the microscopic picture obtained toward mesoscale or continuum level theories. The chapter is organized as follows: in Sect. 2, we present the computational approach we have recently developed for investigating the dynamics and the mechanics in a soft gel model. In particular this approach has been helpful to unravel the different nature of dynamical fluctuations in soft gels, depending on the internal stresses and the elasticity of the material, as discussed in Sect. 3. The mechanical response and its connection with the microstructure are discussed in detail in Sect. 4. Finally, in Sect. 5, we draw conclusions of the progress made so far and review some of the outstanding questions.

2 Computational Approach and Numerical Model

In the need to deepen our understanding of the interplay between the complex microstructure, the emerging dynamics, and the mechanical response of soft gels, numerical simulations play an increasingly important role. A key reason is the unique capability of computational methods to integrate different levels of complexity in a controlled manner. In condensed matter physics, the theories aimed at explaining macroscopic experimental observations are very rarely ab initio (Marx and Hutter 2009). They are based on simplified hypotheses, integrating a conceptual representation, covering a certain scale in time or space. The numerical simulations make it possible either to verify the validity of a theory by simulating the materials within the specific framework of the hypotheses used or to test the validity of the hypothesis of the theory, and to better understand their emergence, by simulating the material on a more microscopic scale. Ab initio numerical approaches are mainly useful when the microscopic details are sufficiently simple, but in most of other cases, coarse-grained approaches are more versatile and powerful. Being able to vary, in a continuous and parametrized way, the different microscopic properties is a big advantage of the numerical methods, because it makes it possible to offer a field of systematic studies out of reach of the experimental systems and, thus, to help to reveal new explanatory concepts.

Particle-based and molecular dynamics inspired approaches have proven effective, in the last few years, to unravel the dynamics and nonlinear mechanics of gel networks, thanks to large-scale simulations that can help disentangle the dynamical

processes at the level of the network structure and their contribution to the material's linear and nonlinear response (Colombo and Del Gado 2014b; Bouzid and Del Gado 2018; Bouzid et al. 2018a; Landrum et al. 2016; Jamali et al. 2017; Boromand et al. 2017; Di Michele et al. 2014; Varga and Swan 2018; Padmanabhan and Zia 2018). In these types of approaches, gels are described in terms of their microscopic discrete building blocks, typically particles, grains, or small aggregates in colloidal gels or flocculated suspensions. With a suitable model for the effective interactions between the particles, the time evolution of the system is obtained by integrating the many-body equations of motion for the whole set of microscopic degrees of freedom (typically, center of mass positions and velocities of the particles). The simulations therefore necessarily contain the following ingredients: (i) a model that describes the interactions between the particles, usually in terms of an interaction potential $\mathscr{U}(\mathbf{r}_i, \ldots, \mathbf{r}_N)$, where $\mathbf{r}_i$ represents the particle coordinates, and (ii) a numerical integrator, i.e., an algorithm that solves efficiently and precisely the equations of motion with a set of boundary conditions that corresponds to the specific problem of interest. The equations of motion can have different amount of details of the microscopic motion of the particles that compose the gel (from Brownian fluctuations to hydrodynamic interactions). With such an approach, one can introduce interactions between particles in a simple and controlled way and ask about the emergence, at various scales, of the constitutive properties of gels.

For the effective interactions, a number of approaches have been used in the literature to model soft gels, including short-range isotropic interactions typical of colloidal suspensions, valence-limited and patchy particle models that mimic small molecule or functionalized nanoparticle gels, and dipolar interactions that can lead to chaining and branching (Koumakis and Petekidis 2011; De Candia et al. 2006; Zaccarelli et al. 2006; Coniglio et al. 2004; Del Gado et al. 2004; Rovigatti and Sciortino 2011; Blaak et al. 2007; Ilg and Del Gado 2011; Eberle et al. 2011; Bianchi et al. 2015; Zia et al. 2014; Varga et al. 2015; Fierro et al. 2008). While these models mainly focus on the two-body short-range interactions that create the mechanical contact between particles at first, the emerging softness of the gel materials is the result of the sparse nature of the network structure, whose connections, on the other hand, need to be fairly rigid to support at least the gel own weight and a finite torque (Del Gado and Kob 2010, 2007; Pantina and Furst 2005, 2006; Ohtsuka et al. 2008; Colombo and Del Gado 2014a; Hsiao et al. 2012). We have therefore focused on a model where a short-range attractive well is combined with a three-body term, which imparts an angular rigidity to the gel branches, as to be expected in such open structures.

2.1   A Microscopic Model with Directional Interactions

We have developed a minimal model, considering particles (or small aggregates represented as particles) of diameter d that interact through a potential composed of two terms:

$$\mathcal{U}(\mathbf{r}_i, \ldots, \mathbf{r}_N) = \varepsilon \left[\sum_{i>j} \mathcal{U}_2 \left(\frac{\mathbf{r}_{ij}}{d} \right) + \sum_i \sum_{j>k}^{j,k \neq i} \mathcal{U}_3 \left(\frac{\mathbf{r}_{ij}}{d}, \frac{\mathbf{r}_{ik}}{d} \right) \right] \tag{1}$$

where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, with $\mathbf{r}_i$ denoting the position vector of the i-th particle and ε the strength of the attraction that sets the energy scale. The first contribution to $\mathcal{U}$ is a two-body potential à la Lennard-Jones, $\mathcal{U}_2$, which consists of a repulsive core and a narrow attractive well that can be expressed in the following dimensionless (and computationally convenient) form :

$$\mathcal{U}_2(\mathbf{r}) = A \left(\frac{a}{r^{18}} - \frac{1}{r^{16}} \right) \tag{2}$$

where r is the distance rescaled by the particle diameter d, while a and A are dimensionless parameters that control, respectively, the width and the depth of the potential well. The second contribution to $\mathcal{U}$ is a three-body term $\mathcal{U}_3$ that confers an angular rigidity to the interparticle bonds, which prevents the formation of dense clusters (Fig. 2). The idea is that once particles start to aggregate, their mechanical contacts can be more complex than the simple geometric contact between two perfect spheres represented by the attractive well $\mathcal{U}_2$ and, in particular, that there may be an energy cost associated to rotation of particles around each other in an aggregate made by more than two of them, due to inhomogeneity of their surfaces. For two particles both bonded to a third one and whose relative positions with respect to it are represented by the vectors $\mathbf{r}$ and $\mathbf{r}'$ (also rescaled by the particle diameter), $\mathcal{U}_3$ takes the following form:

$$\mathcal{U}_3(\mathbf{r}, \mathbf{r}') = B \Lambda(\mathbf{r}) \Lambda(\mathbf{r}') \exp \left[- \left(\frac{\mathbf{r} \cdot \mathbf{r}'}{rr'} - \cos \theta \right)^2 u^{-2} \right] \tag{3}$$

where B, θ, and u are dimensionless parameters. The radial modulation $\Lambda(r)$ that controls the strength of the interaction reads:

$$\Lambda(r) = r^{-10} \left[1 - (r/2)^{10} \right]^2 \mathcal{H}(2 - r) \tag{4}$$

where r is the distance rescaled by the particle diameter d. $\mathcal{H}$ denotes the Heaviside function, which ensures that $\mathcal{U}_3$ vanishes beyond the diameter of two particles. The potential energy in our model depends parametrically on the five dimensionless quantities: A, a, B, θ, and u. Tuning these parameters leads to a variety of mechanically stable porous microstructures (Colombo and Del Gado 2014a), and the values chosen in the following ($A = 6.27$, $a = 0.85$, $B = 67.27$, $\theta = 65°$ and $u = 0.3$) are such that a disordered and thin percolating network starts to self-assemble for low particle volume fractions ($\phi \lesssim 0.1$) at $\varepsilon \simeq 20 k_B T$ (see the
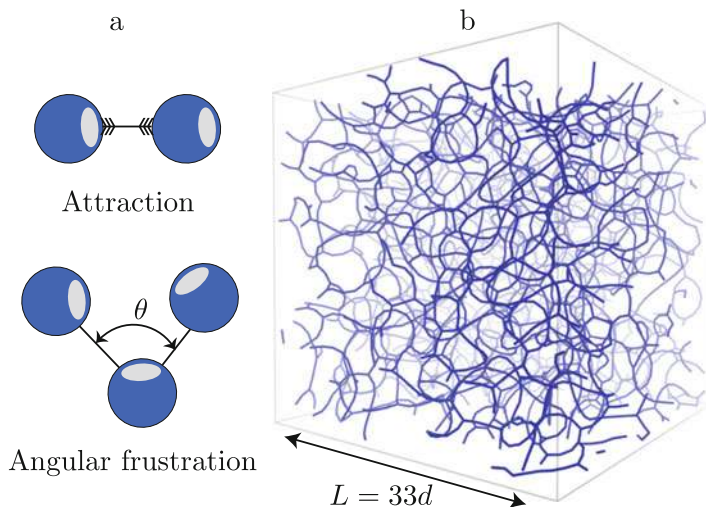
Fig. 2 (**a**) Schematic representation of the interactions: two-body attractive part and the three-body repulsive one. (**b**) Snapshot of a gel network from the simulation with number density $\rho = 0.14$, which approximately corresponds to a solid volume fraction $\phi \simeq 7.3\%$ (the sample size here is $N = 5324$ particles). Each bond is represented by a segment, when the distance d_{ij} between two particles i and j is $d_{ij} \leq 1.3d$

snapshot in Fig. 2), with k_B the Boltzmann constant and T the room temperature. Typical values of d and ε for colloidal particles range, respectively, from $d \simeq 10$ to $100\,\text{nm}$ and from $\varepsilon \simeq 10$ to $100\,k_B T$.

With this model, we have performed molecular dynamics simulations, where the gel is composed of N particles each with a mass m in a cubic simulation box of size L. The simplicity of the model allows us to run large-scale simulations with up to $N \simeq 10^6$ particles, which is essential for a statistical analysis of the microscopic dynamics. The initial gel configuration is prepared with the protocol described in (Colombo and Del Gado 2014b), which consists in starting from a gaseous configuration at $k_B T/\varepsilon = 0.5$ and letting the gel self-assemble upon slow cooling down to $k_B T/\varepsilon = 0.05$ using Newton or Langevin equations of motion. In this temperature range, the self-assembled structures do not meaningfully depend on the specific microscopic dynamics used, and the Newtonian dynamics allows for shorter computational times. Nevertheless, once the gel has formed, the Langevin dynamics is obviously a much better choice for the dynamics overdamped by the solvent, which is typical of these materials (Colombo and Del Gado 2014a; Bouzid et al. 2017).

Once the gel is assembled, the kinetic energy is then completely drawn from the system (down to $10^{-24}\varepsilon$) by means of a dissipative microscopic dynamics:

$$m\frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_{\mathbf{r_i}}\mathcal{U} - \eta_f \frac{d\mathbf{r}_i}{dt}, \tag{5}$$

where η_f is the damping coefficient associated with coupling of the particle motion to the surrounding fluid. We note that here and in the following, we always solve equations of motion that contain explicitly an inertial term with the mass m for computational convenience: integrating the equations with the inertial term allows us to use effective and precise numerical integrators (Frenkel and Smit 2002). On the other hand, the limit $m \rightarrow 0$ is the one relevant to colloidal gels in experiments, since in those systems the particle motion is completely overdamped and inertial effects are negligible. The timescales over which the particle motion is affected by inertia in our simulations are of the order of $1\tau_0 - 10\tau_0$ (for the values of m and η_f chosen). For a spherical colloidal (silica) particle of diameter $d \simeq 100$ nm and interaction strengths $\varepsilon \simeq 10k_BT$, the inertial timescale $\tau_0 = \sqrt{md^2/\varepsilon} \simeq 10^{-6}$ s corresponds to timescales (and lengthscales, in terms of particle displacements) that are not relevant to experiments (Bouzid et al. 2018a). In the following, we make sure to be in the overdamped regime of the dynamics by choosing large enough values of η_f (Colombo and Del Gado 2014a; Puosi et al. 2014).

We also note that what is still missing in the equations of motion we discuss (here and in the following) is a proper treatment of the hydrodynamic interactions that in principle can have a strong impact on the microscopic dynamics of soft gels. While a well-established rigorous framework is still under construction, recent works indicate that the role of (long range) hydrodynamics on the emerging microscopic dynamics of the gel network at rest and in the linear response regime may be relatively negligible (Varga et al. 2015; Varga and Swan 2018; Royall et al. 2015). The main difference detected so far is a slight shift of the gelation threshold (Varga et al. 2015), which could be also due to the limited size of the numerical samples, and a tendency to favor the formation of more open and anisotropic aggregates in the early stages of the aggregation, which may significantly modify the gel formation but seems to have little consequences for the long-time relaxation dynamics in the solidified gel (Varga et al. 2015; Royall et al. 2015). The explanation for such little effect could be in the fact that once the gel is formed, the main contribution to the stresses is through the gel structure, its elasticity, and the extended soft modes present. The latter are determined by the disordered network topology and involve lengthscales of the order of the sample size. Such contributions seem to dominate the gel response with respect to the hydrodynamics coupling, which would dominate only after a significant modification/rupture/destabilization of the gel structure has taken place. Further developments for the treatment of hydrodynamic interactions in such complex materials will help elucidate this point.

In all the simulations we discuss here, the timestep δt used for the numerical integration is $\delta t = 0.005$. Distances are expressed in terms of the particle diameter d, masses in units of m, the energy in terms of the strength of the attraction ε, and the time in the units of the characteristic timescale $\tau_0 = \sqrt{md^2/\varepsilon}$. From the particle number density N/L^3, we can compute an approximate solid volume fraction $\phi \simeq N\pi d^3/6L^3$. All simulations discussed here have been performed using a version of LAMMPS suitably modified by us (Plimpton 1995).

2.2   Computing Stresses and Mechanical Response

2.2.1   Stress Calculation

The average state of stress of the gel is given by the virial stresses as $\sigma_{\alpha\beta} = -\frac{1}{L^3} \sum_i s_{\alpha\beta}^i$, where the Greek subscripts stand for the Cartesian components x, y, z, L represents the size of the system, and $s_{\alpha\beta}^i$ represents the contribution to the stress tensor of all the interactions involving the particle i (Irving and Kirkwood 1950; Thompson et al. 2009). The latter contribution is calculated for each particle, by splitting the contributions of the two-body and the three-body forces according to the following equation:

$$s_{\alpha\beta}^i = -\frac{1}{2} \sum_{n=1}^{N_2} (r_\alpha^i F_\beta^i + r_\alpha' F_\beta') + \frac{1}{3} \sum_{n=1}^{N_3} (r_\alpha^i F_\beta^i + r_\alpha' F_\beta' + r_\alpha'' F_\beta'') \qquad (6)$$

The first term denotes the contribution of the two-body interaction, where the sum runs over all the N_2 pairs of interactions that involve the particle i. The couples (r^i, F^i) and (r', F') denote, respectively, the position and the forces on the two interacting particles. In the same way, the second term indicates the three-body interactions involving the particle i and two neighbors denoted by the prime and double prime quantities.

To evaluate the stress tensor at a mesoscale, we consider a coarse-graining volume v_{cg} centered around the point of interest $\mathbf{r}$ and containing around 10 particles on average and define the local coarse-grained stress based on the per-particle virial contribution as $\bar{\sigma}_{\alpha\beta}(\mathbf{r}) = - \sum_{i \in v_{cg}} s_{\alpha\beta}^i / v_{cg}$. For a typical starting configuration of the gel, the local normal stress $\bar{\sigma}_n = (\bar{\sigma}_{xx} + \bar{\sigma}_{yy} + \bar{\sigma}_{zz})/3$ reflects the heterogeneity of the structure and tends to be higher around the nodes, due to the topological frustration of the network. Note that here and in the following, we have been neglecting the kinetic contribution $-(1/L^3) \sum_i m_i \delta v_\alpha^i \delta v_\beta^i$ to the global stress tensor (from the fluctuations of the particle velocities with respect to the average), since we are mainly concerned with $\varepsilon \gg k_B T$.

2.2.2   Aging Protocol

We consider that breaking of network connections underlying the aging of the gel is more prone to happen in the regions where local stresses tend to be higher, as found also in Colombo et al. (2013) and Colombo and Del Gado (2014b). Hence aging at rest can be due to sudden ruptures of particle connections in the regions where the local tensile stresses are higher, as also suggested by experiments on ultraslow aging (Ramos and Cipelletti 2001). Such events become hardly observable on a reasonable simulation time window making it extremely challenging to study the aging regime (Colombo et al. 2013; Bouzid et al. 2017). In order to investigate the consequences of the ruptures on a timescale computationally affordable, we have devised the following dynamics: we periodically scan the local stresses in the gel structure and remove particle bonds where the local stress is the highest, with a fixed

rate Γ. In between breaking events, the gel evolution follows from the Langevin equations of motion:

$$m \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_{\mathbf{r_i}} \mathscr{U} - \eta_f \frac{d\mathbf{r}_i}{dt} + \xi_i(t), \tag{7}$$

where i refers to the particle and $\mathscr{U}$ is the interaction potential in Eq. (1). $\xi_i(t)$ is a random white noise that models the thermal fluctuations and is related to the drag coefficient η_f through its variance $\langle \xi_i(t)\xi_j(t') \rangle = 2\eta_f k_B T \delta_{ij}\delta(t-t')$. To be in the overdamped limit of the dynamics, η_f is set to 10 m/τ_0.

Recombinations of the gel branches are possible but just not observed at the volume fractions and for the time window explored here. The bonds progressively removed (by turning off the well in $\mathscr{U}_2$) are those whose contribution to the local normal stress $s_n^i = (s_{xx}^i + s_{yy}^i + s_{zz}^i)/3$ is the largest (prevalently bonds between particles belonging to the network nodes) (Bouzid et al. 2017) (see Fig. 3a). As the simulation proceeds, local internal stresses redistribute in the aging structure of the gel, and the locations of more probable connection rupture (as well as their number) change over time. All simulations presented here for the aging have been performed with a rate $\Gamma = 0.04\tau_0^{-1}$, corresponding to removing only $\sim$5% of the total network connections over the whole simulation time window.
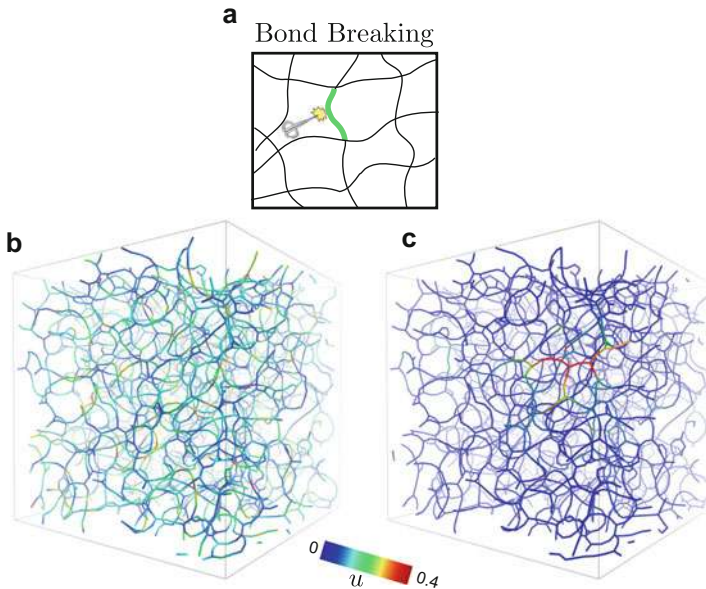
Fig. 3 (**a**) Illustration representing an irreversible elementary process in the aging simulations: breaking of a bond highlighted in green. A snapshot of the gel network is visualized using the bonds between particles, and the colors refer to the displacement field u after the bond breaking, for a network with thermal fluctuations (**b**) and for an athermal gel (**c**). (Reproduced with permission from Bouzid et al. 2017. Creative Commons license available at https://creativecommons.org/licenses/by/4.0/. Copyright 2017 Nature Publishing Group)

We have identified the range of parameters for which changing the rate (having kept all other parameters constant) does not modify the outcomes and the emerging physical picture discussed in the following. In addition to the data shown here, we have also varied the density of events, keeping fixed the criterion used for cutting (i.e., we cut the network connections corresponding to the largest tensile contributions to virial stresses). In fact, the rate of cutting events chosen for the data shown here has been chosen in a range such that the typical interval between two cutting events Γ is quite smaller than the relaxation times associated to the stress redistribution following the bond cutting. Hence many breaking events contribute to the structural relaxation that we discuss. Since for the aging we are interested to the case in which all dynamics are extremely slow (i.e., to very low temperatures in the simulations), the range of values for Γ that satisfy this condition is relatively wide. Increasing the rate corresponds to breaking a larger fraction of the total bonds initially present and hence also to increasing the density of events underlying the network rearrangements and the stress relaxation. Overall, the same results are recovered as long as the $\tau_r = 1/\Gamma$ between two rupture events allows for at least partial stress relaxation, varying Γ over nearly two orders of magnitudes. If the cutting rate is much higher, the rate may eventually disrupt the elastic strain field in the material leading to damage accumulation and spreading (Bouzid et al. 2017).

2.2.3   Start-Up Shear

To determine the gel mechanical response to a start-up shear test, each particle configuration can be submitted to a series of incremental strain steps in simple shear geometry (Colombo and Del Gado 2014b). In each step, we increase the cumulative shear strain by an amount $\delta\gamma$ by first applying an instantaneous affine deformation $\Gamma_{\delta\gamma}$, corresponding to simple shear in the xy plane, to all particles:

$$\mathbf{r}'_i = \Gamma_{\delta\gamma}\mathbf{r}_i = \begin{pmatrix} 1 & \delta\gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{r}_i \qquad (8)$$

The Lees-Edwards boundary conditions (Lees and Edwards 1972) are updated as well, to comply with the increase in the cumulative strain. The configuration $\{\mathbf{r}'_i\}$ is no longer a minimum of the potential energy (being the material amorphous) (Alexander 1998), and the small deformation step induces unbalanced internal forces. Hence we relax the affinely deformed configuration by letting the system free to evolve in time while keeping the global strain constant:

$$\mathbf{r}''_i = \mathscr{T}_{\delta t}\mathbf{r}'_i \,. \qquad (9)$$

where $\mathscr{T}_{\delta t}$ is the time evolution operator for a specified time interval δt and given by the damped dynamics:

$$m\frac{d^2\mathbf{r}_i}{dt^2} = -\eta_f\frac{d\mathbf{r}_i}{dt} - \nabla_{\mathbf{r}_i}U\,, \tag{10}$$

where m is the particle mass and η_f the coefficient of friction. After n steps, the cumulative strain is $\gamma_n = n\,\delta\gamma$, and the gel configuration is

$$\mathbf{r}_{i,n} = (\mathscr{T}_{\delta t}\Gamma_{\delta\gamma})^n\,\mathbf{r}_{i,0}\,, \tag{11}$$

where $\{\mathbf{r}_{i,0}\}$ denotes the configuration of the starting inherent structure.

   This procedure is similar to the *athermal quasistatic* (AQS) approach extensively used to investigate the deformation behavior of amorphous solids (Tanguy et al. 2002; Maloney and Lemaître 2006; Fiocco et al. 2013), with the main difference that, instead of using an energy minimization algorithm after each affine deformation step, we follow the natural dynamics of the system (with a viscous energy dissipation) for a prescribed time interval δt. We can therefore define a finite shear rate $\dot{\gamma} = \delta\gamma/\delta t$ for the deformation we apply. Disregarding effects due to the particle inertia, the microscopic dynamics (10) introduce a natural timescale $\tau = \eta_f d^2/\varepsilon$, corresponding to the time it takes a particle subjected to a typical force ε/d to move a distance equal to its size. Indicatively, if we consider a typical aqueous solution of colloidal particles with a diameter $d \approx 100\,$nm and an interaction energy $\varepsilon \approx 10 k_B T$ (Koumakis and Petekidis 2011), the characteristic time is $\tau \approx 10^{-4}\,$s; in such a system, imposing a shear rate of $0.1\,\mathrm{s}^{-1}$ would correspond to a numerical shear rate of $\dot{\gamma}_s = 10^{-5}\tau^{-1}$.

2.2.4 Small Amplitude Oscillatory Rheology

To determine the gel viscoelastic properties, the particles are submitted to a continuous shear strain $\gamma(t)$ in the xy plane according to the following equation:

$$m\frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_{\mathbf{r}_i}\mathscr{U} - \eta_f\left(\frac{d\mathbf{r}_i}{dt} - \dot{\gamma}(t)y_i\mathbf{e_x}\right) \tag{12}$$

In order to measure the frequency and the strain dependence of the first-harmonic storage G' and loss modulus G'', we impose an oscillatory shear strain on the system, i.e., the shear strain is modulated periodically according as $\gamma(t) = \gamma_0\sin(\omega_i t)$, and we use Lees-Edwards boundary conditions while applying the deformation.

   By monitoring the shear stress response of the material $\sigma(t)$ over time, we can extract the viscoelastic moduli. The storage and loss moduli can be computed from the stress response with the following expressions:

$$\begin{aligned} G'(\omega_i) &= \mathscr{R}e\left\{\frac{\tilde{\sigma}(\omega_i)}{\tilde{\gamma}(\omega_i)}\right\} \\[2mm] G''(\omega_i) &= \mathscr{I}m\left\{\frac{\tilde{\sigma}(\omega_i)}{\tilde{\gamma}(\omega_i)}\right\} \end{aligned} \tag{13}$$

where $\tilde{\sigma}$ and $\tilde{\gamma}$ are the Fourier transforms of the stress and strain signals, respectively (Macosko 1994). The whole viscoelastic spectrum is then reconstructed by performing a discrete series of tests at various frequencies, also known as "frequency sweep."

We note that the drag term used in Eq. (12) is chosen for computational convenience, but a more accurate (Galilean invariant) option would be to have the drag term depend on the relative velocities of particles within a certain range from a given one (Salerno et al. 2012; Vasisht et al. 2019), which requires longer computations. Nevertheless, the correction due to this inaccuracy seems to be negligible in all cases discussed here, since the stresses, and the related dynamics, are dominated by the interparticle interactions and the structure topology, and the correction due to the different drag term does not significantly modify the behaviors analyzed in the following.

3   Elastically Driven Dynamics upon Aging

After the gelation, the microscopic dynamics of gels slows down considerably. The nanometric size of their building blocks makes gels sensitive to thermal fluctuations, resulting in a rich and complex relaxation processes that are associated with spontaneous and thermally activated processes. The study of how these different dynamical processes emerge at rest and how the mechanical response depends on the microstructure is relevant to several applications, since the progressive aging that such materials undergo over time has a dramatic impact on their functionalities. In aging soft solids such as gels, the microscopic dynamics is expected to be akin to slower than exponential (or stretched exponential) dynamics in super-cooled liquids or glasses, due to its strongly cooperative nature fundamentally controlled by the structural disorder, as shown in several experimental and simulation studies (Segre et al. 2001; Manley et al. 2005; Bandyopadhyay et al. 2004; Jabbari-Farouji et al. 2007; Krall and Weitz 1998; Del Gado et al. 2004; Del Gado and Kob 2007; Fierro et al. 2008; Abete et al. 2008; Angelini et al. 2014; Chaudhuri et al. 2015). Nevertheless, over the last few years, time- and space-resolved measurements have often found evidence of dynamics faster than exponential (so-called compressed exponential dynamics), intermittency, and abrupt microstructural changes, raising the question of whether the aging dynamics in gels may be controlled, instead, by stress relaxation through elastic rebound of parts of the material, after local breakages occur in its structure (Cipelletti et al. 2000; Ramos and Cipelletti 2001; Bellour et al. 2003; Angelini et al. 2013; Mansel and Williams 2015; Ruta et al. 2014; Harden et al. 2012; Chung et al. 2006; Bouchaud 2008; Ferrero et al. 2014; Gao et al. 2015; Godec et al. 2014; Chaudhuri and Berthier 2017).

The difficulty to study the aging and characterize the slow relaxation processes in gels (both in experiments and in simulations) comes from the required timescales which are extremely long. The procedure briefly described in Sect. 2.2.2 allowed us to overcome some of the difficulties: we could scan the gel network and artificially remove the connections where tensile stresses are higher, to mimic the aging process

(Fig.3a) (Bouzid et al. 2017). Using that procedure, we could characterize the stress redistribution in the network during the aging and systematically vary the amount of thermal fluctuations to elucidate how the coupling between elasticity and thermal fluctuations may qualitatively change the microscopic dynamics. Figure 3b, c show the differences in the displacement field caused by breaking connections in the same gel network in the presence of different amounts of thermal fluctuations, whereas Fig. 4a–c display the initial stress heterogeneities, the stress redistribution, and the stress fluctuations during the aging. The microscopic picture that emerged from our study points to a major role played by the architecture of the network, its elasticity, and the stress heterogeneities built in upon solidification (Fig. 4a). The timescales governing stress relaxation, respectively, through thermal fluctuations and elastic recovery are key: when thermal fluctuations are weak with respect to enthalpic stress heterogeneities, the stress can still partially relax, after the breaking of network connections, through elastically driven fluctuations. Such fluctuations are intermittent, because of strong spatiotemporal correlations that persist well beyond the timescale of the simulations (Fig. 4b, 4c) or of the experiments. Hence in these conditions ($\varepsilon/k_B T \simeq 0$) the elasticity built into the solid structure controls the microscopic displacements, and the time correlations of density fluctuations decay faster than exponential, as reported in experiments and hypothesized by theories (Cipelletti and Ramos 2005; Bouchaud 2008). Brownian motion driven by thermal fluctuations, in fact, disrupts the spatial distribution of local stresses and their persistence in time, therefore favoring a gradual loss of correlations and a slow evolution of the material properties. The insight gained in this study helped us rationalize apparently contrasting findings and clarify that in elastic soft materials the presence of large stress heterogeneities can favor faster than exponential and intermittent microscopic dynamics. In addition to affecting the time evolution and the material properties rest, these dynamic processes interact with an imposed mechanical deformation and can therefore be decisive for the mechanical response of this class of soft solids.

4   Mechanical Response

A macroscopic constitutive law for soft gels, i.e., the theory capturing how stresses are related to deformations, velocities, and possibly density in these materials, is still fundamentally lacking. As an example, we give the constitutive law of a Hookean homogeneous elastic solid. In this case, the stress tensor takes the form:

$$\sigma_{\alpha\beta} = K\varepsilon_{\ell\ell}\delta_{\alpha\beta} + 2G\varepsilon_{\alpha\beta} \tag{14}$$

K represents the bulk modulus and G the shear modulus. ε_{ij} is the strain tensor of the material with respect to its reference state (Landau and Lifshitz 1986), while α and β indicate the Cartesian components. Following the same reasoning, when describing the flow of the materials in response to a rate of deformation
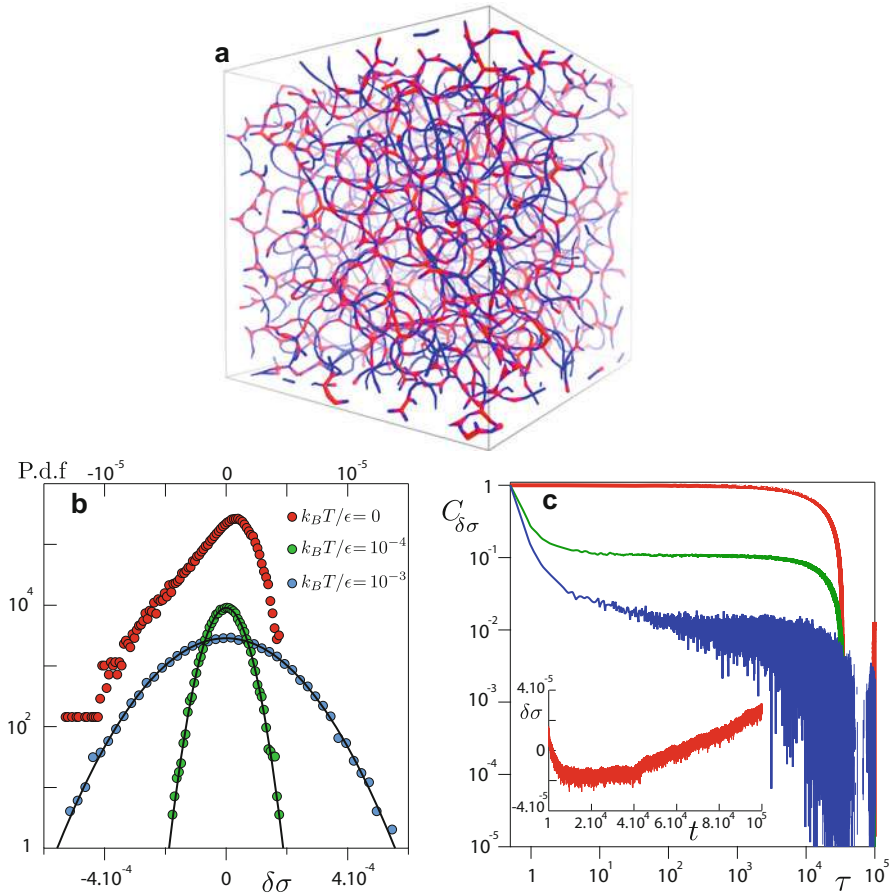
Fig. 4 (**a**) A snapshot of the initial colloidal gel network for $k_B T/\varepsilon = 0$ and at low volume fraction $\phi = 7\%$, showing the interparticle bonds represented by a segment. The colors indicate the local normal stress $\tilde{\sigma}_n$, using red for tension and blue for compression. (**b**) The corresponding P.d.f of the time series of the normal stress fluctuations $\delta\sigma = \sigma_{xx} - \langle\sigma_{xx}\rangle$: in the regime dominated by frozen-in stresses (red), the stress fluctuations are elastically driven and intermittent in nature. (**c**) Main frame: The stress fluctuation autocorrelation function measured over the whole simulation for the fully thermal regime $k_B T/\varepsilon = 10^{-3}$ (blue), the intermediate $k_B T/\varepsilon = 10^{-4}$ (green), and the for $k_B T/\varepsilon = 0$ (red). Inset: Time series of the normal stress fluctuations $\delta\sigma$ over all the simulation for the athermal sample showing the aging of the structure. (Reproduced with permission from Bouzid et al. 2017. Creative Commons license available at https://creativecommons.org/licenses/by/4.0/. Copyright 2017 Nature Publishing Group)

(e.g., due to imposed shear Fig. 7), the constitutive relation can be constructed from the strain rate tensor:

$$\dot{\gamma}_{\alpha\beta} = \frac{1}{2}\left(\frac{\partial u_\alpha}{\partial x_\beta} + \frac{\partial u_\beta}{\partial x_\alpha}\right) \tag{15}$$
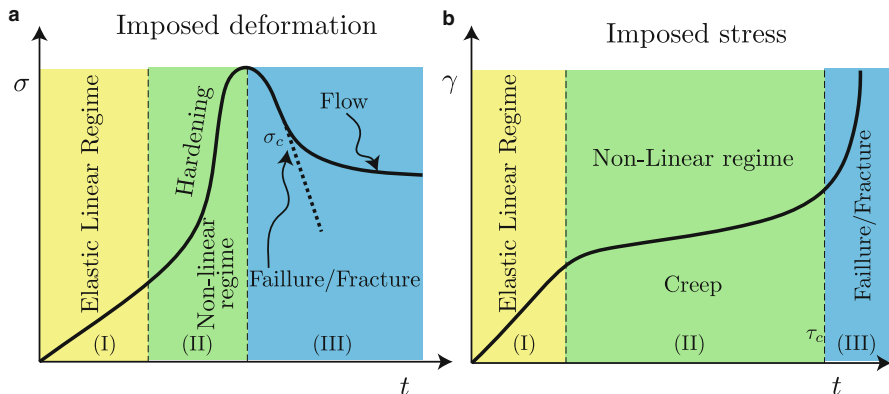
Fig. 5 Log-log scale: schematic representation of the different macroscopic responses of gels to an imposed shear deformation rate (**a**) or shear stress (**b**)

Where u_α and x_α are, respectively, the components of the velocity vector and the position along the α axis. A fluid is characterized by its viscosity η, defined as the ratio between the stress and the strain rate, and the simplest example is a Newtonian fluid, for which the stress tensor takes the form:

$$\sigma_{\alpha\beta} = (g\dot{\gamma}_{\ell\ell} - P)\delta_{\alpha\beta} + 2\eta\dot{\gamma}_{\alpha\beta} \tag{16}$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol, P is the normal stress, and g is zero for an incompressible fluid. For Newtonian fluids, the viscosity is independent of the shear rate as well as the time. The only stresses created by the flow are the shear stresses that cancel out immediately when the flow stops. Any deviation from these properties is a sign of non-Newtonian behavior.

Once a gel has solidified from an initial fluid solution, its mechanical response is typically visco-elasto-plastic (Alexander 1998; Nicolas et al. 2018), with the prevalence of one behavior or the other depending on the nature of the deformation and on the observation time (Fig. 5).

4.1   Linear Response of Soft Gels

For small deformations, the response of gels is viscoelastic (regime I in Fig. 5), often exhibiting a power law frequency dependence of the elastic (G') and the viscous (G'') moduli over a wide range of frequencies (Fig. 6), reflecting a wide distribution of time and lengthscales over which the internal structures of the gel can also relax some of the residual stresses frozen-in upon solidification. These viscoelastic characteristics have been observed experimentally in a wide range of gel materials for which the porous nature of the structure can promote contraction or dilation during shear, suggesting a nontrivial coupling between the transmission of normal
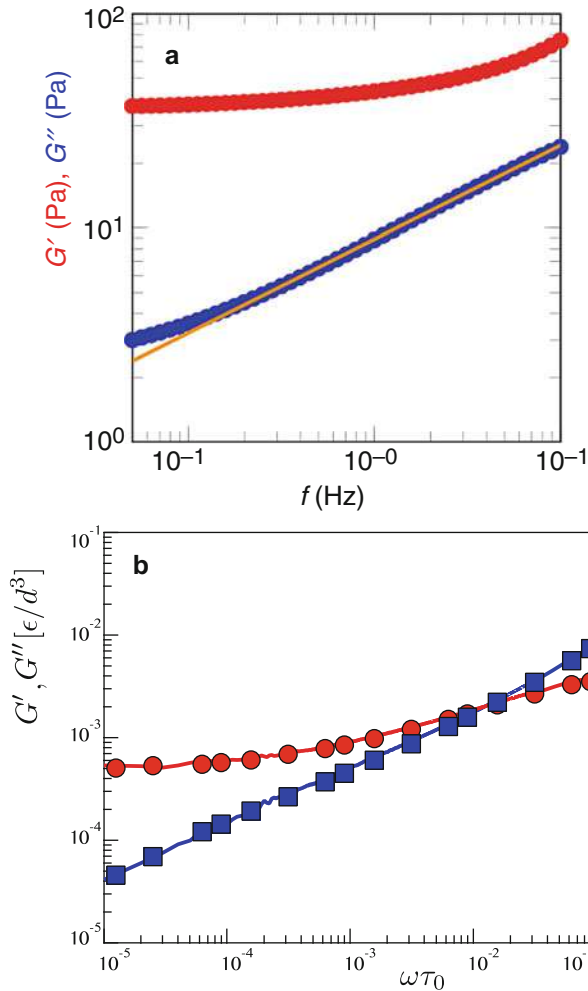
Fig. 6 (**a**) Elastic modulus G' (red) and viscous modulus G'' (blue) vs. frequency f for a carbopol gel. (Reprinted from Lidon et al. 2017). (**b**) Elastic modulus G' (squares) and viscous modulus G'' (circles) from the numerical simulations in Bouzid et al. (2018b)

and tangential forces (Ng and McKinley 2008; Caggioni et al. 2007; Jaishankar and McKinley 2014). For biopolymers networks such as collagen, G' decreases (the system becomes softer) if the system has been previously subjected to compression and increases if it is subjected to tension (Van Oosten et al. 2016). On the other hand, it has been shown that agar gels contract upon solidification (Mao et al. 2016) and that such tendency to contract is usually associated to a tendency of the material to stiffen under larger strains (Tabatabai et al. 2015; Broedersz and MacKintosh 2014; Feng et al. 2015). The changes in volume during gelation are typically coupled to

heterogeneities of internal stresses, whose consequences on the mechanical response of the material could then be controlled by adjusting the processing or preparation conditions (e.g., through the rate of confinement or gelation).

Using the model introduced in Sect. 2, recent numerical investigations have been able to reproduce these linear mechanical features (Fig. 6) and also provide a constitutive parametric model able to capture the linear response (Bouzid et al. 2018a). Classical constitutive models based on combinations of Hookean springs and dashpots, such as the Maxwell or Kelvin-Voigt model, are not able to capture the frequency dependence of the moduli (Macosko 1994). To describe these power laws, one can add different mechanical elements (Maxwell or Kelvin-Voigt) in series or in parallel, which will result in creating additional modes of relaxation of the system. This method is questionable because of the large number of elements necessary to model all the relaxation modes. A different approach is based on proposing a *fractional* model, which uses instead a spring-pot element, originally introduced by Scott Blair (Blair and Veinoglou 1944; Blair 1944) and recently applied with success to a broad variety of soft viscoelastic materials (Jaishankar and McKinley 2012; Bouzid et al. 2018a; Aime et al. 2018a).

4.2   Nonlinear Response

Upon increasing the imposed deformation, the rheological response of soft gels has strongly nonlinear characteristics, corresponding to regimes II and III in Fig. 5: soft gels can flow as yield stress fluids or fracture as solids (see examples in the cartoons of Fig. 7). The strongly nonlinear transient regimes preceding yielding or fracture are poorly understood and constitute an active area of research. In addition, the yielding or fracturing can be reversible, with the material regaining its elasticity once brought to rest, or irreversible.

Numerical simulations are a powerful tool to overcome experimental limitations and challenges, complementing them effectively. For example, imaging the structure and, at the same time, applying a deformation is not always possible, whereas simulations can be obviously more effective in connecting the structure to the
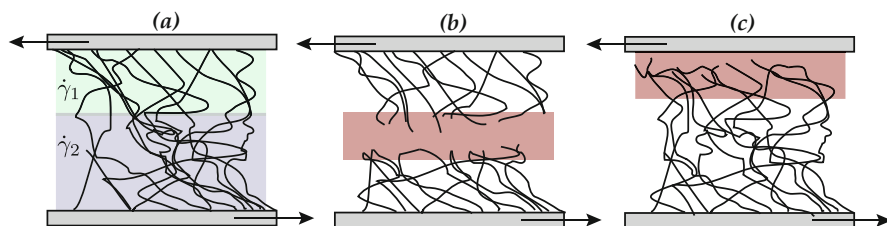
Fig. 7 Illustrations of different non-linear behaviors: (**a**) Shear band with two distinct shear rates $\dot{\gamma}_1 \neq \dot{\gamma}_2$. (**b**) Fracture in the bulk. (**c**) Loss of adhesion between the gel and the wall: sliding on the walls

mechanical response and in addressing specific microscopic mechanisms. Using the approach described in Sect. 2, for example, we have recently provided novel insight into how regimes (I–III) in Fig. 5 are sensitive to the complexity of the architecture of the network and dependent on the rate of deformation. The simulations (Fig. 8) have shown that for tightly connected networks, with smaller and more homogeneously distributed pores, the stresses can be redistributed more uniformly under the action of a mechanical load, promoting almost simultaneously rupture of many connections and favoring crack growth. Conversely, a softer and more sparsely connected network, in the same type of gel, favors the localization of the stresses and their redistribution through the abundant low frequency modes, possibly delaying microstructural damage and cracks growth to much larger deformations
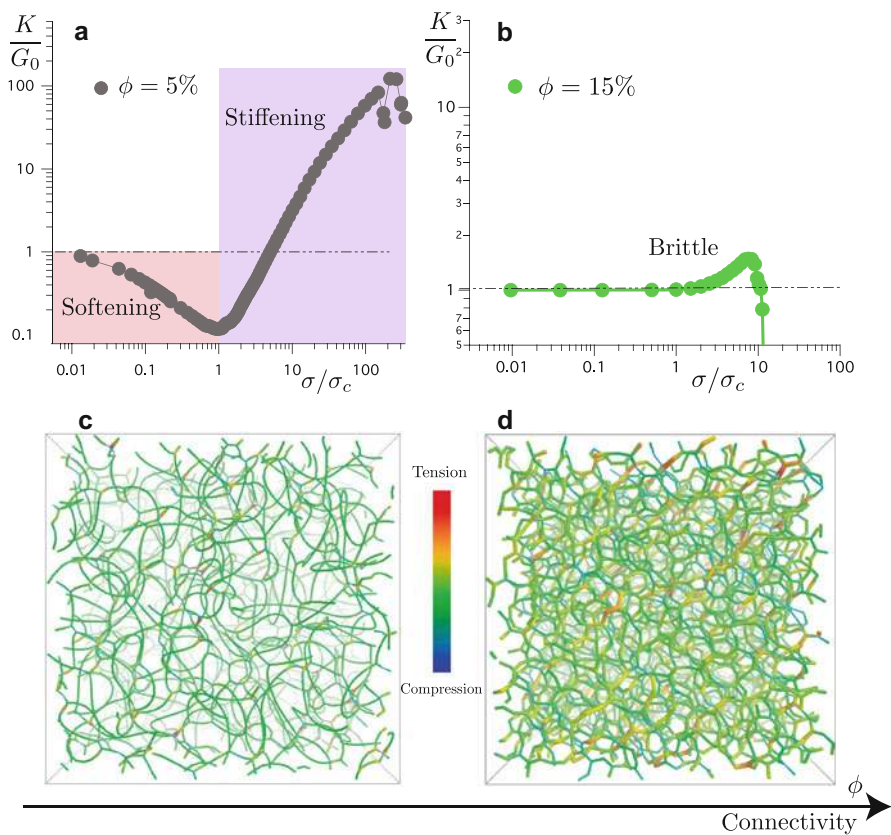
Fig. 8 The differential modulus K normalized by the elastic modulus G_0 as a function of the normalized shear stress σ/σ_c, for $\phi = 5\%$ (**a**) and $\phi = 15\%$ (**b**). Snapshots of the gel network extracted from the simulations at a volume fraction $\phi \simeq 15\%$ (**c**) and $\phi \simeq 5\%$ (**d**). Each bond is represented by a segment, when the distance d_{ij} between two particles i and j is $d_{ij} \leq 1.3d$. The color code shows the value of local tensile or compressive stresses, while the thickness is proportional to the stress amplitude. (See also Bouzid and Del Gado 2018)

(Bouzid and Del Gado 2018). Such distinct behaviors could be associated, in regime (II), to distinct dependencies of the nonlinear modulus K (obtained by deriving the shear stress σ with respect to the imposed deformation γ in the start-up shear tests described in Sect. 2) on the shear stress generated in the model materials under load (see Fig. 8). In sparse networks, extended soft modes can favor softening followed by a localization of the stresses and a progressive stiffening of the material (Figs. 8a, c). In our softer gels, the stiffening is characterized by a power law of the form $K \sim \sigma^{3/2}$, very similar the behavior of semiflexible polymer networks (Broedersz and MacKintosh 2014; Feng et al. 2016). For more homogeneously connected networks, where the gel branches are less flexible, the nonlinear response features also a progressive hardening (which is strongly rate dependent) and $K \sim \sigma$, indicating an exponential increase of the stress σ with increasing the imposed deformation γ, similar to the one observed in collagen fiber networks (Licup et al. 2015; Arevalo et al. 2015). Finally, denser networks composed of small homogeneous pores have a stronger tendency to be brittle (Fig.8b, d). It has also been shown that the ductility of the mechanical response and the reversibility of the damage accumulated under deformation can be modified by changing the flexibility of the gel branches (for a fixed network topology) and by pre-stressing the gel structures (Feng et al. 2018). Interestingly enough, in all these cases, the amount of microscopic structural anisotropy that can be induced in the gel network through deformation (and that can be quantified, for example, via a fabric tensor or a nematic order parameter suitably defined) seems a promising microscopic indicator of the nonlinear macroscopic response of different gels preceding yielding or fracture (Bouzid and Del Gado 2018; Jamali et al. 2017; Feng et al. 2015).

In the last plastic regime (III) of Fig. 5, the mechanical response depends more dramatically on the nature of the system, the specific mode of deformation (e.g., imposed stress vs deformation rate), but also on the nature of the interactions between the material and the boundaries (walls), as previously sketched in Fig. 7. Experiments show that the rupture can take place in bulk (casein-type, (bio)-polymer gels) or at the wall with a loss of adhesion and slip (agar gels, carbopol microgels, microgels of p-NIPAm, collagen, etc.) (Meeker et al. 2004; Divoux et al. 2015; Bonn et al. 2017). In spite of such complex diversity, experimental data also provide several indications that the coupling and the dynamics of localized (and correlated) plastic events may be the underlying common denominator in the transition from regime II to III. Indeed, in a variety of colloidal gels, it has been shown that in the transient regime (II) before flowing or fracturing (III), the dynamics is very intermittent in time and heterogeneous in space, featuring avalanche patterns (Kurokawa et al. 2015): in start-up shear experiments, for example, the yielding may proceed via successive drops during which the flow is spatially and temporally heterogeneous (Fig. 9a–c).

Recent experiments were able to detect a local acceleration of the microscopic dynamics before fracture and an enhancement of the plastic activity (Aime et al. 2018b). Start-up shear numerical simulations, using the approach described above, have indeed shown that the pre-yielding or pre-failure regime is dramatically rate
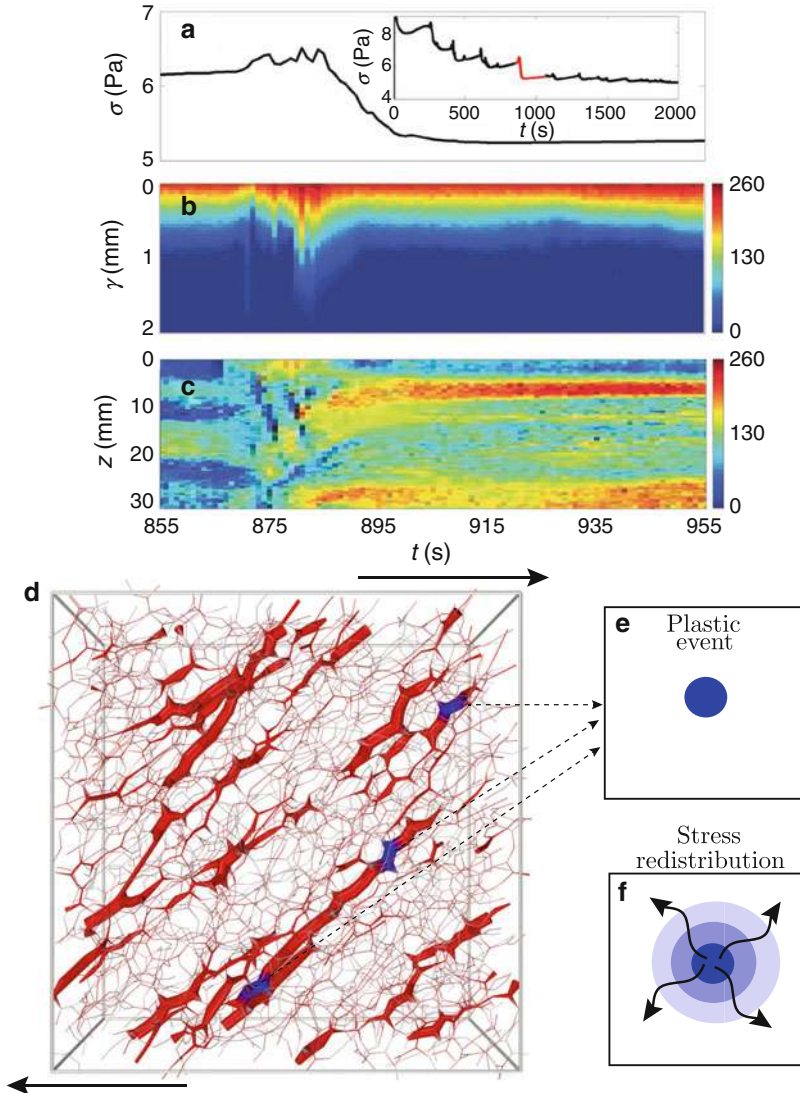
Fig. 9 Shear start-up experiment (imposed deformation) on a colloidal gel performed in a polished Taylor-Couette cell (Kurokawa et al. 2015). (**a**) The shear stress as a function of time for one stress drop (the red part of the full-time evolution shown in the inset), preceding the material fluidization. (**b**) and (**c**) are spatiotemporal intensity maps of the velocity field in the radial (**b**) and vertical (**c**) direction. (**d**) Snapshot from a numerical simulation of the model colloidal gel network discussed in Sect. 2 under shear. The thicker strands are the ones where the local tensions are larger than 60% of the maximum tensile stress. The blue indicates where breaking will happen. (Reprinted from Colombo and Del Gado 2014b). (**e**) Schematic view of the stress localization before a localized plastic event occurs inducing a nontrivial stress redistribution (**f**) across the network
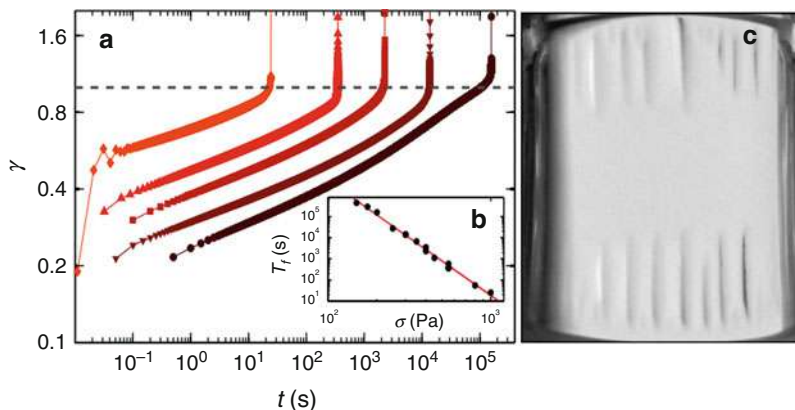
Fig. 10 Creep experiment (imposed stress σ) on a casein gels. The temporal evolution of the strain γ, for different values of shear stress σ (**a**). Three regimes are observed: an elastic regime at short time, followed by a visco-plastic regime, and finally an acceleration before rupture. (**b**) The time associated with the rupture τ_f as a function of the applied stress in a logarithmic scale, with each point corresponds to an independent creep experiment, carried out at a constant stress. The gel breaks faster when the applied stress is higher. (**c**) Snapshot of the gel before failure at $t = 0.99\tau_f$, showing regularly spaced cracks. (Reprinted from Leocmach et al. 2014)

dependent because of the coupling between the deformation rate and the dynamics of stress redistribution following localized plastic events (typically irreversible rupture of the gel connections) in a given gel network (Fig. 9d–f). Because of such coupling, the evolution at moderate or higher rate can lead to a more gradual yielding, where the dynamic correlation between the plastic events is mainly dominated by the imposed rate. Upon deforming the material at slower rate, instead, the stress redistribution dynamics in the network can itself trigger localized plastic events which are even more spatially and dynamically correlated and can enhance density fluctuations or mechanical heterogeneities in the material, eventually promoting fracture or flow instabilities (Colombo and Del Gado 2014b; Zhang et al. 2017) (Figs. 9 and 10).

In protein gels like casein, a plastic regime characterized by the clear emergence of cracks follows the initial linear viscoelastic response (Fig. 9). The resulting deformation grows as a power law over time, similar to Andrade's creep law for hard solids (Andrade 1910), and the dynamics eventually accelerates toward a catastrophic rupture (Leocmach et al. 2014, Fig. 10). The fracture time also decreases as a power law with applied stress, reminiscent of the Basquin's law of fatigue typical of hard solid as well (Basquin 1910). If the case of crystalline solids seems well understood in terms of the interaction of defects (dislocations) within the material (Csikor et al. 2007), the microscopic origin in the case of amorphous materials, and in particular for soft gels, is far from being clear, but it can be connected to the existence of localized plastic events, which act as microscopic precursors of the macroscopic failure (Aime et al. 2018b; van der Kooij et al. 2018).

5 Conclusions

Computational studies of judiciously designed models can uniquely complement new experimental approaches to provide insight into the microscopic underpinnings of the mechanical response of soft gels. Here we have given an overview and examples of recent progress in this area. The key idea is to use relatively simple but microscopic models that can include structural disorder and have complex spatiotemporal fluctuations and correlations naturally emerge from the microscopic interactions. The use of computational methods allows for solving the many-body equations of motion of interest and obtains the full spatiotemporally resolved dynamics that underlies the mechanical response of these complex materials. This comes at the cost of a few simplifications (from the very limited account of the chemical details of the specific materials to the very crude description of the solvent in which the gel matrix is embedded). We have shown that, in spite of them, one can obtain significant novel insight into the nature of the stress fluctuations acting upon aging of the material structure, into the nontrivial microstructural origin of the linear response, and into how the gel connectivity and network topology can modify dramatically the nonlinear behavior all the way to the material failure. On the one end, current challenges from the computational and theoretical side are in developing further models and approaches that can better and more efficiently capture the hydrodynamics interactions, include factors that are usually neglected but very important in experiments such as geometry and boundaries in the mechanical tests, and obtain a more quantitative understanding of the different effective interactions relevant to different material chemistry, beyond existing theories. On the other end, the results discussed here already constitute an interesting starting point to construct mesoscopic constitutive models that naturally emerge from the spatiotemporal evolution of the microstructure and its interplay with the external deformation and the deformation rate. The outstanding questions, for building such models, concern the relevant microstructural variables and the associated dynamics to be included in a continuum mechanics description of the constitutive behavior. The analysis of the spatiotemporal fluctuations, non-affine rearrangements, and local plastic events that we have briefly discussed in this chapter can help understand what are such microstructural variables and dynamics. Finally, the insight gained through studies of the type discussed here has the potential to unravel the interplay and dynamical coupling among microstructural evolution and stress redistribution in soft gels when crossing over from regime II to III in Fig. 5, providing important cues into long-standing issues such as microscopic precursors of failure, origin of creep, and hardening/toughening/self-healing mechanisms in soft materials.

References

Abete T, de Candia A, Del Gado E, Fierro A, Coniglio A (2008) Dynamical heterogeneity in a model for permanent gels: different behavior of dynamical susceptibilities. Phys Rev E 78(4):041404

Aime S, Cipelletti L, Ramos L (2018a) Power law viscoelasticity of a fractal colloidal gel. J Rheol 62:1429–1441. https://doi.org/10.1122/1.5025622

Aime S, Ramos L, Cipelletti L (2018b) Microscopic dynamics and failure precursors of a gel under mechanical load. Proc Nat Acad Sci 115(14):3587–3592

Alexander S (1998) Amorphous solids: their structure, lattice dynamics and elasticity. Phys Rep 296(2–4):65–236

Andrade EdC (1910) Proceedings of the Royal Society of London Series A, containing papers of a mathematical and physical character, 84:1–12

Angelini R, Zulian L, Fluerasu A, Madsen A, Ruocco G, Ruzicka B (2013) Dichotomic aging behaviour in a colloidal glass. Soft Matter 9(46):10955–10959

Angelini R, Zaccarelli E, de Melo Marques FA, Sztucki M, Fluerasu A, Ruocco G, Ruzicka B (2014) Glass–glass transition during aging of a colloidal clay. Nat Commun 5:4049

Arevalo RC, Kumar P, Urbach JS, Blair DL (2015) Stress heterogeneities in sheared type-I collagen networks revealed by boundary stress microscopy. PloS one 10(3):e0118021

Bandyopadhyay R, Liang D, Yardimci H, Sessoms D, Borthwick M, Mochrie S, Harden J, Leheny R (2004) Evolution of particle-scale dynamics in an aging clay suspension. Phys Rev Lett 93(22):228302

Basquin O (1910) The Exponential Law of Endurance Tests. In: Proceedings of ASTM, vol 10, pp 625–630

Bellour M, Knaebel A, Harden J, Lequeux F, Munch JP (2003) Aging processes and scale dependence in soft glassy colloidal suspensions. Phys Rev E 67(3):031405

Bianchi E, Capone B, Kahl G, Likos CN (2015) Soft-patchy nanoparticles: modeling and self-organization. Faraday Discuss 181:123–138

Blaak R, Miller MA, Hansen JP (2007) Reversible gelation and dynamical arrest of dipolar colloids. EPL (Europhys Lett) 78(2):26002

Blair GS (1944) Analytical and integrative aspects of the stress-strain-time problem. J Sci Inst 21(5):80

Blair GS, Veinoglou B (1944) A study of the firmness of soft materials based on Nutting's equation. J Sci Inst 21(9):149

Bonn D, Denn MM, Berthier L, Divoux T, Manneville S (2017) Yield stress materials in soft condensed matter. Rev Mod Phys 89(3):035005

Boromand A, Jamali S, Maia JM (2017) Structural fingerprints of yielding mechanisms in attractive colloidal gels. Soft Matter 13(2):458–473

Bouchaud JP (2008) Anomalous relaxation in complex systems: from stretched to compressed exponentials. In: Klages R, Radons G, Sokolov IM (eds) Anomalous transport: foundations and applications. Wiley-Blackwell, pp 327–345

Bouzid M, Del Gado E (2018) Network topology in soft gels: hardening and softening materials. Langmuir 34(3):773–781

Bouzid M, Colombo J, Barbosa LV, Del Gado E (2017) Elastically driven intermittent microscopic dynamics in soft solids. Nat Commun 8:15846

Bouzid M, Keshavarz B, Geri M, Divoux T, Del Gado E, McKinley GH (2018a) Computing the linear viscoelastic properties of soft gels using an optimally windowed chirp protocol. J Rheol 62(4):1037–1050. https://doi.org/10.1122/1.5018715

Bouzid M, Keshavarz B, Geri M, Divoux T, Del Gado E, McKinley GH (2018b) Computing the linear viscoelastic properties of soft gels using an optimally windowed chirp protocol. J Rheol 62:1037. https://doi.org/10.1122/1.5018715

Broedersz CP, MacKintosh FC (2014) Modeling semiflexible polymer networks. Rev Mod Phys 86(3):995

Caggioni M, Spicer P, Blair D, Lindberg S, Weitz D (2007) Rheology and microrheology of a microstructured fluid: the gellan gum case. J Rheol 51(5):851–865

Cerbino R, Trappe V (2008) Differential dynamic microscopy: probing wave vector dependent dynamics with a microscope. Phys Rev Lett 100(18):188102

Chaudhuri P, Berthier L (2017) Ultra-long-range dynamic correlations in a microscopic model for aging gels. Phys Rev E 95(6):060601

Chaudhuri P, Hurtado PI, Berthier L, Kob W (2015) Relaxation dynamics in a transient network fluid with competing gel and glass phases. J Chem Phys 142(17):174503

Chen DT, Wen Q, Janmey PA, Crocker JC, Yodh AG (2010) Rheology of soft materials. Annu Rev Condens Matter Phys 1:301

Chung B, Ramakrishnan S, Bandyopadhyay R, Liang D, Zukoski C, Harden J, Leheny R (2006) Microscopic dynamics of recovery in sheared depletion gels. Phys Rev Lett 96(22):228301

Cipelletti L, Ramos L (2005) Slow dynamics in glassy soft matter. J Phys Condens Matter 17(6):R253

Cipelletti L, Manley S, Ball R, Weitz D (2000) Universal aging features in the restructuring of fractal colloidal gels. Phys Rev Lett 84(10):2275

Colombo J, Del Gado E (2014a) Self-assembly and cooperative dynamics of a model colloidal gel network. Soft Matter 10(22):4003–4015

Colombo J, Del Gado E (2014b) Stress localization, stiffening, and yielding in a model colloidal gel. J Rheol 58(5):1089–1116

Colombo J, Widmer-Cooper A, Del Gado E (2013) Microscopic picture of cooperative processes in restructuring gel networks. Phys Rev Lett 110(19):198301

Coniglio A, De Arcangelis L, Del Gado E, Fierro A, Sator N (2004) Percolation, gelation and dynamical behaviour in colloids. J Phys Condens Matter 16(42):S4831

Coniglio A, De Arcangelis L, De Candia A, Del Gado E, Fierro A, Sator N (2006) Clusters in attractive colloids. J Phys Condens Matter 18(36):S2383

Csikor FF, Motz C, Weygand D, Zaiser M, Zapperi S (2007) Dislocation avalanches, strain bursts, and the problem of plastic forming at the micrometer scale. Science 318(5848):251–254

De Candia A, Del Gado E, Fierro A, Sator N, Tarzia M, Coniglio A (2006) Columnar and lamellar phases in attractive colloidal systems. Phys Rev E 74(1):010403

de Cagny HC, Vos BE, Vahabi M, Kurniawan NA, Doi M, Koenderink GH, MacKintosh FC, Bonn D (2016) Porosity governs normal stresses in polymer gels. Phys Rev Lett 117(21):217802

Del Gado E, Kob W (2007) Length-scale-dependent relaxation in colloidal gels. Phys Rev Lett 98(2):028303

Del Gado E, Kob W (2010) A microscopic model for colloidal gels with directional effective interactions: network induced glassy dynamics. Soft Matter 6(7):1547–1558

Del Gado E, Fierro A, de Arcangelis L, Coniglio A (2004) Slow dynamics in gelation phenomena: from chemical gels to colloidal glasses. Phys Rev E 69(5):051103

Del Gado E, Ioannidou K, Masoero E, Baronnet A, Pellenq RM, Ulm FJ, Yip S (2014) A soft matter in construction–statistical physics approach to formation and mechanics of c–s–h gels in cement. Eur Phys J Spec Top 223(11):2285–2295

Di Michele L, Fiocco D, Varrato F, Sastry S, Eiser E, Foffi G (2014) Aggregation dynamics, structure, and mechanical properties of bigels. Soft Matter 10(20):3633–3648

Divoux T, Mao B, Snabre P (2015) Syneresis and delayed detachment in agar plates. Soft Matter 11(18):3677–3685

Duduta M, Ho B, Wood VC, Limthongkul P, Brunini VE, Carter WC, Chiang YM (2011) Semi-solid lithium rechargeable flow battery. Adv Energy Mater 1(4):511–516

Eberle AP, Porcar L (2012) Flow-sans and rheo-sans applied to soft matter. Curr Opin Colloid Interface Sci 17(1):33–43

Eberle AP, Wagner NJ, Castañeda-Priego R (2011) Dynamical arrest transition in nanoparticle dispersions with short-range interactions. Phys Rev Lett 106(10):105704

Feng D, Notbohm J, Benjamin A, He S, Wang M, Ang LH, Bantawa M, Bouzid M, Del Gado E, Krishnan R et al (2018) Disease-causing mutation in α-actinin-4 promotes podocyte detachment through maladaptation to periodic stretch. Proc Nat Acad Sci 115:1517–1522

Feng J, Levine H, Mao X, Sander LM (2015) Alignment and nonlinear elasticity in biopolymer gels. Phys Rev E 91(4):042710

Feng J, Levine H, Mao X, Sander LM (2016) Nonlinear elasticity of disordered fiber networks. Soft Matter 12(5):1419–1424

Ferrero EE, Martens K, Barrat JL (2014) Relaxation in yield stress systems through elastically interacting activated events. Phys Rev Lett 113(24):248301

Fielding SM (2014) Shear banding in soft glassy materials. Rep Prog Phys 77(10):102601

Fierro A, Del Gado E, de Candia A, Coniglio A (2008) Dynamical heterogeneities in attractive colloids. J Stat Mech Theory Exp 2008(04):L04002

Fiocco D, Foffi G, Sastry S (2013) Oscillatory athermal quasistatic deformation of a model glass. Phys Rev E 88:020301. https://doi.org/10.1103/PhysRevE.88.020301

Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications, Elsevier, Amsterdam

Gallot T, Perge C, Grenard V, Fardin MA, Taberlet N, Manneville S (2013) Ultrafast ultrasonic imaging coupled to rheometry: principle and illustration. Rev Sci Inst 84(4):045107

Gao Y, Kim J, Helgeson ME (2015) Microdynamics and arrest of coarsening during spinodal decomposition in thermoreversible colloidal gels. Soft Matter 11(32):6360–6370

Godec A, Bauer M, Metzler R (2014) Collective dynamics effect transient subdiffusion of inert tracers in flexible gel networks. New J Phys 16(9):092002

Guo H, Ramakrishnan S, Harden JL, Leheny RL (2011) Gel formation and aging in weakly attractive nanocolloid suspensions at intermediate concentrations. J Chem Phys 135(15):154903

Harden J, Guo H, Ramakrishnan S, Leheny R (2012) Gel formation and aging in weakly attractive nanocolloid suspensions. In: APS meeting abstracts

Helal A, Divoux T, McKinley GH (2016) Simultaneous rheoelectric measurements of strongly conductive complex fluids. Phys Rev Appl 6(6):064004

Hsiao LC, Newman RS, Glotzer SC, Solomon MJ (2012) Role of isostaticity and load-bearing microstructure in the elasticity of yielded colloidal gels. Proc Nat Acad Sci 109(40):16029–16034

Ilg P, Del Gado E (2011) Non-linear response of dipolar colloidal gels to external fields. Soft Matter 7(1):163–171

Irving J, Kirkwood J (1950) The statistical mechanical theory of transport processes. IV. The equations of hydrodynamics. J Chem Phys 18(6):817–829

Jabbari-Farouji S, Wegdam GH, Bonn D (2007) Gels and glasses in a single system: evidence for an intricate free-energy landscape of glassy materials. Phys Rev Lett 99(6):065701

Jaishankar A, McKinley GH (2012) A fractional K-BKZ constitutive formulation for describing the nonlinear rheology of multiscale complex fluids. Power-law rheology in the bulk and at the interface: quasi-properties and fractional constitutive equations. Proc R Soc A Math Phys Eng Sci 469:20120284

Jaishankar A, McKinley GH (2014) A fractional k-bkz constitutive formulation for describing the nonlinear rheology of multiscale complex fluids. J Rheol 58(6):1751–1788

Jamali S, McKinley GH, Armstrong RC (2017) Microstructural rearrangements and their rheological implications in a model thixotropic elastoviscoplastic fluid. Phys Rev Lett 118(4):048003

Koumakis N, Petekidis G (2011) Two step yielding in attractive colloids: transition from gels to attractive glasses. Soft Matter 7(6):2456–2470

Krall A, Weitz D (1998) Internal dynamics and elasticity of fractal colloidal gels. Phys Rev Lett 80(4):778

Kurokawa A, Vidal V, Kurita K, Divoux T, Manneville S (2015) Avalanche-like fluidization of a non-Brownian particle gel. Soft Matter 11(46):9026–9037

Landau LD, Lifshitz E (1986) Theory of elasticity. Course of theoretical physics, vol 7, 3rd edn, Elsevier, Amsterdam

Landrum BJ, Russel WB, Zia RN (2016) Delayed yield in colloidal gels: creep, flow, and re-entrant solid regimes. J Rheol 60(4):783–807

Lees A, Edwards S (1972) The computer study of transport processes under extreme conditions. J Phys C Solid State Phys 5(15):1921

Leocmach M, Perge C, Divoux T, Manneville S (2014) Creep and fracture of a protein gel under stress. Phys Rev Lett 113(3):038303

Licup AJ, Münster S, Sharma A, Sheinman M, Jawerth LM, Fabry B, Weitz DA, MacKintosh FC (2015) Stress controls the mechanics of collagen networks. Proc Nat Acad Sci 112(31):9573–9578

Lidon P, Villa L, Manneville S (2017) Power-law creep and residual stresses in a carbopol gel. Rheol Acta 56(3):307–323

Lieleg O, Kayser J, Brambilla G, Cipelletti L, Bausch A (2011) Slow dynamics and internal stress relaxation in bundled cytoskeletal networks. Nat Mater 10(3):236

Lu PJ, Zaccarelli E, Ciulla F, Schofield AB, Sciortino F, Weitz DA (2008) Gelation of particles with short-range attraction. Nature 453:499

Maccarrone S, Brambilla G, Pravaz O, Duri A, Ciccotti M, Fromental JM, Pashkovski E, Lips A, Sessoms D, Trappe V et al (2010a) Ultra-long range correlations of the dynamics of jammed soft matter. Soft Matter 6(21):5514–5522

Maccarrone S, Brambilla G, Pravaz O, Duri A, Ciccotti M, Fromental JM, Pashkovski E, Lips A, Sessoms D, Trappe V, Cipelletti L (2010b) Ultra-long range correlations of the dynamics of jammed soft matter. Soft Matter 6(21):5514–5522

Macosko C (1994) Rheology. Principles, measurements, and applications. Wiley – VCH, New York

Maloney CE, Lemaître A (2006) Amorphous systems in athermal, quasistatic shear. Phys Rev E 74(1):016118

Manley S, Wyss H, Miyazaki K, Conrad J, Trappe V, Kaufman L, Reichman D, Weitz D (2005) Glasslike arrest in spinodal decomposition as a route to colloidal gelation. Phys Rev Lett 95(23):238302

Mansel BW, Williams MA (2015) Internal stress drives slow glassy dynamics and quake-like behaviour in ionotropic pectin gels. Soft Matter 11(35):7016–7023

Mao B, Divoux T, Snabre P (2016) Normal force controlled rheology applied to agar gelation. J Rheol 60(3):473–489

Mao B, Divoux T, Snabre P (2017) Impact of saccharides on the drying kinetics of agarose gels measured by in-situ interferometry. Sci Rep 7:41185

Marx D, Hutter J (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Cambridge

Meeker SP, Bonnecaze RT, Cloitre M (2004) Slip and flow in pastes of soft particles: direct observation and rheology. J Rheol 48(6):1295–1320

Ng TSK, McKinley GH (2008) Power law gels at finite strains: the nonlinear rheology of gluten gels. SOR, J Rheol 52(2):417–449

Nicolas A, Ferrero EE, Martens K, Barrat JL (2018) Deformation and flow of amorphous solids: an updated review of mesoscale elastoplastic models. Rev Mod Phys 90:45006

Ohtsuka T, Royall CP, Tanaka H (2008) Local structure and dynamics in colloidal fluids and gels. EPL (Europhys Lett) 84(4):46002

Padmanabhan P, Zia R (2018) Gravitational collapse of colloidal gels: non-equilibrium phase separation driven by osmotic pressure. Soft Matter 14(17):3265–3287. https://doi.org/10.1039/c8sm00002f

Pantina JP, Furst EM (2005) Elasticity and critical bending moment of model colloidal aggregates. Phys Rev Lett 94(13):138301

Pantina JP, Furst EM (2006) Colloidal aggregate micromechanics in the presence of divalent ions. Langmuir 22(12):5282–5288

Plimpton S (1995) Fast parallel algorithms for short–range molecular dynamics. J Comput Phys 117:1–19

Puosi F, Rottler J, Barrat JL (2014) Time-dependent elastic response to a local shear transformation in amorphous solids. Phys Rev E 89(4):042302

Ramos L, Cipelletti L (2001) Ultraslow dynamics and stress relaxation in the aging of a soft glassy system. Phys Rev Lett 87(24):245503

Rovigatti L, Sciortino F (2011) Self and collective correlation functions in a gel of tetrahedral patchy particles. Mol Phys 109(23–24):2889–2896

Royall CP, Eggers J, Furukawa A, Tanaka H (2015) Probing colloidal gels at multiple length scales: the role of hydrodynamics. Phys Rev Lett 114:258302. https://doi.org/10.1103/PhysRevLett.114.258302

Ruta B, Chushkin Y, Monaco G, Cipelletti L, Pineda E, Bruna P, Giordano VM, Gonzalez-Silveira M (2012) Atomic-scale relaxation dynamics and aging in a metallic glass probed by x-ray photon correlation spectroscopy. Phys Rev Lett 109:165701

Ruta B, Czakkel O, Chushkin Y, Pignon F, Nervo R, Zontone F, Rinaudo M (2014) Silica nanoparticles as tracers of the gelation dynamics of a natural biopolymer physical gel. Soft Matter 10(25):4547–4554

Salerno KM, Maloney CE, Robbins MO (2012) Avalanches in strained amorphous solids: does inertia destroy critical behavior? Phys Rev Lett 109(10):105703

Segre P, Prasad V, Schofield A, Weitz D (2001) Glasslike kinetic arrest at the colloidal-gelation transition. Phys Rev Lett 86(26):6042

Tabatabai AP, Kaplan DL, Blair DL (2015) Rheology of reconstituted silk fibroin protein gels: the epitome of extreme mechanics. Soft Matter 11(4):756–761

Tanguy A, Wittmer J, Leonforte F, Barrat JL (2002) Continuum limit of amorphous elastic bodies: a finite-size study of low-frequency harmonic vibrations. Phys Rev B 66(17):174205

Thompson A, Plimpton S, Mattson W (2009) General formulation of pressure and stress tensor for arbitrary many-body interaction potentials under periodic boundary conditions. J Chem Phys 131:154107

Trappe V, Prasad V, Cipelletti L, Segre P, Weitz D (2001) Jamming phase diagram for attractive particles. Nature 411(6839):772

Tsurusawa H, Leocmach M, Russo J, Tanaka H (2018) Gelation as condensation frustrated by hydrodynamics and mechanical isostaticity. arXiv preprint arXiv:180404370

van Doorn JM, Verweij JE, Sprakel J, van der Gucht J (2018) Strand plasticity governs fatigue in colloidal gels. Phys Rev Lett 120(20). https://doi.org/10.1103/PhysRevLett.120.208005

Van Oosten AS, Vahabi M, Licup AJ, Sharma A, Galie PA, MacKintosh FC, Janmey PA (2016) Uncoupling shear and uniaxial elastic moduli of semiflexible biopolymer networks: compression-softening and stretch-stiffening. Sci Rep 6:19270

van der Kooij HM, Dussi S, van de Kerkhof GT, Frijns RAM, van der Gucht J, Sprakel J (2018) Laser speckle strain imaging reveals the origin of delayed fracture in a soft solid. Sci Adv 4(5). https://doi.org/10.1126/sciadv.aar1926, http://advances.sciencemag.org/content/4/5/eaar1926.full.pdf

Varga Z, Swan JW (2018) Normal modes of weak colloidal gels. Phys Rev E 97(1):012608

Varga Z, Wang G, Swan J (2015) The hydrodynamics of colloidal gelation. Soft Matter 11(46):9009–9019

Varrato F, Di Michele L, Belushkin M, Dorsaz N, Nathan SH, Eiser E, Foffi G (2012) Arrested demixing opens route to bigels. Proc Nat Acad Sci 109(47):19155–19160

Vasisht VV, Roberts G, Del Gado E (2019) Shear start-up in jammed soft solids: a computational study. arXiv:1908.03943. https://arxiv.org/abs/1908.03943

Youssry M, Madec L, Soudan P, Cerbelaud M, Guyomard D, Lestriez B (2013) Non-aqueous carbon black suspensions for lithium-based redox flow batteries: rheology and simultaneous rheo-electrical behavior. Phys Chem Chem Phys 15(34):14476–14486

Zaccarelli E, Saika-Voivod I, Buldyrev SV, Moreno AJ, Tartaglia P, Sciortino F (2006) Gel to glass transition in simulation of a valence-limited colloidal system. J Chem Phys 124(12):124908

Zhang L, Rocklin DZ, Sander LM, Mao X (2017) Fiber networks below the isostatic point: fracture without stress concentration. Phys Rev Mater 1(5):052602

Zia RN, Landrum BJ, Russel WB (2014) A micro-mechanical study of coarsening and rheology of colloidal gels: cage building, cage hopping, and Smoluchowski's ratchet. J Rheol 58(5):1121–1157

Mesoscale Structure and Mechanics of C-S-H

71

Katerina Ioannidou

Contents

Abstract

Cement is the most used building material on earth, yet its properties are inexactly understood and not fully controlled. Cement hydrates named C-S-H (Calcium-Silicate-Hydrate) are the most abundant phase in hydrated cement paste and are responsible for gluing all other hydration products and unreacted cement together. A source of complexity in modelling C-S-H is that the structure is amorphous, multiscale with fully interconnected porosity spanning from a few nm up to mm. The focus of this chapter is modeling approaches that allow connecting structure and mechanics of C-S-H at the mesoscale (the scale that spans from few nm up to the micron) from the early stages of hydration, the setting and up to the hardened cement paste. The modelling approach reviewed here is of reduced complexity based on coarse-graining with emphasis on the

K. Ioannidou (✉)

Department of Civil and Environmental Engineering, MIT Energy Initiative, Multiscale Materials Science for Energy and Environment, Joint MIT-CNRS laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: hekate@mit.edu

effective interactions between C-S-H particles. It addresses the mesoscale of C-S-H and has provided a unified framework for understanding the microstructure of C-S-H and reconciling data from many different experimental techniques. A consistent picture is presented covering (1) the reactive solidification of cement, (2) the origin of the observed microstructure of C-S-H, and (3) its link to mechanics. This provides a powerful predictive tool for nanoscale design of cement.

1   Introduction

Cement paste is a multiscale porous material that exhibits distinct physical features at different length scales. The most abundant phase in hardened cement and responsible for its remarkable mechanical strength is the C-S-H (Calcium-Silicate-Hydrate). Initially C-S-H forms a soft gel that eventually hardens to glue all phases together. It is a rather complex phase (1) at atomistic level as it has variant stoichiometry, not well-defined molecular size or shape and (2) at meso- and macroscopic levels as it exhibits complicated topology of pores. A source of complexity in cement is that the structure is out-of-equilibrium, amorphous, multiscale with interconnected porosity. This complexity has major implications on the mechanical properties of cement and hence on the stability and durability of buildings.

The initial amount of water influences significantly the overall porosity of the material, the higher the w/c ratio the more porous the material. Initially the voids in cement are big and connected; however, as the hydration proceeds and the solid structure percolates, the pores reduce in size. Initially big capillary pores of μm are filling up with aggregating precipitates that contain a large amount of much smaller pores (1–100 nm). Proton nuclear magnetic resonance (NMR) spectroscopy has given interesting insight into the pore size distribution (Levitz et al. 2003; Korb et al. 2007; Chemmi et al. 2010). More specifically, experiments indicate the presence of pores of $2-5$ nm and $14-30$ nm, both associated with the C-S-H phase.

In the beginning of the previous century, Le Chatelier proposed that the anhydrous cement compound was dissolved in water and that the hydrated products precipitated from this solution in the form of interlocking crystals (Le Chatelier 1887). In the same period, Michaëlis attributed the cementing action to the formation and desiccation of a gel (Michaelis 1893). In the 1950s based on X-ray diffraction evidence, Bernal suggested that C-S-H has a layered structure related to a crystalline mineral, 1.1-nm tobermorite, and called it tobermorite gel (Bernal 1954). Today all these observations have been proved partially correct. Both small angle neutron scattering and statistical nanoindentation technique indicate that the C-S-H meso-texture is comprised of nanoscale colloidal particles that precipitate from the pore solution and aggregate in interlocking units of few nm, initially forming a gel and eventually an amorphous solid (Allen et al. 2004; Vandamme and Ulm 2009).

The structure of C-S-H does not show much order neither at the atomistic scale nor at the mesoscale. In the atomistic scale, tobermorite – a crystalline

natural mineral – might form locally during cement hydration. However, most of C-S-H realizations have little order –expressed as calcium-silicate layers – and much disorder in the interlayers filled with calcium and water molecules (Pellenq et al. 2009). In certain chemical conditions, calcium-silicate layers may grow for several tenths or even hundreds of nm giving the impression under the microscope of a continuum uniform C-S-H phase organized in needles or sheets. In most cases, nevertheless, C-S-H appears chemically and texturally heterogeneous, showing amorphous aggregates interrupted by voids. This dyadic nature of C-S-H between particle and continuum has been the reason for debate over many years in the cement community.

Powers and Brownyard described the nano-colloid of C-S-H as a solid particle with inner crystalline structure and peculiar shape that is the basic building blocks of the C-S-H gel. This was based on evidence from water sorption isotherms and total and nonevaporable water (Powers 1958). Small angle neutron scattering (SANS) experiments of Allen provided the main indication of the colloidal nature of C-S-H (Allen et al. 1987). The colloidal particles in the cement paste have a characteristic length of ∼5 nm after the induction time and a correlation length ∼40 nm developing due to aggregation. Based on this, Thomas and Jennings introduced the idea that C-S-H is composed by particles aggregating with high and low packing density (Thomas and Jennings 2006). They also expanded their model to the formation of gel with pores of 1–2 nm, the so-called gel pores. Aggregation of C-S-H particles was also confirmed by atomic force microscopy imaging of a hydrating surface of alite and recent SANS experiments confirming the lamellar structure and determining the thickness of calcium silicate layers and of the water (Lesko et al. 2001; Chiang et al. 2012). From a modeling perspective, the introduction of a coarse-graining length, the C-S-H particle, facilitates the simulation of the material at larger length scales and over longer timescales.

In the following sections, a brief overview on cement hydration is given, focusing on the coupling of cement hydration to the mechanics of the C-S-H gel, and the mesoscale modelling approach is presented. The validation of the model with experiments is provided through the kinetics of growth, the mesoscale texture, and mechanics of C-S-H. This modelling approach for reactive solidification was thoroughly validated for hardened cement paste. Its main strength being the linking of cement's chemistry, via the effective interactions, to the growth of the microstructure and mechanics. Such an approach could be applied for other reactive solidification processes and also to address durability issues of cement due to in-pore solidification.

2   Cement Hydration: From Liquid to Stone

The transformation of fresh cement paste into soft and finally hard solid is intimately connected to the formation of C-S-H. During the process of hydration, numerous chemical reactions take place and different hydration products are formed. The main hydration processes are the dissolution of the reactive species contained

Fig. 1 Schematic representation of the time dependence of (**a**) the heat flow during hydration, typically obtained in isothermal calorimetry measurements and (**b**) the shear stiffness of cement paste during hydration and setting. The vertical red dashed line show the time that setting occurs
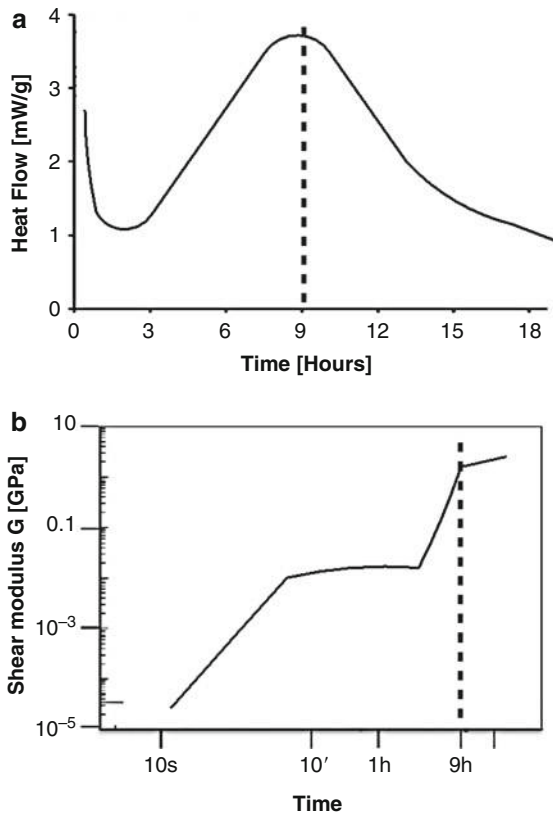
in the clinker (tri-calcium and di-calcium silicates – indicated as C3S and C2S, respectively) and precipitation of C-S-H with other minor hydration products. To fully react with water, cement needs to be mixed at least in a proportion of 42% (water to cement ratio – w/c). The full hydration process is completed in 28 days. However, the material has already solidified by the end of the acceleration period – the rheological setting – after which it reacts at very slow pace. A typical technique to follow the hydration reaction is by measuring the heat flow in isothermal calorimetry (Fig. 1a). From the changes of the heat curve, the hydration is usually divided into (1) an initial (∼30–60 mins), (2) an accelerating (∼8 h), and (3) a decelerating period of heat flow (Taylor 1997).

The growth of C-S-H changes significantly the rheology and the mechanics of the cement paste. Controlling the rheology of cement is crucial for several applications such as casting and transportation of cement. During setting, cement is a viscoelastic material, exhibiting characteristics of both solids and viscous fluids. In particular the way the mechanical strength develops and the crossing from the fluid-like properties to solid-like behavior are intimately connected to the hydration process. The complex time-dependent rheological response of cement

during the hydration is usually monitored by dynamic mode rheometry where the shear modulus versus the hydration time is obtained (Lootens et al. 2004). Three different regimes are observed in the measurements of the shear modulus of cement, as shown in Fig. 1b, the first increase of stiffness, a dormant period where the change of the elastic modulus is much weaker and the second increase of stiffness.

Almost immediately after cement is mixed with water, the shear modulus increases to 10^7Pa, which corresponds to the first increase of stiffness of Fig. 1b, and it is a clear sign of an initial gelation. At this point there are no reaction products, but the cement paste has a measurable yield stress. This first increase occurs when the clinker particles stick together and form a weak network, due to high ionic strength of the interstitial solution. At this stage, the gel structure is relatively weak and is easily destroyed by vibration, although, if not disturbed, it does not change during the next hours (Nachbaur et al. 2001). After this first weak gelation, the shear modulus stays at the plateau of 10^7Pa for nearly 5 h, which correspond to the plateau of Fig. 1b. The second increase of stiffness signals the setting of cement and its transformation into a solid material. The second increase of the shear modulus happens at the same time as the acceleration period of the hydration (see red vertical dashed line in Fig. 1a and b), when the initial network of clinker particles gets reinforced with C-S-H. The densification of the C-S-H around the clinker particles continues and, by the end of the second increase, when the acceleration period also ends, the shear modulus has reached the range of GPa due to the mechanical percolation of the C-S-H gel among what is left unhydrated of the clinker particles. Finally, further hardening occurs over periods from weeks to years due to further hydration and long-term rearrangements of the structure over different length scales.

3 Mesoscale Modeling for Reactive Solidification

The aim of mesoscale modeling is to capture the solidification process and reproduce a realistic porous texture that can be validated by different experimental observations of structural and mechanical characteristics of cement. This approach takes into account the out-of-equilibrium reactive nature of the material. By "mesoscale" one indicates lengthscales neither small enough to use atomistic approaches nor large enough to employ mean-field homogenization theories to investigate the relevant physics. Modeling techniques based on coarse-grained statistical physics provided a unified framework for understanding the structure and mechanics of a material as complex as C-S-H and can even suggest nano-scale design for enhanced and multifunctional cement formulations. Coarse-grained models allow investigating large lengthscales and longer timescales by reducing the degrees of freedom of the system and maintaining only the minimal essential constituents (Doi 2013).

3.1 The Colloidal Approach

C-S-H is known to precipitate as nanoscale particles in the pore solution upon mixing cement powder with water. During cement hydration, new C-S-H units are produced, due to the ongoing reactions; they interact with the existing ones, stick together, and form aggregates that eventually create a stress-bearing network, a gel that acts as a spontaneously densifying glue. The structural and mechanical properties of C-S-H gels are the results of a complex interplay between the precipitation kinetics, controlled by the evolution of the chemical composition and the dissolution of the clinker, and the effective interactions between the hydrates.

This approach uses coarser units than atomistic simulations in order to obtain the mesoscale structure of C-S-H that can span up to the micron scale. The building blocks are the C-S-H hydrates represented by colloidal particles of few nm. We consider an open system of precipitating and interacting colloidal particles within a small volume of a capillary pore of the fresh cement paste. The ionic solution of the capillary pore is supersaturated such that C-S-H hydrates can precipitate.

The statistical ensemble that captures this process is the grand canonical ensemble of constant $\mu V T$ condition, where μ is the chemical potential, V is the volume, and T the temperature. The simulations consist of a coupled Grand Canonical Monte Carlo (GCMC) and Molecular Dynamics (MD) scheme that accounts for particle insertion and the dynamics of the formed structures.

A particle can be inserted or deleted with a probability given by a Metropolis Monte Carlo (MC) scheme (Frenkel and Smit 2001)

$$P = \min\left[1, e^{\pm\beta(\mu_{exc}-\Delta U)}\right], \tag{1}$$

where k_b is the Boltzmann constant, T is the temperature, μ_{exc} is the excess chemical potential, and ΔU is the difference in the total potential energy of the particle configurations due to the insertion or removal of a particle. The chemical potential provides an effective description of the energy gain to create a new hydrate and thus accounts for the chemical environment of the pore solution. The probability of precipitation is obtained as the difference, at a fixed temperature T, between the change in the internal energy (due to inter-particle interactions, ΔU) and the free energy gain corresponding to the formation of a C-S-H hydrate (expressed through μ_{exc}). The chemical potential sets the equilibrium particle density that can be reached in the pore. Assuming that during cement hydration and C-S-H precipitation, the pore solution is always supersaturated, the chemical potential is chosen such that favors insertion and drives the system towards progressive densification.

To follow the microscopic dynamics of the microstructure growth, a hybrid scheme couples the GCMC insertion/deletion events to MD simulation (Ioannidou et al. 2014). The number of GCMC attempts over the number of MD steps defines the kinetic rate $R = N_{GCMC}/N_{MD}$, a parameter that characterizes the kinetics of precipitation in the experimental system. Structure and mechanics depends on how

fast precipitation takes place with respect to the timescale associated to structural relaxation of the C-S-H gel or amorphous solid that is progressively formed. The MD part of the simulations allows following the aggregation process and the restructuring as the precipitation proceeds.

Overall, the number of precipitated particles in the simulation box is the results of the effective interactions between particles, the chemical potential, and the imposed precipitation rate. This coarse-grained approach allows rationalizing the role of the rate (and therefore of the parameters that control it) in the development of the physical properties of the system and its interplay with the emerging cooperative processes due to the densification of the microstructure.

3.2   Effective Interactions

The choice of the effective interactions between precipitating particles is very important as it determines the local packing of particles and the mesoscale organization. For complex systems such as cement or clays, the potential of mean force (PMF) computed from atomistic models of two interacting C-S-H particles or clay platelets offers a route to upscale the chemistry of the particles surface and of the solution into a two-body interaction (Ioannidou et al. 2017a; Masoumi et al. 2017; Ebrahimi et al. 2014). The PMF is a free energy function of two interacting objects for specific thermodynamic conditions, i.e., pressure, temperature, ionic concentration, etc. Although an expensive computationally approach, PMF is useful for chemically complex materials as cement. For charged colloids, the electric double layer theory of Gouy and Chapman (GC), based on the solution of the Poisson–Bolzmann (PB) equation, describes the formation of the electric double layer when a homogenously charged surface is in contact with an electrolyte solution. The high surface charge of C-S-H particles and the divalent calcium ions in the pore solution make mean-field approaches such as PB nonapplicable (Pellenq and Van Damme 2004).

A combination of evidences points to the fact that the effective interactions between C-S-H nano-colloids result from a competition between a short-range attraction with a longer-range repulsion expressed as an attractive well and a shoulder (Ioannidou et al. 2014). Two similarly charged particles may generate attractive interaction because of correlations in the local concentration of ions due to thermal fluctuations. This has been known for more than 20 years (Kjellander et al. 1988; Valleau et al. 1991), and the suggestion that this may contribute to the cohesion of set cement was made more than two decades ago (Pellenq et al. 1997). Sufficiently strong electrostatic coupling in the electrolyte is necessary for attraction. The electrostatic coupling is evaluated from the colloidal surface charge density, the ionic charge and radius, the solvent dielectric constant (in an implicit solvent description), and the temperature. The PMF approach takes into account all these parameters and was able to predict the short-range attraction followed by a repulsive shoulder (Bonnaud et al. 2016). This was also confirmed by direct force measurements performed using a C-S-H covered atomic force microscopy (AFM)

tip and a C-S-H flat surface (Plassard et al. 2005). Investigation of different solution parameters showed that the calcium ion concentration controls the attraction to repulsion ratio. Recent measurement of hydrating cement volume changes showed that cement paste initially expands and later shrinks (Abuhaikal et al. 2018). They also suggest that the reason for this is the competition of attractive and repulsive interactions.

The effective interactions in cement are not homogeneous in space and time, but they depend strongly on the ionic solution surrounding the C-S-H particles. The AFM measurements suggest that in the beginning of hydration when calcium ionic concentration is low the repulsion is stronger, whereas eventually the attraction is taking over. In the mesoscale approach just introduced (Ioannidou et al. 2014), the effective interaction of two spherical C-S-H particles with diameter σ is defined as a combination of a short-range generalized Lennard-Jones (LJ) attraction and a long-range Yukawa repulsion (Israelachvili 1992):

$$
V(r) = 4 \epsilon \left[\left(\frac{\sigma}{r} \right)^{2\gamma} - \left(\frac{\sigma}{r} \right)^{\gamma} \right] + A \frac{e^{-\kappa r}}{r}, \tag{2}
$$

where r is the inter-particle distance, γ the exponent in the generalized LJ interaction, and κ the inverse screening length. The parameters ε and A quantify the relative strength of the attractive and repulsive interaction terms. The functions of Eq. 2 with all possible parameter values define a family of interaction potentials with an attractive minimum at $r = r_{min}$ and a repulsive barrier at $r = r_{max}$. A subfamily of potentials was considered with fixed $\gamma = 12$ and $\kappa^{-1} = 0.5\sigma$ and combinations of A and ε values such that the depth of the attraction well at r_{min} is fixed to $V(r_{min}) = -1k_BT$ and the height of the repulsive shoulder is between $0 < V(r_{max}) \leq 0.5\ k_BT$ (Ioannidou et al. 2016a).

The exact parameters depend on the time of hydration and the specific thermodynamic condition. In the next sections, we discuss results for three cases: the high shoulder HS ($A = 4$, $\varepsilon = 1.5$) representative of early stages hydration, the low shoulder LS ($A = 12$, $\varepsilon = 2.4$) representative of later stages hydration, and no Yukawa term LJ ($A = 0$, $\varepsilon = 1.5$) representative of the hardened cement paste. The three interaction potentials are displayed in Fig. 2. In all cases under consideration the attraction is short-ranged, i.e., the width of the attraction well is in the range between 0.1σ and 0.3σ. Monodisperse (Ioannidou et al. 2016a) and polydisperse (Ioannidou et al. 2016b, 2017b; Masoero et al. 2012) size systems were consider. In this approach, polydispersity is a way to introduce further heterogeneity in the system.

In the next section, we show that this type of approach made it possible to connect the time evolution of C-S-H growth and densification to textural and mechanical properties of hardened cement paste. In addition, although the hybrid GCMC-MD precipitation scheme was conceived to explore the nonequilibrium conditions, it helped to reveal the crucial influence of the underlying thermodynamically stable and metastable states (dictated by the effective interactions) on the out-of-equilibrium process of precipitation (Ioannidou et al. 2014, 2016a).

Fig. 2 Effective interaction of C-S-H particle in different thermodynamic conditions. A short-range 12–24 Lennard-Jones attraction is combined with a long-range Yukawa repulsion. The blue dashed-dotted line corresponds to HS, the red dashed line to LS and the black solid line to 12–24 LJ. The attraction to repulsion ratio ε/A depends upon the concentration of calcium cations
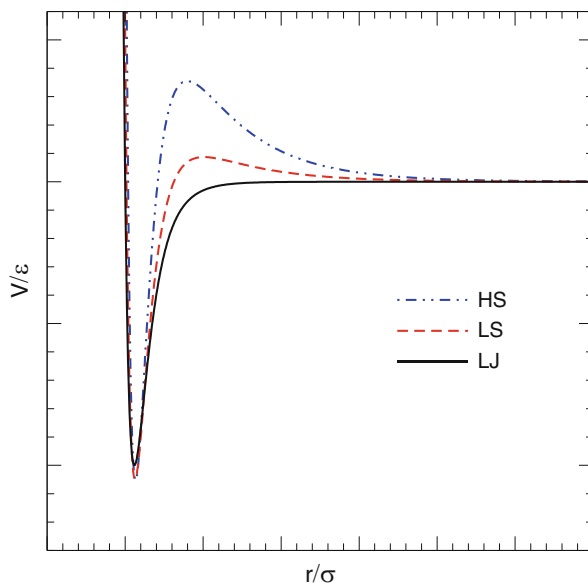
4   The Kinetics and Rheology of C-S-H Growth

The colloidal mesoscale approach for C-S-H precipitation captured for the first time consistently the kinetics of C-S-H growth and densification and at the same time provided a realistic time evolution of the microstructure (Ioannidou et al. 2014, 2016b). The simulations provide the volume fraction of the box occupied by C-S-H hydrates, and by analyzing the coordinates of the particles, correlation in the microstructure can be computed. Starting from C-S-H particles of size 3–10 nm, the mesoscale structures assembled by the precipitation simulation technique can reach up to 600 nm.

In more detail, the results display hydrates precipitation curves (time evolution of the volume fraction) that indeed reproduce the acceleration and deceleration regime typically observed in experiments and were able to correctly capture the effect of Ca^{+2} concentration on the hydration kinetics and the gel morphology (Ioannidou et al. 2014). The structural analysis of the C-S-H gels formed under precipitation showed that the different densification regimes are directly related to the aggregation process that is controlled by the underlying thermodynamics (Ioannidou et al. 2016a). As the volume fraction increases the local packing of the aggregates changes facilitating further precipitation, due to the change in the effective interactions. When the system is dilute, large elongated clusters quickly form and merge into a gel, whereas when the system becomes denser compact and locally denser domains are prevalent (Ioannidou et al. 2014, 2016a).

Figure 3 shows how the different effective interaction influences the formation of C-S-H gel at low volume fractions, as it appears in simulation snapshots. HS poten-
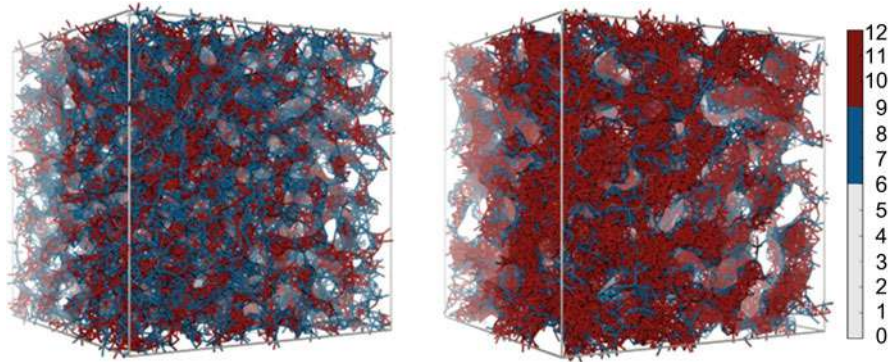
Fig. 3 Snapshots of C-S-H gels, with HS (left) and LS (right) effective interaction at volume fraction $\eta = 0.25$ produced by particle precipitation in the μVT ensemble. Only the bonds between the particles are depicted using lines. The color code corresponds to the number of nearest neighbors

tial, with a higher repulsive shoulder than LS, favors the formation of elongated thin strands, whereas LS enhances the close packing of particles. This has an effect on the pore size distributions since HS produced many smaller pores and LS instead favors large pores. Furthermore, tuning the out-of-equilibrium conditions via the kinetic rate R changes the degree of local order in the system (Ioannidou et al. 2014).

The mesoscale colloidal approach for reactive solidification allows to connect the changes of the physico-chemical environment, which dictate the evolution of the effective interactions, to the sizes of pores. Moreover, in the case of cement the fact that the effective interactions evolve from HS to LS and to LJ eventually specifically favors the early gel formation and its continuous densification. Such an approach helps rationalizing how cement attains its unique strength and contributes in the science-based design of the properties of cement and related materials.

A comparison with rheological measurement can also be achieved by evaluating the evolution of the shear modulus during the precipitation simulations of C-S-H particles. Shear tests were performed on configurations of different volume fractions and, using the resulting load curves, the evolution of elastic modulus during cement hydration can be determined as shown in Fig. 4. Hence, the mesoscale colloidal approach allowed us to build the connection among gelation kinetics, changes in the morphology of the microstructure, and rheological and mechanical properties of the material.

4.1 Mesoscale Texture and Mechanics of C-S-H in Hardened Cement Paste

A full validation of the C-S-H amorphous solid structures obtained from the colloidal approach was possible against a large set of available experimental
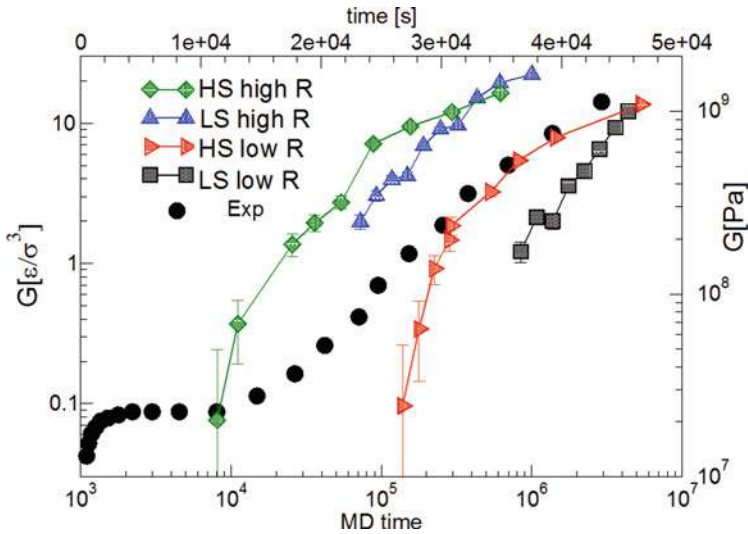
Fig. 4 Shear modulus as a function of the time for HS and LS at a high and low kinetic rate R. The simulation timescale is set up in agreement to Reference (Ioannidou et al. 2014). The experimental data, depicted by black circles, are from reference (Lootens et al. 2004) plotted in linear time

results ranging from electron-microscopy imaging, nano-indentation tests, small angle X-rays (SAXS) and SANS, NMR spectroscopy, and adsorption/desorption of nitrogen and water (Ioannidou et al. 2016b). The complex pore network and the structural heterogeneities inherent to cement paste naturally emerge from the effective cohesive interactions and from the nonequilibrium conditions of the precipitation simulation.

The voids and solids of the configurations obtained via mesoscale simulations were analyzed via pore size distribution and local volume fractions. This approach was able to capture the broad distribution of pore sizes up to two orders of magnitude. A large population of nanoscale gel pores (1–2 nm) and a broad distribution of larger mesopores (3–40 nm) form an extended pore network. Together with the continuous distribution of local solid packing fractions, this observation composes a new physical picture of cement (Ioannidou et al. 2016b).

In more details, the meso-scale texture obtained is consistent with adsorption/desorption experiments, NMR, and with Powers-Brownyard classical relations estimating capillary and gel pore volumes. The SANS signal in cement paste has an intriguing q^{-3} behavior observed over an unusually large range of distances. This behavior was obtained also from calculations of the scattering intensity of the mesoscale configurations. By further analysis of the structure using chord length distributions, the explanation for the q^{-3} dependence was found in the interface of pore and solid matrix. It is due to a unique coexistence of a surface roughness of the capillary pores at the scale of 3–40 nm with a complex (fractal) pore network, all consequences of the interplay between the cohesive nano-scale effective interactions
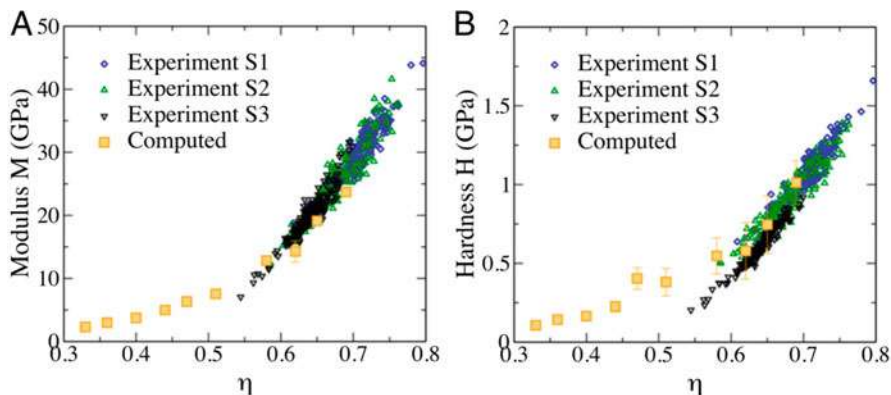
Fig. 5 (**a**) The C-S-H modulus M and (**b**) the hardness H as a function of η for simulations and experiments from reference Ioannidou et al. (2016b))

and the out-of-equilibrium conditions in which the material forms during cement hydration (Ioannidou et al. 2016b). Due to the mesoscale approach for the first time, simulations data were able to explain the extended spatial correlations of cement.

Moreover, calculating the specific surface area using in-pore chord length distribution has allowed us to formulate a new perspective on water and nitrogen adsorption experiments and to reconcile literature data apparently contradictory (Ioannidou et al. 2016b). In fact, the calculations are consistent with the fact that drying conditions may determine which part of the pore population is actually accessible to the adsorbate and the existence of a thick, stable layer of water molecules in C-S-H may block, in some of the experiments, the access to nanoscale gel pores.

Mechanical properties of the C-S-H mesoscale configurations can also be computed. Figure 5 shows the modulus M and hardness H as a function of the volume fraction η in experiments on hardened cement paste and in simulations, showing an excellent agreement. In these experiments, portlandite and all other minor phases of cement have been carefully excluded and the data refer only to the C-S-H phase (Krakowiak et al. 2015). The C-S-H data were isolated from the cement samples by spatially correlating nano-indentation data and energy dispersive X-ray spectrometry and elemental mapping data. The comparison of these data highlights the fact the densest part ($\eta > 0.6$) of the material spans the whole structure in a complex, load-bearing network, which dominates the nano-indentation experiments (Ioannidou et al. 2016b, 2017b).

5   Conclusions

In this chapter, we provided an effective approach to reactive solidification applied to the problem of the formation and evolution of the C-S-H microstructure, the main binding phase of cement, at the mesoscale level (1–500 nm). A source of complexity

in cement, and more specifically in C-S-H, is that the structure is amorphous and develops over a range of length scales under out-of-equilibrium conditions. This issue has major implications in the mechanical properties of cement and hence in the stability and durability of buildings. In order to tackle this complexity, a minimalistic coarse-grained nanoscale model and computation approach for the C-S-H gel formation during the early stages of cement hydration was discussed. The use of a coarse-grained model makes this investigation possible by reducing the degrees of freedom of the system and maintaining only the minimal essential constituents. The main characteristics of the colloidal mesoscale approach are: the material consists of particles of few nanometer representing the main units of the C-S-H gel, the particles interact with effective forces arising from atomistic interactions, the formation and evolution of the gel microstructure is determined by the combination of statistical mechanics laws with the out-of-equilibrium conditions stemming from the complex chemical hydration reactions. This type of model can provide the time evolution of the C-S-H gel microstructure which can be analyzed and used for computing directly its structural and mechanical properties in excellent agreement with experimental measurements.

References

Abuhaikal M, Ioannidou K, Petersen T, Pellenq RJ-M, Ulm FJ (2018) Le Châtelier's conjecture: measurement of colloidal Eigenstresses in chemically reactive materials. J Mech Phys Solids 112:334–344

Allen AJ, Oberthur RC, Pearson D, Schofield P, Wilding CR (1987) Development of the fine porosity and gel structure of hydrating cement systems. Phil Mag B 56:263–268

Allen AJ, McLaughlin JC, Neumann DA, Livingston RA (2004) In situ quasi-elastic scattering characterization of particle size effects on the hydration of tricalcium silicate. J Mater Res 19(11):3242–3254

Bernal JD (1954) The structures of cement hydration compounds. In: Proceedings of the 3rd international symposium on the chemistry of cements, pp 216–236

Bonnaud PA, Labbez C, Miura R, Suzuki A, Miyamoto N, Hatakeyama N, Miyamoto A, Van Vliet KJ (2016) Interaction grand potential between calcium–silicate–hydrate nanoparticles at the molecular level. Nanoscale 8(7):4160–4172

Chemmi H, Petit D, Levitz P, Korb JP (2010) NMR control of aging and durability of hardened cement pastes. Comptes Rendus Chim 13(4):405–408

Chiang W, Fratini E, Baglioni P, Liu D, Chen S (2012) Microstructure determination of calcium-silicate-hydrate globules by small-angle neutron scattering. J Phys Chem C 116(8):5055–5061

Doi M (2013) Soft matter physics. Oxford University Press, Oxford

Ebrahimi D, Whittle AJ, Pellenq RJ-M (2014) Mesoscale properties of clay aggregates from potential of mean force representation of interactions between nanoplatelets. J Chem Phys 140(15):154309

Frenkel D, Smit B (2001) Understanding molecular simulation: from algorithms to applications. Elsevier, Amsterdam

Ioannidou K, Pellenq RJ-M, Del Gado E (2014) Controlling local packing and growth in calcium-silicate-hydrate gels. Soft Matter 10:1121–1133

Ioannidou K, Kanduc M, Li L, Frenkel D, Dobnikar J, Del Gado E (2016a) The crucial effect of early-stage gelation on the mechanical properties of cement hydrates. Nat Commun 7:12106

Ioannidou K, Krakowiak KJ, Bauchy M, Hoover CG, Masoero E, Yip S, Ulm F-J, Levitz P, Pellenq RJ-M, Del Gado E (2016b) Mesoscale texture of cement hydrates. Proc Natl Acad Sci 113(8):2029–2034

Ioannidou K, Carrier B, Vandamme M, Pellenq R (2017a) The potential of mean force concept for bridging (length and time) scales in the modeling of complex porous materials. EPJ Web Conf 140:1009

Ioannidou K, Del Gado E, Ulm F-J, Pellenq RJ-M (2017b) Inhomogeneity in cement hydrates: linking local packing to local pressure. J Nanomech Micromech 7(2):4017003

Israelachvili JN (1992) Intermolecular and surface forces, second edition: with applications to colloidal and biological systems (colloid science). Academic, Boston

Kjellander R, Marcelja S, Pashley RM, Quirk JP (1988) Double-layer ion correlation forces restrict calcium-clay swelling. J Phys Chem 92(23):6489–6492

Korb J-P, Monteilhet L, McDonald PJ, Mitchell J (2007) Microstructure and texture of hydrated cement-based materials: a proton field cycling Relaxometry approach. Cem Concr Res 37(3):295–302

Krakowiak KJ, Wilson W, James S, Musso S, Ulm F-J (2015) Inference of the phase-to-mechanical property link via coupled X-ray spectrometry and indentation analysis: application to cement-based materials. Cem Concr Res 67:271–285

Le Chatelier H (1887) Recherches expérimentales sur la constitution des mortiers hydrauliques. Ann. des Mines, Huitième S, pp 345–465

Lesko S, Lesniewska E, Nonat A, Mutin JC, Goudonnet JP (2001) Investigation by atomic force microscopy of forces at the origin of cement cohesion. Ultramicroscopy 86(1–2):11–21

Levitz P, Korb JP, Petit D (2003) Slow dynamics of embedded fluid in mesoscopic confining systems as probed by NMR relaxometry. Eur Phys J E 12(1):29–33

Lootens D, Hébraud P, Lécolier E, Van Damme H (2004) Gelation, shear-thinning and shear-thickening in cement slurries. Sci Technol 59(1):31–40

Masoero E, Del Gado E, Pellenq RJ-M, Ulm F, Yip S (2012) Nano-structure and -mechanics of cement : polydisperse colloidal packing. Phys Rev Lett 109(15):3–6

Masoumi S, Valipour H, Qomi MJA (2017) Intermolecular forces between nanolayers of crystalline calcium-silicate-hydrates in aqueous medium. J Phys Chem C 121(10):5565–5572

Michaelis W (1893) Über den Portland Cement. J für Prakt Chemie Chem Zeitung 17:982–986

Nachbaur L, Mutin JC, Nonat A, Choplin L (2001) Dynamic mode rheology of cement and Tricalcium silicate pastes from mixing to setting. Cem Con Res 31(2):183–192

Pellenq RJ-M, Van Damme H (2004) Why does concrete set?: the nature of cohesion forces in hardened cement-based materials. MRS Bull 29(5):319–323

Pellenq RJ-M, Caillol JM, Delville A (1997) Electrostatic attraction between two charged surfaces: a (N,V,T) Monte Carlo simulation. J Phys Chem B 101(42):8584–8594

Pellenq RJ-M, Kushima A, Shahsavari R, Van Vliet KJ, Buehler MJ, Yip S, Ulm F-J (2009) A realistic molecular model of cement hydrates. Proc Natl Acad Sci USA 106(38):16102–16107

Plassard C, Lesniewska E, Pochard I, Nonat A (2005) Nanoscale experimental investigation of particle interactions at the origin of the cohesion of cement. Langmuir 21(16):7263–7270

Powers TC (1958) Structure and physical properties of hardened Portland cement paste. J Am Ceram Soc 41(1):1–6

Taylor H (1997) Cement chemistry. Thomas Telford Publishing, London

Thomas JJ, Jennings HM (2006) A colloidal interpretation of chemical aging of the C-S-H gel and its effects on the properties of cement paste. Cem Con Res 36(1):30–38

Valleau JP, Ivkov R, Torrie GM (1991) Colloid stability: the forces between charged surfaces in an electrolyte. J Chem Phys 95(1):520–532

Vandamme M, Ulm F-J (2009) Nanogranular origin of concrete creep. Proc Natl Acad Sci 106(26):10552–10557

Nanoscale Composition-Texture-Property Relation in Calcium-Silicate-Hydrates

Mohammad Javad Abdolhosseini Qomi, Mathieu Bauchy, and Roland J.-M. Pellenq

Contents

M. J. A. Qomi (✉)
Advanced Infrastructure Materials for Sustainability Laboratory (AIMS Lab), Department of Civil and Environmental Engineering, Henry Samueli School of Engineering, E4130 Engineering Gateway, University of California, Irvine, Irvine, CA, USA
e-mail: mjaq@uci.edu

M. Bauchy
Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA
e-mail: bauchy@ucla.edu

R. J.-M. Pellenq
Concrete Sustainability Hub, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

MSE2, the MIT/CNRS/Aix-Marseille University Joint Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Centre Interdisciplinaire de Nanosciences de Marseille, CINaM, CNRS/Aix-Marseille Université, Marseille, France

UMI MSE, CNRS/MIT/AMU Joint Laboratory "MultiScale Materials Science for Energy and Environment", Marseille, France
e-mail: pellenq@mit.edu

Abstract

The more than 20 billion tons of concrete, produced every year, is responsible for 5–7% of global anthropogenic carbon dioxide emissions. Yet, there is no other viable material that can substitute concrete to meet the need for civil infrastructure in the developed and developing countries. This leaves reducing concrete's carbon footprint as the only path forward to meet environmental targets. The strength and durability properties of concrete rely on the calcium-silicate-hydrate (CSH) phase that forms during cement hydration. Controlling the structure and properties of CSH phase is challenging, due to the intrinsic multiscale complexity of this hydration product that spans several orders of magnitude in length scale (from nanometers to microns). The existing lack in scientifically consistent insights into structure and properties of CSH has been the major obstacle to the development of greener formulations of modern concrete. In this chapter, we review how bridging general concepts from condense matter physics to cement and concrete research has revolutionized our contemporary understanding of the CSH phase and its making-up at the nanoscale, redefining this ubiquitous material described simultaneously as a spanning space continuous matrix and as a cohesive granular material that degrades and creeps over time.

1   Introduction

At the 21st Conference of Parties Climate Summit in Paris, the cement industry launched an action plan aimed to reduce its carbon emissions by 1 Gt by 2030. This plan identifies a range of actions that tackle emissions of the sector, including expanding the use of alternative fuels, developing new low-carbon cement formulations, looking into avoided emissions in concrete's use phase, and exploring novelties in the production process. This plan is envisioned to engage the building and infrastructure value chains to reduce the environmental footprint of cement and concretes as a sustainable and resilient construction solution.

Cement clinker particles are produced from a high-temperature (around 1500 °C) process that mixes limestone (calcium carbonate) and clay (alumino-silicate minerals) followed by a sintering step to form hard nodules called clinker that get finally ground into a fine powder. During this high-temperature treatment, several chemical
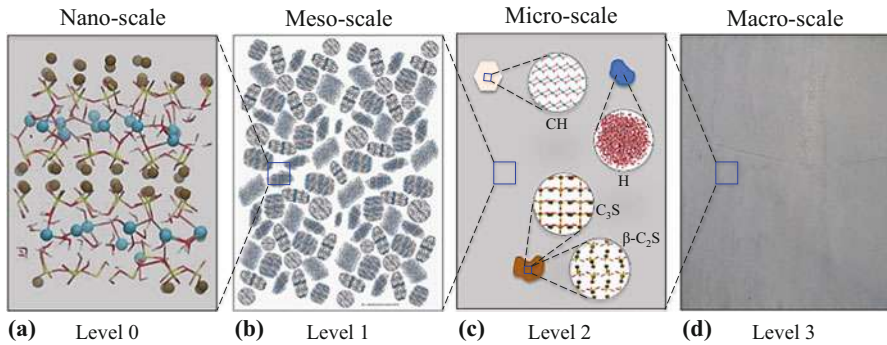
Fig. 1 A four-level thought model of cement paste spanning from the nanoscale to the engineering scale. (**a**) Snapshot of CSH's nanostructure at Ca/Si = 1.5. The brown and cyan spheres represent intra- and interlayer calcium ions. Yellow-red bars show Si-O bonds. Red-white bars depict water molecules and hydroxyl groups. (**b**) CSH's mesotexture viewed as the agglomeration of randomly oriented CSH particles with size polydispersity. (**c**) Hardened cement paste's microtexture encompassing anhydrous clinker phases, hydration products, and capillary porosity. (**d**) Cement paste's texture at the engineering scale (Qomi et al. 2015)

processes, from the loss of free water to thermal decomposition, take place (Bensted and Barnes 2002; Taylor 1997). The clinker is then ground in presence of gypsum into a fine powder to produce the "Ordinary Portland Cement" (OPC) sac, the most commonly used cement type.

The hydration process begins soon after mixing cement powder with water. The clinker is rich in crystalline calcium silicates, mainly tricalcium silicate $(CaO)_3 SiO_2$ (or $C_3 S$ in cement chemistry notation, also called alite, that is, a chemically impure form of $C_3 S$) and dicalcium silicate $(CaO)_2 SiO_2$ (or $C_2 S$ also called belite, that is, a chemically impure form of $C_2 S$), and reacts with water to form the calcium-silicate-hydrate (CSH) phase, calcium hydroxide $(Ca(OH)_2)$, and hydrated calcium aluminates (AFm and AFt phases) (Taylor 1997). Note that, in cement chemistry notation, "C" stands for CaO, "S" for SiO_2, and "H" for H_2O. It is now clear that the CSH phase is a *cohesive*, *glassy*, *porous*, *multiscale*, *nonstoichiometric compound* (by nonstoichiometric, we indicate that different CSH chemistries can coexist within the same set paste). See Fig. 1 for CSH and cement paste's schematic representation across length scales. CSH's stoichiometry is usually accounted for through the Ca/Si ratio also called the C/S ratio that ranges between 1.0 and 2.0 and distributed around C/S = 1.7. At the very nanoscale, CSH is often described as an analog of a defective tobermorite, a layered hydrated calcio-silicate mineral (with C/S ratio ranging from 0.66 to 1.0, with several possible polymorphs) (Gartner 2004).

At the mesoscale (100 nm to 1000 nm), CSH appears to be an assembly of nanograins (on the order on 5 nm in size, with some size polydispersity (Gartner and Sui 2018; Popescu et al. 2003; Quillin 2001)) with two distinct textures called low-density (around 64%) and high-density (above 76%) – the first being located at the surface of the hydrating clinker grains, while the second corresponds to regions in-between clinker grains. From SEM experiments, the low-density phase is often described as being fibrillar, although the vacuum conditions of the SEM

observations might affect the texture of a fresh paste (Neubauer and Jennings 1996). Interestingly, the small-angle scattering response of CSH to X-ray or neutron diffraction exhibits an unusual Q^{-3} behavior that was explained only recently (Q being the norm of the momentum transfer vector corresponding to a $d = 2\pi/Q$ characteristic distance) (Santos et al. 2015). The 3D percolation of CSH in-between initial reactive clinker grains during the hydration process determines the overall rigidity threshold of the paste. Alite reacts fast with water so that, after 28 days, 70% of it has been replaced by hydration products (Taylor 1997). On the contrary, belite hydrates much slower – only 30% has reacted within the same amount of time. Note that the hydration of alite or belite produces CH, in addition to CSH; CH represents 15–20% of the cement paste volume under the form of large (10s of microns) size crystallites immersed in CSH.

In this chapter, we review how bridging general concepts from colloidal and glass physics to cement and concrete research has revolutionized our contemporary understanding of the CSH phase and its making-up at the nanoscale, redefining this ubiquitous material described simultaneously as a spanning space continuous matrix and as a cohesive granular material that degrades and creeps over time. This new route in understanding the CSH phase grew out from a combination of new experimental techniques together with an important effort in modeling and numerical simulations using statistical condensed matter physics as a unifying knowledge background allowing tackling the structural and complexity of the material over critical time and length scales. This framework provides a novel opportunity to scientifically design cement material properties and hence identify strategic issues that have the potential to transform cement production and use. The purpose of this chapter is (i) to give a holistic overview of the physical chemistry and mechanics of cement hydrate, as it develops during cement hydration and progressively solidifies; (ii) to frame the outstanding questions on cement hydrate structure and mechanics within the context of modern condensed matter physics of amorphous materials; (iii) to review the recent research developments in experiments, modeling, and simulations; and (iv) to analyze the new emerging questions and their possible implications for reducing the environmental impact of cement and concrete. In contrast to the existing reviews on this subject (see, e.g., J. Thomas et al. (2011)), we are addressing the large community of condensed matter physicists and material scientists, well beyond the traditional concrete research community. Our scope is to promote fundamental scientific understanding of cement hydrates' physics to encourage new research directions.

2 The Nanoscale Process of Hydration = Dissolution of Clinker + Hydrate Precipitation

In this section, we focus on new and promising simulation approaches that advance the understanding of different phases of cement hydration products, mainly CSH. Figure 2 provides a schematic illustration of different stages of cement hydration
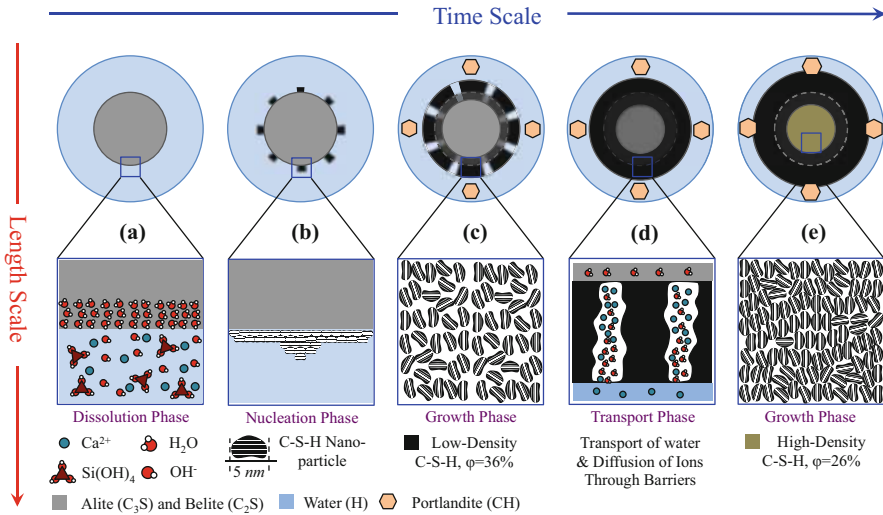
Fig. 2 The schematics of a five-stage idealized thought model of cement hydration that combines transport and boundary nucleation growth mechanisms. (**a**) Formation of a semi-impermeable layer around cement particles during the dormant period. (**b**) A symmetric nucleation of a constant number of CSH particles on the surface of cement particles. (**c**) An isotropic and self-similar growth of CSH nuclei with spatially constant but time-varying growth rate. (**d**) Overlapping of growing CSH nuclei and the coverage of the entire surface with hydrates which terminates the nucleation-growth process. (**e**) The uniform growth of CSH particles that is controlled chiefly by the inward transport of water and the outward diffusion of ions to the solution (see Rahimi-Aghdam et al. 2017)

including dissolution, nucleation, and growth (Rahimi-Aghdam et al. 2017). At the beginning of the hydration, the CSH grows loosely around the clinker phase starting from nuclei; later it densifies and glues together also the crystalline phases. The morphology of CSH can vary from globules to fibril or foil-like depending on the initial conditions and the hydration stage (Richardson 2000). The initial composition of the cement powder, the water-to-cement (w/c, mass basis) ratio, and the relative humidity influence CSH structure. The fibril growth has been suggested to rise from the increased osmotic pressure that ruptures the preexisting CSH layer (Richardson 2000). Jennings and co-workers investigated the size and diameter of the needles formed under different conditions of relative humidity (Fonseca and Jennings 2010). It is also found in microscopy images that CSH grows in different directions. CSH grows both outward and inward the clinker particles, forming the so-called inner and outer product, respectively. The inner product occupies the space of the hydrating grain, and its structure is more compact and amorphous. The outer product is formed in the pore solution, and it is reported to form bumps, needles, or foils.

A complementary way to examine the cement paste's structure is by characterization and distribution of void spaces in the material. The initial amount of water influences significantly the overall porosity of the material; the higher

the w/c ratio, the more porous the material. Initially, the voids in the cement paste are large and connected; however, as the hydration proceeds and the solid structure percolates, the pores reduce in size and get disconnected. Initially large capillary pores of few μm are replaced by pores of few decades of nm at the later stages or cured paste. Proton NMR experiments provide information about the water movement and the porosity in cement paste (Korb et al. 2007b; McDonald et al. 2010; Muller et al. 2013a). More specifically, these experiments indicate that there are consistent pores of 2–5 nm and 14–30 nm, both associated with the CSH gel. In order to interpret the experimental findings of the CSH structure, models were developed based either on the hypothesis of the planar growth of naturally occurring tobermorite, jennite, and portlandite minerals or the aggregation of colloidal particles (Bullard et al. 2010). According to the first hypothesis, the CSH gel is formed by interconnected layers that are similar to mineral crystals (Daimon et al. 1977) following the cartoon representation proposed by Feldman and Sereda (1968), where the layers are irregularly stacked and water is trapped in between them or adsorbed in the free surfaces. This hypothesis could explain some properties of the nanostructure such as the observed variable composition of C/S ratio, the crystalline regions, and the amorphous layers, and they were validated by gas sorption properties, porosity, and shrinkage. Gartner et al. (2017) described the layers as narrow ribbon-like with random defects that make them wavy and cause them to separate and branch. The growth rate of these ribbons was considered exponential assuming that, due to the kinetics of the hydration, nucleation and growth of the new material happen at the extremities of the ribbons. Gartner suggested that the growth of the ribbons and their branching or impingement could correspond to the acceleration and deceleration regime, respectively, of the kinetic hydrations (Gartner 1997). The indication of the colloidal nature of CSH came with the small-angle neutron scattering (SANS) experiments of Allen in the 1980s (Allen et al. 2007). The scattering data of a hydrating cement paste showed a characteristic length of $\sim$5 nm after the induction time and a correlation length $\sim$ 40 nm developing due to aggregation (Allen et al. 1987). Based on these scattering measurements and sorption data, Jennings proposed a hybrid model of globules with a peculiar shape and inner structure. According to his first colloidal model (CM-I), these particles form aggregates with two different packing densities: high-density (HD) CSH and low-density (LD) CSH (Thomas and Jennings 2006). A further refinement of Jennings' model (CM-II) accounted for the smallest pores of the CSH arising from the internal structure of the globules that, like in the picture proposed by Feldman and Sereda (1968), are composed of tobermorite and/or jennite sheets and water. Therefore, the CM-II combines a layer-like structure into the colloidal-like description and accordingly interprets the adsorption isotherm experiments and processes like creep and shrinkage (Jennings 2000).

Recent papers (e.g., Scrivener et al. 2015; Thomas et al. 2011) provide a comprehensive review of modeling efforts on the cement hydration process over the past 40 years, covering different modeling approaches such as single-particle models, mathematical nucleation and growth models, and vector- and lattice-based approaches to simulating microstructure development. In this section, we took the step to consider advances in modeling based on modern solid-state quantum and

statistical physics and importing general concepts from colloid and glass science, which are critical to advance the science and engineering of cement paste.

2.1   Nanoscale Description of Clinker Surface Reactivity

The hydration of mineral surfaces is a critical process that involves multiple coupled chemical reactions as well as topological changes. These features make computational modeling of the hydration process very challenging. For cement clinker, most of the modeling effort concentrates on C_3S. The contact of C_3S with water leads to the hydroxylation of silicate particles and dissolution, releasing calcium and silicate ions in the solution. It has been shown that Ca^{2+} is the main parameter controlling the hydration kinetics and the stoichiometry and structure of the reaction products such as CSH. The concentrations of the solution ([CaO], [SiO$_2$]) do not vary independently during C_3S or cement hydration. In their recent work, Manzano et al. have used reactive force-field simulations to understand the surface properties, hydration, and dissolution of C_3S (Manzano et al. 2015). Furthermore, they showed that the computed static quantities, i.e., surface energies and water adsorption energies, do not provide useful insight into prediction of mineral hydration because these quantities do not account for major structural changes at the interface, when dynamic effects are included. As water reacts with the surface, significant topological changes take place modifying the surface structure, and the correlation with the surface properties is lost. Therefore, the authors concluded that the hydration of minerals couldn't be predicted from the static properties as has been traditionally done. Hence, including dynamics becomes necessary to observe topochemical reactions.

The hydration mechanism of the reactive surfaces of C_3S can be summarized as a reaction and hopping process. Water reacts with the surface oxygen atoms forming hydroxyl pairs. Then, protons penetrate into the crystal, hopping from surface to inner oxygen atoms following a Grotthuss-type mechanism. The protonation reactions break the calcium-oxygen bonds, and calcium desorbs from the bulk to the surface, forming inner- and outer-sphere complexes. Overall, the surface of the crystal evolves toward a disordered calcium-silicate-hydrate phase, and, as a result, the reaction rates homogenize. In addition, Manzano et al. (2015) also observed the formation of a stable icelike water monolayer driven by the peculiar underlying crystalline order of some C_3S surfaces that can inhibit hydration.

First-principles DFT-type calculations were employed for the first time by Durgun et al. (2012) to provide a fundamental understanding of the relationship between the reactivity of synthetic calcium silicate clinker phases and their electronic structure with the aim to shedding light on the wide range of hydration kinetics observed in different phases of calcium silicate. The authors showed that the high reactivity of C_3S is mainly related to the reactive sites around its more ionic oxygen atoms. C_2S does not contain these types of oxygen atoms, although experiments suggest that impurities may play a role in changing the reactivity of these materials. The electronic structure of a wide range of possible substitutions in both C_3S and C_2S shows that while the influence of different types of impurities on structural

properties is similar, their effect on reactivity is very different. These first-principles calculations suggest that the variation of electronic structure is mainly related to the formation of new hybridized orbitals and the charge exchange between the impurity atoms and the bulk material. These differences are shown to influence the reactivity in these materials, in good agreement with available experimental data. By quantifying the charge density localization of the VBM and CBM (valence band maximum and conduction band minimum, respectively) for a large set of potential substitutions, it was predicted that several candidates could influence the reactivity of C_2S.

Manzano et al. have used a combination of classical and quantum mechanical simulation methods to study the detailed physicochemical changes of C_3S and C_2S upon incorporation of Mg, Al, and Fe guest ions (Manzano et al. 2011). Using classical force-field methods, they considered random substitutions among possible sites and different compositions in order to identify the preferential substitution sites on the crystalline structures. Using quantum mechanical density functional theory calculations, the electronic structure of representative configurations has been computed to determine the potential impact of impurities on the reactivity. Simulations show that any crystallographic site within the unit cell is equally probable for the incorporation of guest ions and that there is not an optimal distance between them, when multiple incorporations occur. The rearrangement of the silicate groups' orientation and calcium coordination polyhedron allows for defect accommodation. Despite the change in orientation, the internal structure of the silica groups is kept unaltered, while the change occurs in the more flexible Ca-O distances. Manzano et al. (2011) found that C_2S tighter structure makes it more susceptible to structural changes when guest ions are incorporated. Regarding the reactivity, DFT calculations reveal that the reactive areas under nucleophilic and electrophilic attack change differently depending on the guest ion. Mg docking does not change the electronic structure appreciably, while Al and Fe do, thereby reducing the number of reactive sites. Aluminum, however, creates weak Al-O bonds that are therefore highly reactive, balancing the reduction of dissolution points. By contrast, iron dramatically decreases the reactive sites and introduces a strong Fe-O bond. This difference might induce a noticeable decrease in the dissolution rate with respect to the pure or the Mg- and Al-substituted alite and belite. Along the same lines, Huang et al. (Huang et al. 2015) combined first principles (DFT) and reactive molecular dynamic simulations, to provide a similar picture of the influence of Mg, Al, and Fe impurity species on the hydration reactivity of triclinic C_3S phase.

2.2 Nanoscale Modeling of CSH Precipitation from the Initial Electrolyte Solution

The formation of CSH out on the so-called pore electrolytic solution that exits in-between clinker grains as they dissolve remains an open question. Here, we describe the work of Dolado et al. (2011), who took the step to set a direct simulation

of CSH precipitation using molecular dynamic simulation using the Garofalini reactive potential for silica with the aim of reconciling the existing structural and colloidal/gel-like models. The key result of this work is the prediction of the formation of a branched three-dimensional CSH solid network with segmental branches (SB) of $\sim 3 \times 3 \times 6$ nm in size. These SB units of CSH are essentially glassy and do not reproduce the layered nanoscale texture that has been observed from many TEM experiments over the years. The reason behind this is most probably that the system got stuck into some local potential minima, being unable to overcome large energy barriers. There are two ways to overcome such drawbacks. The first would be to use a replica-exchange molecular dynamic approach (also often called parallel tempering (Sugita and Okamoto 1999)). The second strategy can build on the replica-exchange approach but will implement a coarse-grain description from a full atomistic description to basic relevant structural units such as silica tetrahedra and calcium ions in a dielectric continuum that represents the solvent, i.e., water in the case of CSH. For a good example of this approach, we refer the interested reader to the work of Chien et al. (2015). These authors have performed replica-exchange reaction ensemble Monte Carlo simulations to study the low-energy crystalline structures of a reactive model of silica. The implementation of replica-exchange Monte Carlo involves simulating several system copies, each with its own value of the equilibrium constant controlling silica condensation/hydrolysis reactions, which are essential for building higher-order network structures and eventually crystals. These replica-exchange simulations were found to traverse energy landscapes from amorphous to crystalline phases, yielding the dense silica polymorphs. The good predictive power of this approach is based on advanced statistical physics methods that open the door to future simulation studies of the free energy barriers controlling crystallization processes.

2.3 Origins of Cohesion Between CSH Grains and Its Evolution During Cement Hydration: Bridging from the Nanoscale to the Mesoscale

The bottom-up approach for multiscale modeling of cementitious materials integrates different scales of knowledge to accurately model relevant phenomena across multiple length and time scales. This approach provides a unique opportunity to develop a full predictive modeling and simulation framework for amorphous complex porous materials from molecular to engineering scales. Our starting point is the atomic structure from which, using the state-of-the-art interatomic potentials, one can derive the so-called potential-of-mean-force (PMF) between different phases using free energy perturbation approach within the framework of statistical physics. A PMF is a free energy function representing the interactions between objects at a given temperature along a pre-specified reaction path, sometimes also referred to as a collective variable. The PMF, therefore, is the key information piece allowing to coarse-grained physical-chemistry information within a mesoscale

model formulation. Such a coarse-grained modeling integrates atomistic response into interparticle potentials that fully describe the atomic interactions and enables us to reach much larger time and length scales.

The cohesion and stability of charged layers separated by an electrolyte solvent have long been discussed within DLVO theory framework. The DLVO theory combines the repulsive electrical double-layer interaction with the attractive van der Waals interactions (Israelachvili 2011). The repulsive electrostatic interaction is obtained by analytically solving the Poisson-Boltzmann equation of a system made of positive point charges between negatively charged surfaces. Designed for point charges, it is, therefore, a mean-field approach that ignores correlations in ionic positions and their fluctuations due to thermal agitation. Making use of the Poisson-Boltzmann equation, the electrostatic pressure P (i.e., the force per unit area, i.e., the derivative of the PMF with respect to the inter-colloidal distance) between two charged surfaces in function of the surface separation D is given by $P(D) = k_B T[C_{ion}(D/2) - C(\infty)]$ where k_B is the Boltzmann constant, T is the temperature, and $C_{ion}(D/2)$ is the cation concentration at the mid-separation. This pressure depends only on the increase in the ionic concentration at the center of the pore, and therefore its origin is entropic.

It is now known that a system of charged plates and counterions can generate strong attractive electrostatic interactions, the so-called ionic correlation forces (ICFs) (Pellenq et al. 1997, 2008; Pellenq and Van Damme 2004). The so-called primitive model was introduced in the 1980s to study such attractive forces. In the primitive model, the charged lamellae and the hydrated cations are represented by uniformly charged infinite planar surfaces and by charged hard spheres, respectively, while the solvent is represented by a dielectric continuum. Monte Carlo simulations in the framework of the primitive model showed that fluctuations of the ionic concentration in the interlayer space result in an instantaneous polarization and give rise to an attractive ionic fluctuation interaction, the so-called ionic correlation forces. Thus, the behavior of the system is governed by a competition between repulsive entropic forces and attractive energetic forces. For divalent or trivalent cations, the attractive energetic contribution can exceed the repulsive entropic fluctuations. For example, in the case of calcium counterions, attractive pressures are observed for pore sizes larger than 10 Å and explain the experimentally observed difficulty in complete exfoliation of calcium montmorillonite clay layers.

The primitive model has obviously its own limitations. The surfaces of the slit pore are assumed to be flat and uniformly charged; the ions are supposed to be fully hydrated and mobile, and the solvent is considered to be a dielectric continuum, the dielectric constant of which is that of bulk water. These hypotheses make the primitive model questionable, in particular at low surface separation. In this case, the molecular structures of the solvent and the surfaces have to be directly taken into account. Therefore, numerous works adopted a full-atomistic description to study, for instance, the intralamellar swelling of smectite clays (Ebrahimi et al. 2012, 2014) or cohesion between CSH layers (Bonnaud et al. 2016; Masoumi et al. 2017a, b). However, the computational cost of these atomistic simulations is very significant preempting coarse-grained simulations. Accordingly, it is necessary to

determine more precisely the complexity that needs to be introduced into the models of lamellar systems and the key physical ingredients responsible for the cohesion of such materials.

During cement hydration, the cement paste has a high ionic concentration of hydroxyl ions, resulting in 10 to 13 pH levels. These ions interact with the surface of the already formed CSH particles, thereby giving them high surface charge – on the order of one electron charge per $0.35 \ nm^2$ for fully ionized CSH. To compensate the highly negatively charged CSH surfaces, Ca^{2+} counterions gather in the vicinity of CSH surfaces forming an ionic cloud. These two characteristics of highly charged particle surface and divalent surrounding counterions make the interactions of CSH particles different than other colloidal systems. In weak electrostatic coupling conditions, it has recently become clear that the distribution of the ionic cloud of monovalent counterions cannot be described by the Poisson-Boltzmann equation that constitutes the part of the DLVO theory that deals with electrostatics in a mean-field fashion (Carrier 2013). This conclusion was drawn out from the primitive model describing ionic ICFs using explicit solvent (Israelachvili 2011; Pellenq et al. 1997, 2008; Pellenq and Van Damme 2004). In this case, ion-ion correlations should be considered explicitly. The excess counterions near a CSH surface may be correlated with a deficit of counterions in the opposite layer, giving rise to attractive ion-ion correlation forces. These ion-ion correlated density fluctuations give rise to an attractive force, akin to London dispersion forces that arise from correlations between instantaneous electronic dipoles.

This nonclassical feature of the double layer, inducing an attraction between two similarly charged surfaces immersed in an electrolyte, has been explored within the theory of equilibrium electric double layer, where two regimes can be distinguished (Attard 1996). The first regime is an attractive regime that occurs at small separation distances and in presence of multivalent ions. The forces in this case change from repulsive to attractive as the surface charge and pH increase. The second regime occurs at larger separations, and especially in case of multivalent ions, the force can be alternately attractive and repulsive depending on the electrostatic coupling. For highly charged surfaces such as that of CSH and the divalent counterions (Ca^{2+}), simulations and theory predict strong attractive force more than one order of magnitude larger than the van der Waals interactions. To account for the ion-ion correlations, the following models and simulations were performed treating ions explicitly (Jönsson et al. 2004, 2005; Pellenq and Van Damme 2004; Pochard et al. 2010). Within the mesoscale model, all ions are treated as point charges and the water as a dielectric continuum. Two parallel walls with smeared surface charge represent the opposite charged surfaces of CSH particles, and, in the pore solution between the two surfaces, there are Ca^{2+} and OH^- ions. All ions interact solely with Coulomb interactions and satisfy electroneutrality within the simulation cell. Monte Carlo simulations were performed in the canonical ensemble in order to calculate the ionic distributions and the osmotic ICF pressure (Jönsson et al. 2004; Pellenq et al. 1997; Pellenq and Van Damme 2004).

At this point, it is worth mentioning some numbers about the ionic concentrations in CSH that will manifest the high ion concentrations near the particle surfaces. The

bulk ionic concentrations at pH >11.6, when charge reversal occurs, hence when the forces become attractive, are of the order of 10^{-3} M, the average separation of the ions is 10 nm, and their electrostatic energy is much lower than their thermal energy. The range of electrostatic interactions in the solution is given by the Debye screening length that in this case is 4 nm and the Gouy-Chapman length that indicates the width of the confinement layer next to the surface is of the order of 0.1 nm (Jönsson et al. 2004, 2005; Pellenq and Van Damme 2004). Hence most counterions are confined in the vicinity of the surfaces, and this layer around the surface has a much higher ionic concentration than the bulk. The main finding of these simulations confirms the finding of AFM experiments, i.e., strongly cohesive forces appear in between the CSH particles.

At the sub-nanoscale, atomistic simulations have been recently performed within the CSH globules to evaluate the cohesion between the layers of tobermorite-like CSH using Hartree-Fock, quantum DFT, or force-field methods (Gmira et al. 2004; Pellenq et al. 2008; Pellenq and Van Damme 2004). The atomistic simulations have elucidated the role of interlayer calcium ions and have been used to compute the hydrogen bonding scheme of crystals such as jennite and tobermorite (Churakov 2009) or to investigate the cohesion of tobermorite layers (Gmira et al. 2004; Pellenq et al. 2008; Pellenq and Van Damme 2004). The mechanical properties and several spectral properties of the sub-nanometric structure of CSH could also be obtained (Manzano et al. 2013; Pellenq et al. 2009). Static ionic interactions govern the sub-nanoscale, whereas at larger length scales, dynamic correlations dominate due to charge fluctuations. Therefore, in the case of CSH, it all comes down to the evolution of the attracto-repulsive PMF that changes as the hydration proceeds: as more and more ions are getting stuck into solid phases (CSH and CH), the potential shoulder decreases, and in the case of the hardened paste, one is left with a single well PMF. Note that in the case of clay, the PMF could exhibit up to three minima corresponding to different water contents, in agreement with experiments (Ebrahimi et al. 2014).

Interestingly enough, recent full atomistic simulations have confirmed the role of ICFs in the case of tobermorite and CSH and provided reference PMFs that can be readily used for mesoscale simulations of CSH (Bonnaud et al. 2016; Masoumi et al. 2017a, b). See Fig. 3 for an all-atom PMF calculation between tobermorite layers using free energy perturbation technique. These PMFs, which are in quantitative agreement with AFM data (Plassard et al. 2005), measure the cohesion with CSH grains, as shown in the table below.

	Contact surface area (nm^2)	Cohesive pressure (GPa)
CSH model A (Bonnaud et al. 2016)	26.15	−0.9
CSH model B (Bonnaud et al. 2016)	26.15	−0.07
CSH model C (Bonnaud et al. 2016)	9.95	−1.0
Tobermorite model (Masoumi et al. 2017a, b)	26.54	−6.0
Exp. (Plassard et al. 2005)	64.00	From −0.6 to −3.0
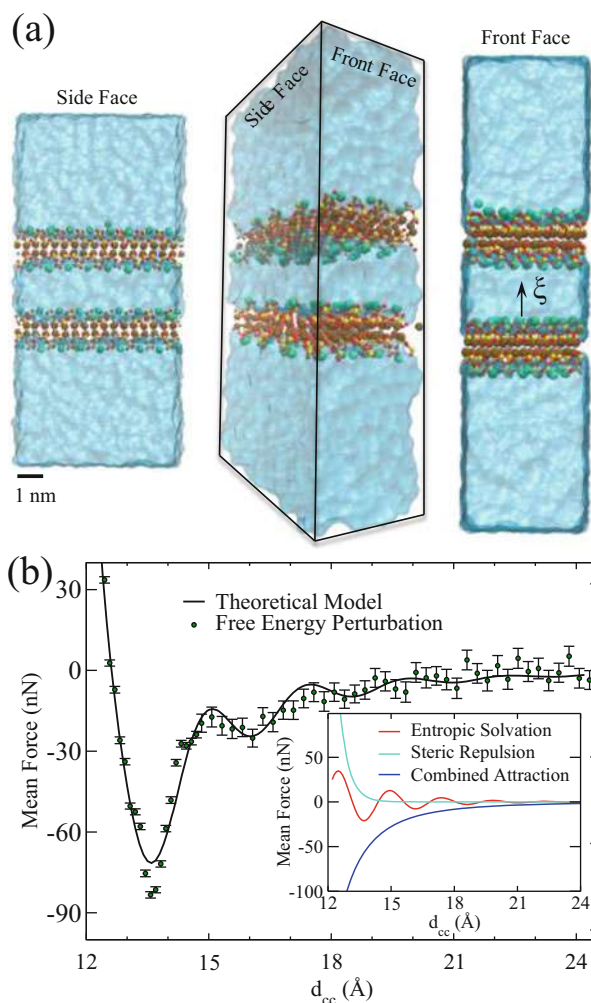
Fig. 3 All-atom free energy perturbation calculation of PMF between nanolayers of tobermorite. (**a**) The configuration of CSH layers solvated in water. (**b**) The mean force between CSH layers as a function of separation distance. The inset shows the force decomposition; see Masoumi et al. (2017a, b) for details

Combining advanced statistical physics approaches with realistic PMFs for the simulation of cement precipitation and settings, Ioannidou et al. (2014, 2016) investigated at the mesoscale the development of CSH gels under out-of-equilibrium conditions using a realistic CSH grain-grain PMF (Ioannidou et al. 2014, 2016). These authors proposed a new model and numerical approach to follow the gel formation upon precipitation and aggregation of nanoscale colloidal hydrates, whose effective PMF interactions are quantitatively consistent with forces measured in

experiments at fixed lime concentrations and also with recent molecular simulations (Bonnaud et al. 2016; Masoumi et al. 2017a, b). In the work of Ioannidou et al., a Grand Canonical Monte Carlo (GCMC) scheme was introduced for the first time in order to mimic the precipitation events (of CSH grains) by *NVT* molecular dynamic simulation steps, which mimic a rate effect corresponding to the hydrate production set by the chemical environment. The simulated hydrate precipitation curve indeed reproduces the acceleration and deceleration regime typically observed in experiments correctly capturing the effect of lime concentration on the hydration kinetics and the gel morphology. The analysis of the evolution of the gel morphology indicates that the acceleration is related to the formation of optimal local dense packing that allows for growth of large and elongated aggregates. It is important to mention that this is the first prediction of the heat curve with no pre-assumed kinetic models that always contain ill-defined and/or unmeasurable parameters (Thomas et al. 2011). Overall, Ioannidou et al.'s efforts clarify the connection between the hydration kinetics and the changes of the gel morphology.

3 Hardened Cement Paste

At the microscale, many crystalline phases are observed in hardened cement paste together with the amorphous CSH. The most common crystal is the portlandite, which can grow up to several micrometers in the pore solution and close to the clinker surface. The size of the crystals depends on the temperature and w/c ratio. The ones growing in the pore solution under ideal conditions form hexagonal plates, whereas the ones growing close to the clinker surface are not perfectly shaped. Other crystal phases are the AFm and AFt that form needles of around 1 μm close to the clinker particles (Bullard et al. 2010).

3.1 Chemical Composition and Nanoscale Structure of CSH

Many experimental techniques elucidate the structure of CSH at the nanoscale (Fig. 4). The main findings of the CSH's nanostructure come from analog models that interpret X-ray diffraction (XRD), sorption, shrinkage, NMR, and AFM experiments. In the 1950s, based on XDR evidence, Bernal (1954) suggested that CSH has a layered structure related to a crystalline mineral – 1.1 nm tobermorite – and called it tobermorite gel (Taylor 1993). In general, different types of CSH can coexist within a same cement hydrate paste with variable stoichiometry characterized by the water content and the calcium to silicon (Ca/Si) ratio that ranges from 1.2 to 2.3 with a maximum abundance at 1.7. The reason for the coexistence of all these different chemical compositions has to be established on proper thermodynamic grounds based on free energy arguments. CSH is usually described as amorphous due to its X-ray diffraction (XRD) pattern that shows few (and rather broad) peaks (Meral et al. 2011; Skinner et al. 2010). Tobermorite minerals have lamellar structure of calcium oxide layers connected on both sides by silicate chains
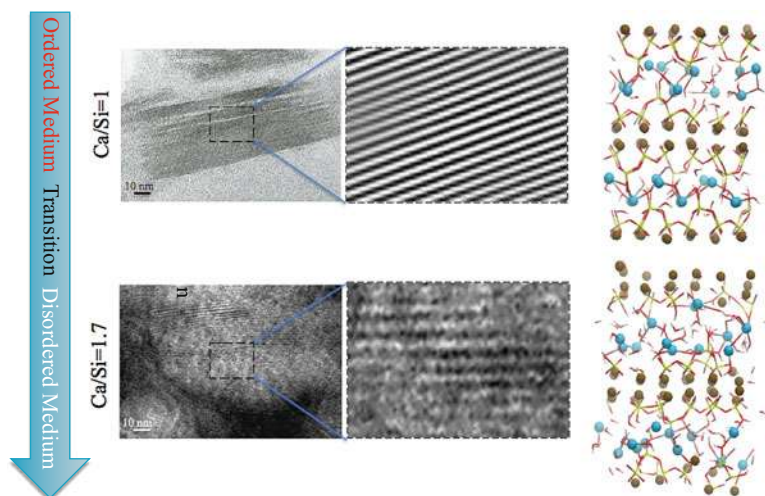
Fig. 4 The textural evolution of CSH with increasing Ca/Si ratio. At low Ca/Si ratio (Ca/Si = 1), both the TEM and MD simulations show ordered crystalline structure, in which calcium-silicate chain layers are distinctively very long. However, at high Ca/Si ratios (Ca/Si = 1.7), both TEM and MD simulations indicate defective structures that lead to a disordered glassy structure. For details, see Pellenq et al. (2009) and Qomi et al. (2014b)

following dreierketten arrangement, i.e., based on three silicate tetrahedra – two in the vicinity of the calcium oxide layer and one bridging to the next dreierketten group (Bonaccorsi et al. 2005). Note that, at C/S > 0.66, tobermorite layers acquire a nonzero negative charge with the hydrolysis of O-H bonds attached to silicate groups. This surface charge is then compensated by additional hydrated interlayer calcium ions.

It has been proposed that CSH could be made of crystalline tobermorite-like layers mixed with jennite, another layered calcio-silicates with C/S = 1.5, which is the base of the so-called T-J model of Richardson (Richardson 2008). Obviously, a mixture of compounds with C/S altogether less than 1.5 cannot account for the fact that the C/S distribution in CSH can average at 1.7. This explains why the idea of nanocrystalline portlandite was put forward. To date, no experimental evidence of such nanocrystalline portlandite has been found. More experimental findings for the chemical composition of CSH were provided by Nonat et al. when they investigated alite hydration with a special setting that allowed them to control the calcium hydroxide concentration in solution (Garrault et al. 2005, 2006; Nonat 2004; Plassard et al. 2005). They found that the calcium hydroxide concentration determines the Ca/Si ratio of the CSH and the formation of the CSH network. The evolution of the stoichiometry of CSH can be linked to the lime concentration of the solution. According to this, Nonat et al. categorized CSH into three classes with different C/S ratios and developed a predictive thermodynamic equilibrium model

based on bulk and surface mass laws that are reliable at local equilibrium conditions (Nonat 2004).

Solid-state nuclear magnetic resonance measurements of ^{29}Si-enriched triclinic C$_3$S enabled the quantitative monitoring of the hydration process in terms of transient local molecular composition, extent of silicate hydration, and polymerization. This provided insights into the relative influence of surface hydroxylation and hydrate precipitation on the hydration rate. When the rate drops, the amount of hydroxylated C$_3$S decreases, thus demonstrating the partial passivation of the surface during the deceleration stage. Moreover, the relative quantities of monomers, dimers, pentamers, and octamers in the CSH structure can be measured (Pustovgar et al. 2016).

In 2009, Pellenq et al. (2009) proposed a molecular model of CSH based on a bottom-up atomistic simulation approach that considers only the chemical specificity of the system as the overriding constraint. By allowing for short silica chains distributed as monomers, dimers, and pentamers, this CSH archetype of a molecular description of interacting CaO, SiO$_2$, and H$_2$O units provides for a realistic value of the C/S ratio (set at 1.65 and corresponding to the most abundant form of CSH in a cement paste). The mass density computed after Grand Canonical Monte Carlo simulation of water adsorption at 300 K is 2.43 g/cm^3 that comes very close to the CSH density measured by NMR (Muller et al. 2013b). Note that the densities of both tobermorite (Ca/Si = 0.83, ρ = 2.18 g/cm^3) and jennite (Ca/Si = 1.5, ρ = 2.27 g/cm^3) are significantly lower than that of CSH. Jennings et al. (Allen et al. 2007) using SANS measurements also determined the mean C/S ratio to be at 1.7 and the density of CSH at 2.6 g/cm^3 – a value that is a bit larger than that reported by McDonald et al., i.e., 2.45 g/cm^3 (Muller et al. 2013b). This discrepancy can be rationalized from the fact that it is necessarily an upper bound as the technique made no distinction between in-CSH grain structural water and adsorbed water at the surface of capillary pores. The simulated model of CSH, with a chemical composition of (CaO)$_{1.65}$(SiO$_2$)(H$_2$O)$_{1.75}$ (see Fig. 4), also predicts other essential structural features and fundamental physical properties amenable to experimental validation, which suggests that the CSH gel structure includes both glass-like short-range order and layered structure features of the mineral tobermorite. Additionally, the predicted mechanical stiffness, strength, and hydrolytic shear response compare well to experimentally measured properties of CSH using nanoindentation.

This work was then extended to studying all the possible chemical compositions of CSH measured in terms of C/S spanning from 1.2 to 2.1 (Qomi et al. 2014a). A database for the combinatorial screening of mechanical behavior was then obtained by creating an ensemble of atomic structures of CSH with each member characterized by a given value of the C/S ratio according to the multi-statistical physics-based method approach shown in Fig. 5. All the (150+) atomistic models were created consistently, according to the procedure in which randomly chosen electroneutral SiO$_2$ groups are removed from a pristine 11 Å tobermorite crystal structure under the constraint of a given C/S ratio. Subsequent molecular simulation steps in the course of this procedure accounted for the chemical reactivity of
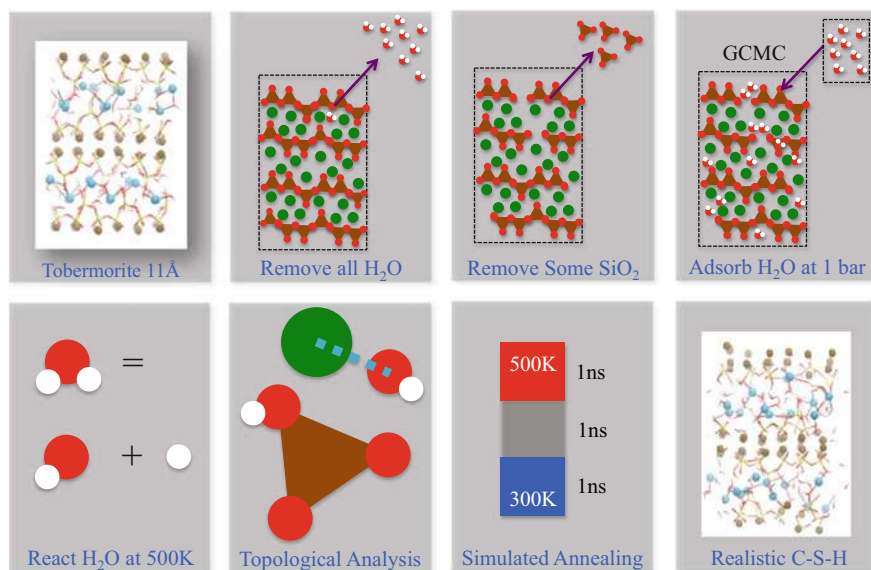
Fig. 5 The eight stages of model construction. These stages are strictly followed 150 times to produce CSH samples with varying Ca/Si ratio ranging from 1.1 to 2.1

the silicate chains with water through the use of the ReaxFF empirical reactive potential.

There models provide a quantitative means to relate the composition of CSH to its mechano-physical properties at the nanoscale. Figure 6 shows the dependence of mechanical properties of CSH to C/S ratio. Both the indentation modulus and hardness decrease as a function of C/S. The superior strength and stiffness required for reducing cement consumption can be envisioned with the altercation of the chemistry of CSH to lower C/S ratios. Recently, using high-pressure X-ray diffraction, it was shown that at low C/S ratios, the densification can increase the mechanical properties of CSH despite the increase in the defect content, and its possibilities were verified using atomistic simulations (Geng et al. 2017). Furthermore, it was shown that incorporation of aluminum in the molecular structure of CSH creates crosslinking between CSH layers, increasing its mechanical properties (Qomi et al. 2012). There is an important combinatorial aspect of this work as it is clear that, for a given C/S ratio, a (finite) number of atomistic configurations could be built, relaxed, and subjected to GCMC simulation of water adsorption. We mentioned earlier that the fact that the same CSH paste could host different local composition characterized by its C/S ratio had to be explained by an argument. Indeed, it is one of the main outcomes of the work of Qomi et al. (Qomi et al. 2014a; Qomi et al. 2015) that all CSH polymorphs for a given C/S ratio turn out to exhibit the same free energy (see Fig. 7 (Qomi et al. 2014b)). As expected, low C/S CSH polymorphs are found to be more stable than those at larger C/S indicating
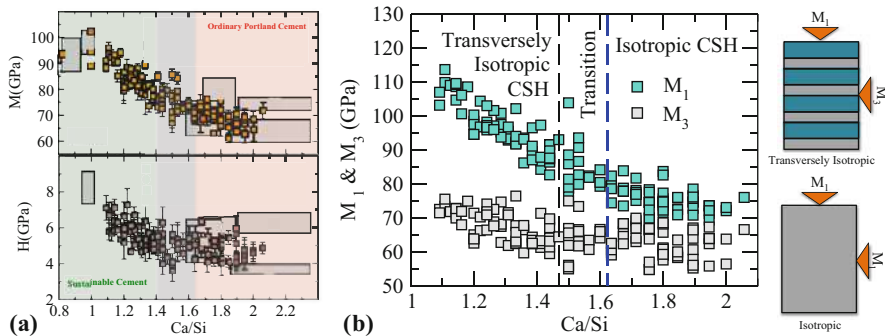
Fig. 6 Mechanical properties of CSH. (**a**) The indentation modulus (M, orange squares) and hardness (H, brown squares) of CSH as a function of Ca/Si ratio. The results of the atomistic simulations are compared against nanoindentation results (gray squares). (**b**) CSH's anisotropy. At low Ca/Si ratios, the CSH has a transversely isotropic structure as indicated in the difference between M_1 and M_3. See Qomi et al. (2014b) for details

Fig. 7 Vibrational free energy of CSH models as function of Ca/Si ratio that indicates the presence of polymorphism in CSH structure (Qomi et al. 2014a and Qomi et al. 2015)
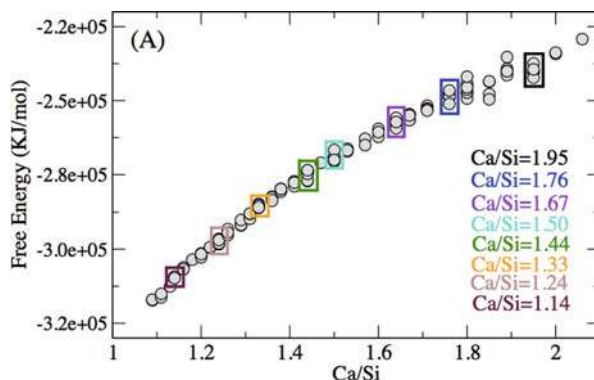
that tobermorite is more stable than CSH and that CSH over time should turn into tobermorite if silica sources are provided.

More interesting, for any given C/S value, all CSH polymorphs have the same free energy indicating that they have the same equal probability to occur whatever is the distribution of silica chunks in terms of monomers (Q^0), dimers (Q^1), and longer silica chains (Q^2). Note that the Q^n notation refers a SiO_4 tetrahedral unit that is surrounded by n bridging oxygen atoms (with $n = 1$–4). Hence, Q^0 indicates an isolated silica tetrahedron; Q^1 refers to end members in a silica chain, i.e., a dimer is therefore made of two Q^1-type silica tetrahedra; and Q^2 refers to silica tetrahedra that are in the core body of a silica chain, hence having two tetrahedra site as immediate neighbors, etc. Importantly, these nanoscale models of CSH were validated against a large body of experimental evidence ranging from water content (from drying experiments), Ca-OH bond content (from neutron scattering experiments), silica mean chain length (from solid-state ^{29}Si NMR), and total scattering function (as obtained from synchrotron-based X-ray spectroscopy) as

shown in Fig. 2 of Qomi et al. 2014a. The agreement between simulations and experimental results is impressive as there are no adjustable parameters in the simulation protocol. In particular, one should focus on the GCMC prediction of water content upon increasing amount of SiO_2 vacancies. This result remains unique and clearly indicates the validity of the approach: for every missing SiO_2 group, two water molecules can be docked in the CSH nanotexture – which is totally meaningful in terms of created volume accessible to water molecules based on a simple van der Waals radius argument. The use of the ReaxFF reactive force field further allowed some water molecules initially inserted in GCMC fashion to react and split into Ca-OH and Si-OH groups.

3.2 CSH at Large C/S Values: The Nano-portlandite Hypothesis and the Fate of Silica Monomers

Another very interesting point coming out of the texture analysis of the CSH models across the range of C/S values is that CSH's density increases as the C/S ratio increases, which was thoroughly confirmed by synchrotron scattering experiments (Soyer-Uzun et al. 2012). Interestingly, recent molecular simulations based on reactive interatomic potentials of tobermorite crystals doped with $Ca(OH)_2$ moieties contradict these experiments, hence disqualifying the "nano-portlandite hypothesis" that was put forward in order to account for CSH at the large C/S values in the T-J model of Richardson (Richardson 2004). We now focus on analyzing the fate of silica chunks in the nanoscale models of CSH across the span of C/S values. As shown in Fig. 2c of Ref. (Qomi et al. 2014a), the evolution of the silica mean chain length is well accounted for in the simulated models. Despite all prediction of the CSH models across the range of C/S values, the approach was highly criticized, in particular concerning the large number of silica monomers present at large C/S values, apparently contradicting the commonly accepted belief that Q^0 monomers were remains of hydrated clinker grains and that CSH only contains silica dimers, pentamers, and octamers. Figure 8 presents the average number of Q^0 (monomer) for a given family of CSH models at fixed C/S value as a function of the C/S ratio. There is a sharp transition at C/S = 1.5 where this number jumps from 3 +/− 1% to 14 +/− 2%. When convoluting this curve with the experimental C/S distribution, one obtains an average value of the number of monomers of around 6 ± 1%. This value is in very good agreement with the recent high-resolution NMR work of d'Espinose et al. (Pustovgar et al. 2016), who were able to quantity Q^0 fraction in CSH to be at 5 ± 2%. This confirms that silica monomers can exist in CSH up to some large amount and can be stable due to O-H covalent and/or O-Ca ionic bonds. Interestingly, the Si-O and Ca-O pair distribution functions for the CSH models at large C/S value are characteristic of oxide glasses, while at low C/S value, these are close to that found in crystalline tobermorite (Qomi et al. 2014a; Bauchy 2014; Bauchy et al. 2014b, c). The overall picture that emerges is that, at the nanoscale, CSH is a collection (of grains) of different chemistry (measured by the local
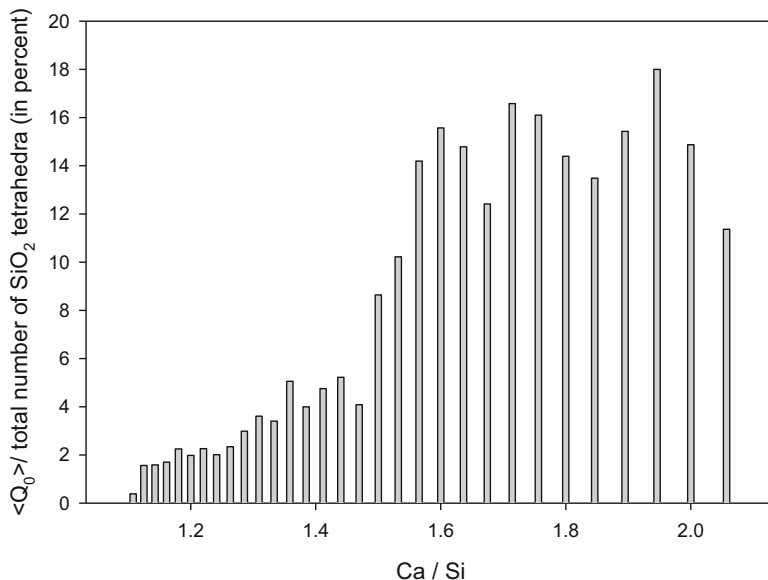
Fig. 8 The number of Q_0 as a function of Ca/Si ratio in CSH models

C/S ratio) – each grain having a structure reminiscent of layered tobermorite that gets more and more defective and hydrated as C/S increases locally inducing glassy atomic arrangements (Abdolhosseini Qomi et al. 2014a; Bauchy et al. 2014c). The transition between ordered and glassy was identified at C/S = 1.5 (Qomi et al. 2014a).

3.3   Anomalous Behavior of Confined Water in CSH

Despite the multidisciplinary interests in understanding water in confined geometries, our knowledge in the field of interfacial properties of water in such extreme confining conditions is still in a fledgling state. This is important as the confinement size and pore geometry directly affect the chemo-physical properties of water, e.g., a shift in dipole moment, hydrogen bonding, and transport properties – which are significantly different from those of bulk water. Equilibrium molecular dynamic (MD) simulations were performed on the 150+ models of CSH samples with the goal of understanding the structure and mobility of unreacted interlayer water molecules using the CSH-FF potential (Shahsavari et al. 2011) as it reproduces the structure and mechanics of CSH at varying stoichiometry. It was found that the mobility of water strongly depends on composition; namely, it increases with increasing C/S ratio by up to one order of magnitude (Fig. 9). However, over the whole range of C/S ratio values, the diffusivity of mobile water remains about 1/1000th of the bulk water. The increase in the probability of escaping the
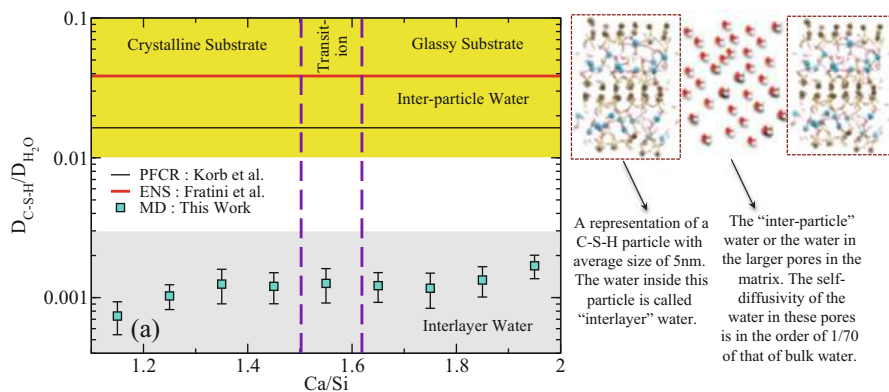
Fig. 9 The self-diffusivity of water in the CSH structure. The self-diffusivity of water in the interlayer of CSH increases with increasing Ca/Si ratio. The diffusion coefficient of the nanoconfined interlayer water is three orders of magnitude less than that of bulk water (Qomi et al. 2014). The nanoconfined self-diffusivity of water is 70 times larger than those measured for CSH by Korb et al. (2007a) and Fratini et al. (2013)

dynamical cage with increasing C/S ratio physically explains the composition-dependent mobility in relation to the amount of protruding silica groups in the interlayer spacing of CSH. This brings forward the notion of "pocket-like," "patch-like," and "in-channel" diffusion at low, medium, and high C/S ratios, respectively.

The strong interaction with the CSH hydrophilic surfaces and their sub-nanometric defective texture reduces the self-diffusivity of mobile water molecules. Furthermore, the self-diffusivity of water in the interlayer is roughly ten times smaller than that observed in elastic neutron scattering and proton field-cycling relaxometry experiments, which both only probe the dynamics of water in CSH's mesopores. Hence, these experiments are not yet addressing the water dynamics inside the nanotexture of CSH. Finally, a correlation was found between the structural and dynamical properties, which arise from the anomalous behavior of water under pressure. Similar to the anomalous behavior of bulk water, the self-diffusivity of water in the nanotexture of CSH increases with the density. The confined water in CSH was shown to have higher self-diffusivity due to an increase in the escaping probability from the dynamical cage through reduction of diffusion energy barriers.

The consistent picture that emerges from these studies is that molecular water confined in the nanotexture of CSH is really part of the structure and does not behave at all as liquid water: water molecules vibrate around almost fixed interlayer calcium ion and show very limited translational diffusion. This translates into an anomalous diffusive behavior and very small value of the associated dielectric constant, even affecting the water dipole moment. Spanning across the data bank of atomistic models of CSH, it became clear that molecular water goes through

a hopping mechanism (from interlayer calcium ion to the next) with increasing self-diffusivity constant with increasing C/S values. However, these diffusivity constants remain 1000 less than that of bulk water. Obviously, this almost non-diffusive water does not arrange as ice. Interestingly enough, Korb et al. (2007a, b) measured a self-diffusivity constant for water molecules at the surface of the CSH grains defining capillary pores that is 60 less than that of the bulk. Therefore, it does totally make sense to find even lower self-diffusivity constant for water inside CSH grains. Recently, Morshedifard et al. showed that the non-asymptotic long-term deformation of creep, i.e., the so-called logarithmic creep, originates from increasing nanoconfined water content between the CSH layers (Morshedifard et al. 2018). They demonstrated that the viscoelastic deformation in CSH transitions to logarithmic creep by increasing the lubrication effect on adjacent CSH layers. This places a critical emphasis on the role of nanoconfined water on the durability of cementitious materials.

3.4 Thermal Properties of Cement Paste from the Nano- to the Macroscale

Having a set of realistic structure of CSH, Qomi et al. studied the thermal properties of cement paste via a multiscale bottom-up approach, starting from the nanoscale by calculating properties of individual constituent phases and connecting them to macroscale properties using mean-field homogenization theories (Qomi et al. 2015). At the nanoscale, the phonon density of states of CSHs, CH, β-C_2S, and C_3S and the components' atomic contributions were studied, and it was found that analogous to glassy materials, CSH exhibits excess vibrational states characterized by the boson peak at the low-frequency region. The position of the boson peak shifts to higher frequencies with increasing CSH density or, equivalently, by decreasing the C/S ratio. The specific heat capacity of dry and saturated CSH models increases linearly with the C/S ratio; see Fig. 10a. The apparent heat capacity of nanoconfined water in the interlayer spacing of CSH is determined to be significantly lower than that of bulk water at room temperature due to strong interaction with the calcium-silicate substrate. The thermal-conductivity tensor individual phases using the Green-Kubo relation were calculated and shown to be close to that of silica glass. The cumulative model thermal conductivity measurements are close to that of molecular dynamics and those back-calculated from the experiment (Zhou et al. 2017). Due to the glassy nature of CSH, only 30% of the total heat conduction stems from the phonon-like vibrational modes called propagans. The rest of the 70% heat conduction is carried out via the diffusive vibrational modes; see Fig. 10b. Furthermore, it was demonstrated that the variation of the macroscopic heat capacity of cement paste in terms of the hydration degree using mixture rules. The macroscopic heat capacity and thermal conductivity values predicted via our multiscale bottom-up approach were shown to be in good quantitative agreement with experiments.
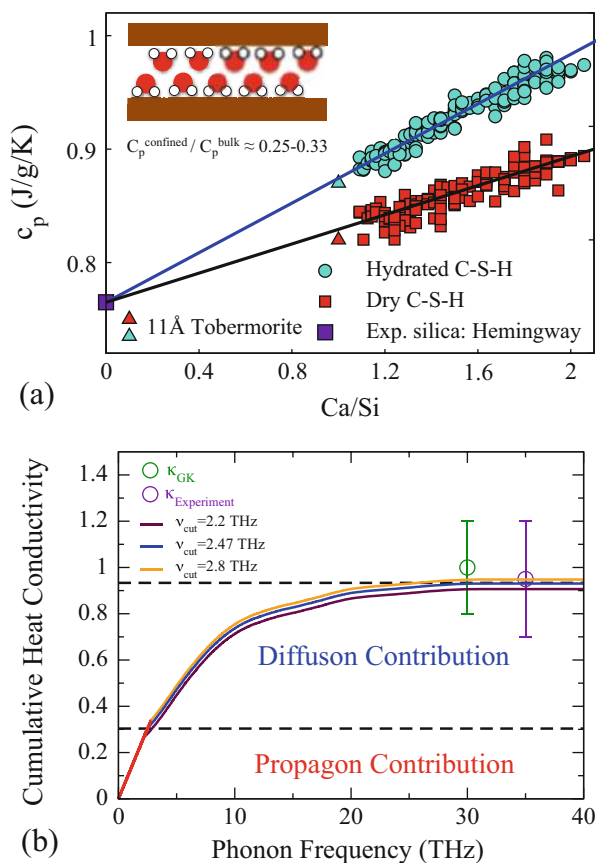
Fig. 10 Thermal properties of CSH. (**a**) The heat capacity of CSH increases linearly with Ca/Si ratio (Qomi et al. 2015). (**b**) The cumulative thermal conductivity of CSH at Ca/Si = 1.7, which is close to that of molecular dynamics and those back-calculated from the experiment (Zhou et al. 2017)

3.5   Topological Constraint Theory

Concrete's mechanical properties are mostly controlled by the mechanical properties of its binding "glue," the CSH phase. However, the complex atomic structure of CSH makes it challenging to reveal the relationship between CSH's composition and its nanoscale mechanical properties. This lack of knowledge has likely severely limited our potential to significantly improve the mechanical properties of CSH. This is an account of the fact that CSH can exhibit a virtually infinite number of possible compositions (e.g., with varying C/S ratio) and that even small variations in its composition can drastically alter the properties of CSH as shown here above.

To this end, recent progress in glass science has recently resulted in some new insights into the relationship between CSH's atomic composition/structure and its

engineering properties. This analogy with glass science is based on the observation that cementitious materials and structural glasses share many similarities: (i) they are largely disordered, (ii) they exhibit fairly similar compositions (e.g., SiO_2, Al_2O_3, CaO, Na_2O, etc.), (iii) they are out-of-equilibrium (i.e., they "relax" toward a more stable state of energy over time), and (iv) they can present a virtually infinite number of possible compositions (i.e., unlike crystals, they do not have to satisfy any stoichiometry rules). Over the past decade, significant efforts have been devoted by the glass industry to nano-engineer novel glasses featuring optimized properties, especially to stimulate the development of fracture- and scratch-resistant glasses used as smartphone protective screens (Mauro et al. 2013).

In particular, topological constraint theory (TCT) has emerged as a powerful technique to predict the engineering properties of glasses based on their composition (Bauchy 2012; Mauro 2011; Phillips 1979). The idea of TCT is to reduce the complex, disordered structure of glasses into simple mechanical trusses, wherein nodes (the atoms) are connected to each other by some mechanical constraints (the chemical bonds). In atomic networks, such constraints comprise (i) the radial bond-stretching constraints, which maintain the bond length fixed around their average values, and (ii) the angular bond-bending constraints – which govern the interatomic angles. As such, TCT captures the important atomic topology that has a first-order impact on their macroscopic properties while filtering out less relevant structural details (Smedskjaer et al. 2010). In analogy with the stability of mechanical trusses, atomic networks can then be classified as flexible, stressed-rigid, or isostatic when the network of topological constraints per atom (n_c) is smaller, larger, or equal to three (i.e., the number of degrees of freedom per atom in 3D), respectively (see Fig. 11). Importantly, TCT was successfully used to predict the optimal composition of Corning® Gorilla® glass, which is now used on billions of smartphones and tablets around the world (Ball 2015).

3.6 Composition-Property Relationships in CSH

Recently, for the first time, TCT was applied to CSH to rationalize the relationship between its composition and properties. The rigidity state of CSH systems of varying compositions was identified by reactive molecular dynamic simulations (Bauchy et al. 2014a, 2015) using a well-established method (Bauchy et al. 2011; Bauchy and Micoulaut 2011). CSH was found to be flexible ($n_c < 3$) and stressed-rigid ($n_c > 3$) for Ca-rich and Si-rich compositions, respectively (see Fig. 12). These regions effectively delimit a series of optimally constrained ($n_c = 3$) CSH samples (gray region in Fig. 12). The knowledge of the topology of the atomic network of CSH was used to predict its hardness, in the very same fashion as the hardness of aluminosilicate glasses was predicted while designing Gorilla® glass (Bauchy et al. 2015; Smedskjaer et al. 2010; Zheng et al. 2017). The hardness was calculated using the new approach proposed by Qomi et al. by applying biaxial deformation to the molecular structure (Qomi et al. 2017). It was found that the hardness of CSH scales linearly with the number of angular bond-bending
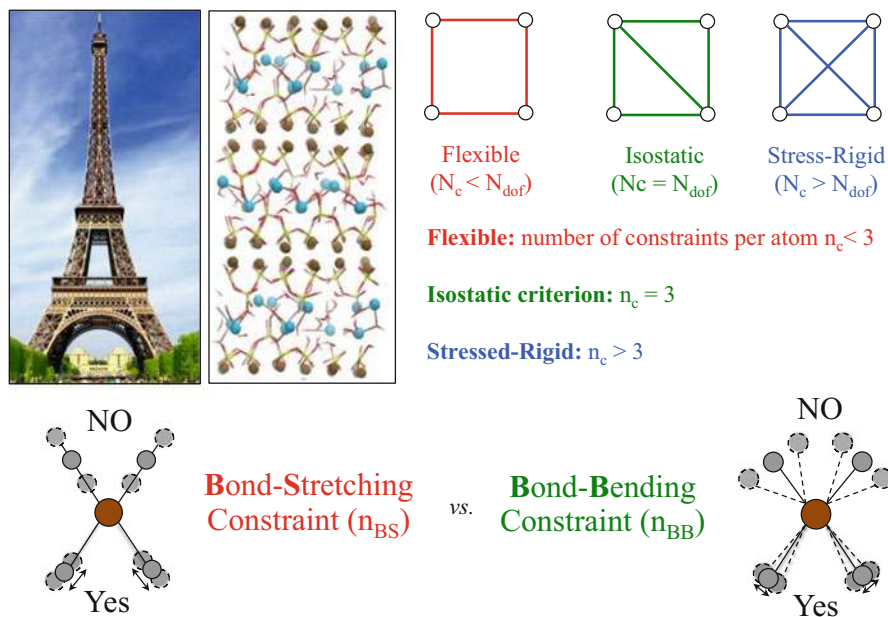
Fig. 11 The notion of topological constraint theory. In this approach, the molecular structure of CSH is analyzed to that of trusses in structural analysis. The three states of rigidity of a mechanical truss, i.e., flexible, isostatic, and stressed-rigid. We also show an illustration of bond-stretching and bond-bending constraints
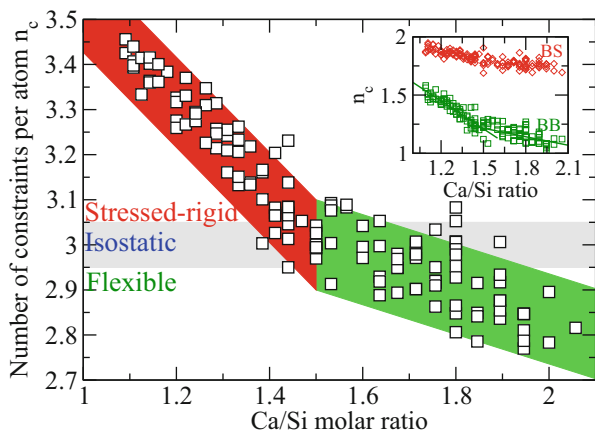
Fig. 12 Number of constraints per atom (n_c) in CSH as a function of the Ca/Si molar ratio. The gray area indicates the extent of the isostatic domain ($n_c = 3$). The inset shows the number of bond-stretching (BS) and bond-bending (BB) constraints per atom as a function of the Ca/Si molar ratio
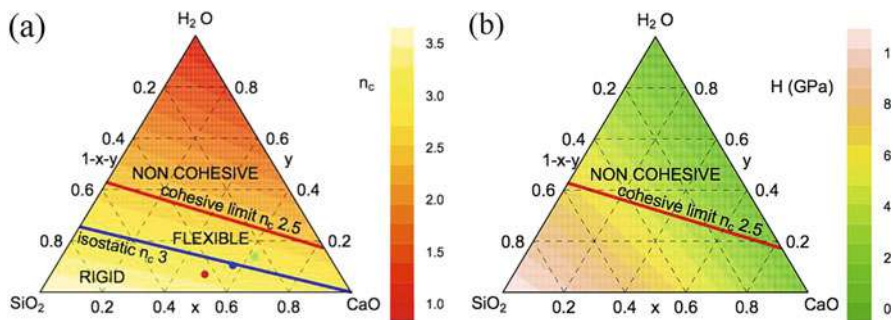
Fig. 13 Ternary diagram showing the compositional dependence of (**a**) the number of constraints per atom and (**b**) the indentation hardness

constraints per atom (Bauchy et al. 2015). Based on this knowledge, the ternary map of CSH hardness was predicted (see Fig. 13). All the predictions were validated by nanoindentation tests. This ternary map of hardness can be used to pinpoint promising CSH compositions exhibiting targeted hardness values (see Fig. 13). Altogether, by simplifying the complex atomic network of CSH, TCT can be used to analytically predict composition-property relationships in CSH. This approach can be used to identify in silico promising CSH compositions that feature desired engineering properties.

3.7   Optimal Properties of Isostatic CSH Compositions

When analyzing the 150+ models of CSH spanning the range of possible C/S values through the prism of the so-called rigidity theory, it was striking to realize that the C/S = 1.5 composition was the turning point from stressed-rigid (at C/S >1.5) to floppy/flexible configurations (at C/S <1.5), that is, the composition threshold at which CSH exhibits a crystalline-to-glassy transition (see above). Due to its isostatic character, the particular composition at C/S = 1.5 is rigid but free of any internal stress (Bauchy et al. 2015). Based on previous observations regarding isostatic glasses, these compositions are expected to show optimal properties. Importantly, it has been shown that isostatic glasses (i.e., $n_c = 3$) feature (i) an optimal space-filling tendency, (ii) a minimum propensity for aging, and (iii) a maximum resistance to fracture (Boolchand et al. 2001; Varshneya and Mauro 2007). This is significant as, despite obvious advantages (low cost, ease of use, etc.), concrete's main drawback lies in the fact that it can break or creep over time.

As mentioned above, CSH was found to be isostatic (optimally constrained, $n_c = 3$) for Ca/Si =1.5. This isostatic state is obtained when the atomic is fully rigid (i.e., no internal modes of deformation) but free of any eigenstress (i.e., no mutually dependent constraints). Interestingly, it was observed that isostatic CSH samples exhibit maximum fracture toughness (see Fig. 14) (Bauchy et al. 2016).

Fig. 14 Computed fracture toughness of CSH as a function of the number of constraints per atom. The gray area indicates the range of composition for which CSH shows an isostatic character
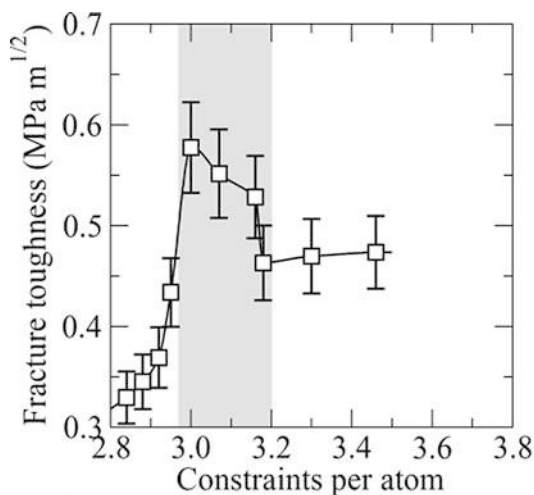
This was shown to arise from the facts that flexible CSH samples ($n_c < 3$) exhibit low surface energy values, which makes them poorly cohesive, whereas stressed-rigid CSH samples ($n_c > 3$) show high brittleness since the high atomic connectivity prevents any local reorganization of the atoms to release stress. In contrast, isostatic CSH samples ($n_c = 3$) present the best compromise between cohesion and ability to show local plastic deformations and, hence, feature maximum fracture toughness (in line with what is observed for structural glasses).

In addition, isostatic CSH samples were also found to exhibit maximum resistance to long-term relaxation deformations (Bauchy et al. 2017; Pignatelli et al. 2016). This was explained by the facts that flexible atomic networks present some internal modes of deformation, which facilitate the activation of low-energy modes of relaxation upon loading. On the contrary, stressed-rigid systems are free of such floppy modes but present some internal eigenstress, which acts as a driving force stimulating relaxation (i.e., to achieve a more stable configuration that is free of any internal stress). Isostatic CSH compositions were also found to exhibit unusual thermal properties, dissolution kinetics, and resistance to irradiation (Krishnan et al. 2016, 2017; Pignatelli et al. 2016). Altogether, these results support the idea that CSH might behaves as a conventional silicate glass.

4   Conclusion

Overall, these results offer a self-consistent description of concrete at different scales: (i) electronic scale (DFT simulations), (ii) atomic scale (classical and reactive molecular dynamic simulations), (iii) colloidal scale (mesoscale simulations), and (iv) continuum scale (i.e., the scale that is accessible to experiments). Various upscaling techniques (e.g., "potential of mean force") or theories (e.g., topological

constraint theory) can now be used to seamlessly connect one scale to the next – by capturing the essential feature of a given scale and modeling it as a reduced-dimensionality input for the next scale. This approach offers a complete bottom-up description of calcium-silicate-hydrate – from electrons to continuum – with no adjustable parameters. We expect these models to facilitate the optimization of existing cementitious binders and accelerate the discovery of novel formulations with drastically improved properties. Altogether, these results also exemplify how a synergetic combination of experiments, simulations, and theoretical models – which mutually feed into and advance each other – can unlock the "genome" of concrete, that is, by identifying and deciphering how the basic structural building blocks of its atomic structure control its macroscopic properties. This effort also largely benefited from advances offered by various fields, including material science, condensed matter, statistical physics, and chemistry. Such an interdisciplinary is critically needed to solve the "material problems" facing the civil engineering community – a field that had largely relied on empiricism-driven routes in the past.

References

Allen AJ, Oberthur RC, Pearson D, Schofield P, Wilding CR (1987) Development of the fine porosity and gel structure of hydrating cement systems. Philos Mag B 56(3):263–288

Allen AJ, Thomas JJ, Jennings HM (2007) Composition and density of nanoscale calcium-silicate-hydrate in cement. Nat Mater 6(4):311–316

Attard P (1996) Electrolytes and the electric double layer. In: Prigogine I, Rice SA (eds) Advances in chemical physics. Wiley, New York, pp 1–159

Ball P (2015) Material witness: concrete mixing for gorillas. Nat Mater 14(5):472–472

Bauchy M (2012) Topological constraints and rigidity of network glasses from molecular dynamics simulations. Am Ceram Soc Bull 91(4):34–38A

Bauchy M (2014) Structural, vibrational, and elastic properties of a calcium aluminosilicate glass from molecular dynamics simulations: the role of the potential. J Chem Phys 141(2):024507

Bauchy M, Micoulaut M (2011) Atomic scale foundation of temperature-dependent bonding constraints in network glasses and liquids. J Non-Cryst Solids 357(14):2530–2537

Bauchy M, Micoulaut M, Celino M, Le Roux S, Boero M, Massobrio C (2011) Angular rigidity in tetrahedral network glasses with changing composition. Phys Rev B 84(5):054201

Bauchy M, Qomi MJA, Bichara C, Ulm F-J, Pellenq RJ-M (2014a) Nanoscale structure of cement: viewpoint of rigidity theory. J Phys Chem C 118(23):12485–12493

Bauchy M, Qomi MJA, Pellenq RJ-M, Ulm FJ (2014b) Is cement a glassy material? Comput Model Concr Struct 1:169

Bauchy M, Qomi MJA, Ulm F-J, Pellenq RJ-M (2014c) Order and disorder in calcium-silicate-hydrate. J Chem Phys 140(21):214503

Bauchy M, Qomi MJA, Bichara C, Ulm F-J, Pellenq RJ-M (2015) Rigidity transition in materials: hardness is driven by weak atomic constraints. Phys Rev Lett 114(12):125502

Bauchy M, Wang B, Wang M, Yu Y, Qomi MJA, Smedskjaer MM, Bichara C, Ulm F-J, Pellenq R (2016) Fracture toughness anomalies: viewpoint of topological constraint theory. Acta Mater 121:234–239

Bauchy M, Wang M, Yu Y, Wang B, Krishnan NMA, Masoero E, Ulm F-J, Pellenq R (2017) Topological control on the structural relaxation of atomic networks under stress. Phys Rev Lett 119(3):035502

Bensted J, Barnes P (2002) Structure and performance of cements. Spon Press, London/New York

Bernal JD (1954) The structures of cement hydration compounds. In: Proceedings of the 3rd international symposium on the chemistry of cement, pp 216–236

Bonaccorsi E, Merlino S, Kampf AR (2005) The crystal structure of tobermorite 14 A (Plombierite), a C-S-H phase. J Am Ceram Soc 88(3):505–512

Bonnaud PA, Labbez C, Miura R, Suzuki A, Miyamoto N, Hatakeyama N, Miyamoto A, Vliet KJV (2016) Interaction grand potential between calcium-silicate-hydrate nanoparticles at the molecular level. Nanoscale 8(7):4160–4172

Boolchand P, Georgiev DG, Goodman B (2001) Discovery of the intermediate phase in chalcogenide glasses. J Optoelectron Adv Mater 3(3):703–720

Bullard JW, Enjolras E, George WL, Satterfield SG, Terrill JE (2010) A parallel reaction-transport model applied to cement hydration and microstructure development. Model Simul Mater Sci Eng 18(2):025007

Carrier B (2013) Influence of water on the short-term and long-term mechanical properties of swelling clays: experiments on self-supporting films and molecular simulations. PhD thesis, Université Paris-Est

Chien S-C, Auerbach SM, Monson PA (2015) Reactive ensemble Monte Carlo simulations of silica polymerization that yield zeolites and related crystalline microporous structures. J Phys Chem C 119(47):26628–26635

Churakov SV (2009) Structure of the interlayer in normal 11 Å tobermorite from an ab initio study. Eur J Mineral 21(1):261–271

Daimon M, Abo-El-Enein SA, Rosara G, Goto S, Kondo R (1977) Pore structure of calcium silicate hydrate in hydrated tricalcium silicate. J Am Ceram Soc 60(3–4):110–114

Dolado JS, Griebel M, Hamaekers J, Heber F (2011) The nano-branched structure of cementitious calcium-silicate-hydrate gel. J Mater Chem 21(12):4445–4449

Durgun E, Manzano H, Pellenq RJ-M, Grossman JC (2012) Understanding and controlling the reactivity of the calcium silicate phases from first principles. Chem Mater 24(7):1262–1267

Ebrahimi D, Pellenq RJ-M, Whittle AJ (2012) Nanoscale elastic properties of montmorillonite upon water adsorption. Langmuir 28(49):16855–16863

Ebrahimi D, Whittle AJ, Pellenq RJ-M (2014) Mesoscale properties of clay aggregates from potential of mean force representation of interactions between nanoplatelets. J Chem Phys 140(15):154309

Feldman RF, Sereda PJ (1968) A model for hydrated Portland cement paste as deduced from sorption-length change and mechanical properties. Mater Constr 1(6):509–520

Fratini E, Faraone A, Radi F, Chen SH, Baglioni P (2013) Hydration water dynamics in tricalcium silicate pastes by time-resolved incoherent elastic neutron scattering. J Phys Chem C 117(14):7358–7364

Fonseca PC, Jennings HM (2010) The effect of drying on early-age morphology of C-S-H as observed in environmental SEM. Cem Concr Res 40(12):1673–1680

Garrault S, Finot E, Lesniewska E, Nonat A (2005) Study of C-S-H growth on C_3S surface during its early hydration. Mater Struct 38(4):435–442

Garrault S, Behr T, Nonat A (2006) Formation of the C−S−H layer during early hydration of tricalcium silicate grains with different sizes. J Phys Chem B 110(1):270–275

Gartner EM (1997) A proposed mechanism for the growth of C-S-H during the hydration of tricalcium silicate. Cem Concr Res 27(5):665–672

Gartner E (2004) Industrially interesting approaches to 'low-CO_2' cements. Cem Concr Res 34(9):1489–1498. H F W Taylor Commemorative Issue

Gartner E, Sui T (2018) Alternative cement clinkers. Cem Concr Res 114:27–39. https://doi.org/10.1016/j.cemconres.2017.02.002

Gartner E, Maruyama I, Chen J (2017) A new model for the C-S-H phase formed during the hydration of Portland cements. Cem Concr Res 97:95–106

Geng G, Myers RJ, Qomi MJA, Monteiro PJM (2017) Densification of the interlayer spacing governs the nanomechanical properties of calcium-silicate-hydrate. Sci Rep 7(1):10986

Gmira A, Zabat M, Pellenq RJ-M, Damme HV (2004) Microscopic physical basis of the poromechanical behavior of cement-based materials. Mater Struct 37(1):3–14

Huang J, Wang B, Valenzano L, Bauchy M, Sant G (2015) Electronic origin of doping-induced enhancements of reactivity: case study of tricalcium silicate. J Phys Chem C 119:25991

Ioannidou K, Pellenq RJ-M, Del Gado E (2014) Controlling local packing and growth in calcium-silicate-hydrate gels. Soft Matter 10(8):1121–1133

Ioannidou K, Krakowiak KJ, Bauchy M, Hoover CG, Masoero E, Yip S, Ulm F-J, Levitz P, Pellenq RJ-M, Del Gado E (2016) Mesoscale texture of cement hydrates. Proc Natl Acad Sci 113(8): 2029–2034

Israelachvili JN (2011) Intermolecular and surface forces, Revised 3rd edn. Academic, Cambridge, MA

Jennings HM (2000) A model for the microstructure of calcium silicate hydrate in cement paste. Cem Concr Res 30(1):101–116

Jönsson B, Wennerström H, Nonat A, Cabane B (2004) Onset of cohesion in cement paste. Langmuir 20(16):6702–6709

Jönsson B, Nonat A, Labbez C, Cabane B, Wennerström H (2005) Controlling the cohesion of cement paste. Langmuir 21(20):9211–9221

Korb J-P, McDonald PJ, Monteilhet L, Kalinichev AG, Kirkpatrick RJ (2007a) Comparison of proton field-cycling relaxometry and molecular dynamics simulations for proton-water surface dynamics in cement-based materials. Cem Concr Res 37(3):348–350

Korb J-P, Monteilhet L, McDonald PJ, Mitchell J (2007b) Microstructure and texture of hydrated cement-based materials: a proton field cycling relaxometry approach. Cem Concr Res 37(3):295–302

Krishnan NMA, Wang B, Falzone G, Le Pape Y, Neithalath N, Pilon L, Bauchy M, Sant G (2016) Confined water in layered silicates: the origin of anomalous thermal expansion behavior in calcium-silicate-hydrates. ACS Appl Mater Interfaces 8(51):35621–35627

Krishnan NMA, Wang B, Sant G, Phillips JC, Bauchy M (2017) Revealing the effect of irradiation on cement hydrates: evidence of a topological self-organization. ACS Appl Mater Interfaces 9(37):32377–32385

Manzano H, Durgun E, Qomi MJA, Ulm F-J, Pellenq RJ, Grossman JC (2011) Impact of chemical impurities on the crystalline cement clinker phases determined by atomistic simulations. Cryst Growth Des 11(7):2964–2972

Manzano H, Masoero E, Lopez-Arbeloa I, Jennings HM (2013) Shear deformations in calcium silicate hydrates. Soft Matter 9(30):7333–7341

Manzano H, Durgun E, López-Arbeloa I, Grossman JC (2015) Insight on tricalcium silicate hydration and dissolution mechanism from molecular simulations. ACS Appl Mater Interfaces 7(27):14726–14733

Masoumi S, Valipour H, Qomi MJA (2017a) Intermolecular forces between nanolayers of crystalline calcium-silicate-hydrates in aqueous medium. J Phys Chem C 121(10):5565–5572

Masoumi S, Valipour H, Qomi MJA (2017b) Interparticle interactions in colloidal systems: towards a comprehensive mesoscale model. ACS Appl Mater Interfaces 9(32):27338–27349

Mauro JC (2011) Topological constraint theory of glass. Am Ceram Soc Bull 90(4):31–37

Mauro JC, Ellison AJ, Pye LD (2013) Glass: the nanotechnology connection. Int J Appl Glas Sci 4(2):64–75

McDonald PJ, Rodin V, Valori A (2010) Characterisation of intra- and inter-C-S-H gel pore water in white cement based on an analysis of NMR signal amplitudes as a function of water content. Cem Concr Res 40(12):1656–1663

Meral C, Benmore CJ, Monteiro PJM (2011) The study of disorder and nanocrystallinity in C-S-H, supplementary cementitious materials and geopolymers using pair distribution function analysis. Cem Concr Res 41(7):696–710

Morshedifard A, Masoumi S, Qomi MJA (2018) Nanoscale origins of creep in calcium silicate hydrates. Nat Commun 9(1):1785

Muller ACA, Scrivener KL, Gajewicz AM, McDonald PJ (2013a) Densification of C-S-H measured by ^{1}H NMR relaxometry. J Phys Chem C 117(1):403–412

Muller ACA, Scrivener KL, Gajewicz AM, McDonald PJ (2013b) Use of bench-top NMR to measure the density, composition and desorption isotherm of C-S-H in cement paste. Microporous Mesoporous Mater 178:99–103

Neubauer CM, Jennings HM (1996) The role of the environmental scanning electron microscope in the investigation of cement-based materials. Scanning 18(7):515–521

Nonat A (2004) The structure and stoichiometry of C-S-H. Cem Concr Res 34(9):1521–1528. H. F. W. Taylor Commemorative Issue

Pellenq RJ-M, Van Damme H (2004) Why does concrete set?: the nature of cohesion forces in hardened cement-based materials. MRS Bull 29(05):319–323

Pellenq RJ-M, Caillol JM, Delville A (1997) Electrostatic attraction between two charged surfaces: a (N,V,T) Monte Carlo simulation. J Phys Chem B 101(42):8584–8594

Pellenq RJ-M, Lequeux N, van Damme H (2008) Engineering the bonding scheme in C-S-H: the iono-covalent framework. Cem Concr Res 38(2):159–174. Special issue – The 12th International Congress on the Chemistry of Cement, Montreal, 8–13 July 2007

Pellenq RJ-M, Kushima A, Shahsavari R, Vliet KJV, Buehler MJ, Yip S, Ulm F-J (2009) A realistic molecular model of cement hydrates. Proc Natl Acad Sci 106(38):16102–16107

Phillips JC (1979) Topology of covalent non-crystalline solids. 1. Short-range order in chalcogenide alloys. J Non-Cryst Solids 34(2):153–181

Pignatelli I, Kumar A, Alizadeh R, Pape YL, Bauchy M, Sant G (2016) A dissolution-precipitation mechanism is at the origin of concrete creep in moist environments. J Chem Phys 145(5):054701

Plassard C, Lesniewska E, Pochard I, Nonat A (2005) Nanoscale experimental investigation of particle interactions at the origin of the cohesion of cement. Langmuir 21(16):7263–7270

Pochard I, Labbez C, Nonat A, Vija H, Jönsson B (2010) The effect of polycations on early cement paste. Cem Concr Res 40(10):1488–1494

Popescu CD, Muntean M, Sharp JH (2003) Industrial trial production of low energy belite cement. Cem Concr Compos 25(7):689–693

Pustovgar E, Sangodkar RP, Andreev AS, Palacios M, Chmelka BF, Flatt RJ, d'Espinose de Lacaillerie J-B (2016) Understanding silicate hydration from quantitative analyses of hydrating tricalcium silicates. Nat Commun 7:10952

Qomi MJA, Ulm F-J, Pellenq RJ-M (2012) Evidence on the dual nature of aluminum in the calcium-silicate-hydrates based on atomistic simulations. J Am Ceram Soc 95(3):1128–1137

Qomi MJA, Bauchy M, Ulm F-J, Pellenq RJ-M (2014a) Anomalous composition-dependent dynamics of nanoconfined water in the interlayer of disordered calcium-silicates. J Chem Phys 140(5):054515

Qomi MJA, Krakowiak KJ, Bauchy M, Stewart KL, Shahsavari R, Jagannathan D, Brommer DB, Baronnet A, Buehler MJ, Yip S, Ulm F-J, Van Vliet KJ, Pellenq RJ-M (2014b) Combinatorial molecular optimization of cement hydrates. Nat Commun 5:4960

Qomi MJA, Ulm F-J, Pellenq RJ-M (2015) Physical origins of thermal properties of cement paste. Phys Rev Appl 3(6):064010

Qomi MJA, Bauchy M, Ulm F-J, Pellenq R (2015) Polymorphism and its implications on structure-property correlation in calcium-silicate-hydrates. In: Sobolev K, Shah SP (eds) Nanotechnology in construction. Springer International Publishing, Cham, pp 99–108

Qomi MJA, Ebrahimi D, Bauchy M, Pellenq R, Ulm F-J (2017) Methodology for estimation of nanoscale hardness via atomistic simulations. J Nanomech Micromech 7(4):04017011

Quillin K (2001) Performance of belite-sulfoaluminate cements. Cem Concr Res 31(9):1341–1349

Rahimi-Aghdam S, Bažant ZP, Qomi MJA (2017) Cement hydration from hours to centuries controlled by diffusion through barrier shells of C-S-H. J Mech Phys Solids 99:211–224

Richardson IG (2000) The nature of the hydration products in hardened cement pastes. Cem Concr Compos 22(2):97–113

Richardson IG (2004) Tobermorite/jennite- and tobermorite/calcium hydroxide-based models for the structure of C-S-H: applicability to hardened pastes of tricalcium silicate, β-dicalcium silicate, Portland cement, and blends of Portland cement with blast-furnace slag, metakaolin, or silica fume. Cem Concr Res 34(9):1733–1777

Richardson IG (2008) The calcium silicate hydrates. Cem Concr Res 38(2):137–158

Santos RL, Horta RB, Pereira J, Nunes TG, Rocha P, Canongia Lopes JN, Colaço R (2015) Microstructural control and hydration of novel micro-dendritic clinkers with CaO/SiO$_2$=1.4. Cem Concr Res 76:212–221

Scrivener KL, Juilland P, Monteiro PJM (2015) Advances in understanding hydration of Portland cement. Cem Concr Res 78:38–56. Keynote papers from 14th International Congress on the Chemistry of Cement (ICCC 2015)

Shahsavari R, Pellenq RJ-M, Ulm F-J (2011) Empirical force fields for complex hydrated calcio-silicate layered materials. Phys Chem Chem Phys 13(3):1002–1011

Skinner LB, Chae SR, Benmore CJ, Wenk HR, Monteiro PJM (2010) Nanostructure of calcium silicate hydrates in cements. Phys Rev Lett 104(19):195502

Smedskjaer MM, Mauro JC, Yue Y (2010) Prediction of glass hardness using temperature-dependent constraint theory. Phys Rev Lett 105(11):115503

Soyer-Uzun S, Chae SR, Benmore CJ, Wenk H-R, Monteiro PJM (2012) Compositional evolution of calcium silicate hydrate (C-S-H) structures by total X-ray scattering. J Am Ceram Soc 95(2):793–798

Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314(1):141–151

Taylor HFW (1993) Nanostructure of C-S-H: current status. Adv Cem Based Mater 1(1):38–46

Taylor HFW (1997) Cement chemistry. Thomas Telford, London

Thomas JJ, Jennings HM (2006) A colloidal interpretation of chemical aging of the C-S-H gel and its effects on the properties of cement paste. Cem Concr Res 36(1):30–38

Thomas JJ, Biernacki JJ, Bullard JW, Bishnoi S, Dolado JS, Scherer GW, Luttge A (2011) Modeling and simulation of cement hydration kinetics and microstructure development. Cem Concr Res 41(12):1257–1278. Conferences Special: Cement Hydration Kinetics and Modeling, Quebec City, 2009 & CONMOD10, Lausanne, 2010

Varshneya AK, Mauro DJ (2007) Microhardness, indentation toughness, elasticity, plasticity, and brittleness of Ge-Sb-Se chalcogenide glasses. J Non-Cryst Solids 353(13–15):1291–1297

Zheng Q, Yue Y, Mauro JC (2017) Density of topological constraints as a metric for predicting glass hardness. Appl Phys Lett 111(1):011907

Zhou Y, Morshedifard A, Lee J, Qomi MJA (2017) The contribution of propagons and diffusons in heat transport through calcium-silicate-hydrates. Appl Phys Lett 110(4):043104

Emanuela Del Gado, Kirsten Martens, and Roland J.-M. Pellenq

Contents

E. Del Gado (✉)
Department of Physics, Institute for Soft Matter Synthesis and Metrology, Georgetown University, Washington, DC, USA
Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA, USA
e-mail: emanuela.del.gado@georgetown.edu; ed610@georgetown.edu

K. Martens (✉)
Université Grenoble Alpes, CNRS, LIPhy, Grenoble, France
e-mail: kirsten.martens@univ-grenoble-alpes.fr

R. J.-M. Pellenq (✉)
Concrete Sustainability Hub, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

MSE2, the MIT/CNRS/Aix-Marseille University Joint Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Centre Interdisciplinaire de Nanosciences de Marseille, CINaM, CNRS/Aix-Marseille Université, Marseille, France

UMI MSE, CNRS/MIT/AMU Joint Laboratory "MultiScale Materials Science for Energy and Environment", Marseille, France
e-mail: pellenq@mit.edu

Abstract

Translating the insight gained through numerical simulations of microscopic models into information and predictions useful for engineering applications is a formidable task for complex and amorphous materials. Here we propose a strategy that builds on distinct dedicated approaches at different length scales and transfers the relevant information across them, to bridge from an atomistic description to a mesoscale mechanical model. We provide an overview of the fundamental concepts that underlie such strategy and two examples for engineering-relevant materials such as clays and cement hydrates.

1   Introduction

Novel breakthroughs in engineering materials depend crucially on understanding the microscopic physical processes underlying the development of amorphous structures and the related complex mechanical response. Being able to explain material properties in terms of their atomic structure is the ultimate goal for a mature branch of materials science, but it is far from being reached for amorphous materials: the mechanical properties of crystalline alloys are well understood in terms of microscopic heterogeneities in their structure – dislocations, grain boundaries, and interstitials – but an equivalent picture is lacking for amorphous materials. Translating the insight gained through microscopic models into information and predictions useful for engineering applications is therefore a formidable task. The idea that bridging the wide range of length scales involved can be achieved just by extending first-principle calculations over increasingly large samples, thanks to the continuum increase in power and capacity of computational resources, is far too simplistic and set to fail. The conundrum, in fact, is not to compute atomistic trajectories but rather to identify and disentangle distinct dynamical and mechanical processes that emerge at different length scales, depending on the complexity of the material structure. Solving this problem requires constructing models and analysis tools that unravel the fundamental physical mechanisms at the different relevant lengthscales, and devise the right strategy to pass the information across scales. For engineering applications, the ultimate challenge is to use the knowledge obtained at the microscopic level to formulate mesoscopic constitutive models that allow to predict the material behavior at engineering-relevant scales.

2   From Atomistic Approaches to the Potential of Mean Force: The Case of Cement Hydrates and Clays

Clay-based soils and cementitious materials have enormous practical relevance and implications for sustainable industrial technologies. The mechanics of clay-hydrates and calcium-silicate-hydrates (the main binding phase in cement), together with its evolution with changes in the physicochemical environment, is one of

the outstanding questions to control and design durability and sustainability of materials in civil engineering. Even at the atomistic level, those solids are extremely complex and require sophisticated models to capture their subnanoscale structure (Pellenq et al. 1997, 2009). In particular, the mechanics emerges from the mesoscale organization of nanoscale (colloidal) units, whose cohesion, in turn, is the results of subnanoscale effective interactions. Hence the most useful information to be extracted from the atomistic description of the material is the cohesion between the nanoscale units, which can be then translated into effective interactions for a model that describes their aggregation and phase behavior. From the atomic structure of a specific material, one can in fact derive the free energy function representing the effective interactions over a given length scale (the *Potential of Mean Force* or PMF). The PMF is the key piece of information needed in a coarse-grained description of the physical chemistry of a material and offers huge computational and modeling advantages as it allows for the upscaling of models and computational approaches with the possibility to retain the essential physicochemical processes of interest (Doi 2013; Frenkel and Smit 2001; Kumar et al. 1992; Pellenq et al. 1997; Bonnaud et al. 2016; Ioannidou et al. 2017a).

For both clay-hydrates and calcium-silicate-hydrates (at early stages of the hydration, where ions and small aggregates are still quite mobile in solution), the cohesion between the nanoscale units can be rationalized as the result of the net interactions between charged surfaces separated by a solvent containing counterions (Carrier 2013). This problem has been solved analytically, in certain conditions, by the DLVO theory that describes the combination of the repulsive electrical double-layer interaction (due to the accumulation of the counterions close to the charged surfaces) with the attractive van der Waals interactions (Israelachvili 1992). Within the DLVO theory, the analytical form of the repulsive electrostatic interactions is obtained by solving the Poisson-Boltzmann equation for a system made of positive point charges between negatively charged surfaces with the well-framed classical density functional theory approach. Designed for point charges, it is a mean-field approach that ignores the correlations in the fluctuations of the ionic positions. Making use of the Poisson-Boltzmann equation, the electrostatic pressure P (i.e., the force per unit area, or the derivative of the PMF with respect to the distance) between two charged surfaces as a function of the surface separation D is given by

$$P(D) = k_B T [C_{\texttt{ion}}(D/2) - C_{\texttt{ion}}(\infty)] \tag{1}$$

where k_B is the Boltzmann constant, T is the temperature, and $C_{\texttt{ion}}(D/2)$ is the cation concentration at the mid-separation. The PMF is simply the integral form of the pressure curve (typically integrated along the direction orthogonal to the two surfaces). This pressure depends only on the increase of the ionic concentration in the gap between the two surfaces, and its origin is the entropic contribution of an ideal gas of ions to the total stress (Israelachvili 1992). However, counterion fluctuations between charged surfaces can be affected by strong correlations that provide non-negligible contributions to the stresses on the surfaces: these are the so-called ionic correlation forces (ICFs) which can be strongly attractive even between
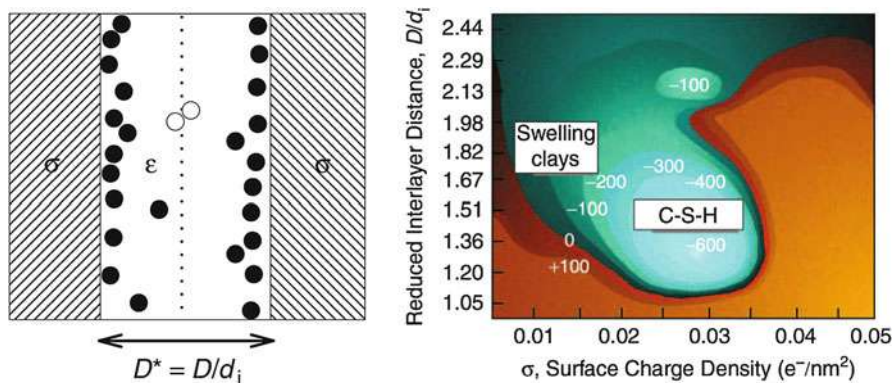
Fig. 1 Schematics and results of the primitive model (PM) from (Pellenq and Van Damme 2004). Left: Scheme showing the two uniformly charged walls separated by a dielectric continuum in which hydrated counterions are free to move. D is the interlayer separation and d is the ion diameter. Right: Isobaric contour map obtained from Monte Carlo simulations of the PM. The map shows both the cases for clays and C-S-H. Reproduced with permission from (Pellenq and Van Damme 2004).

equally charged surfaces and are not accounted for by DLVO theory (Pellenq et al. 1997, 2008; Pellenq and Van Damme 2004). ICFs can be computed numerically: in the primitive model (PM) introduced in the 1980s, the charged plane surfaces and the hydrated cations are represented by uniformly charged, infinite, and structureless planar surfaces and by charged hard spheres, respectively, while the solvent is represented by a dielectric continuum (see Fig. 1(a)). Monte Carlo simulations of PM have shown that fluctuations of the ionic concentration in the interlayer space result in an instantaneous polarization and can give rise to attractive ICFs. In particular, ICFs can become strongly attractive for divalent or trivalent cations, since in those cases, the attractive energetic contribution can exceed the repulsive entropic fluctuations by much. In the case of calcium counterions, present both in clays and calcium-silicate-hydrates, strong attractive pressures can be obtained at separations <10 Å (see Fig. 1(b)).

The primitive model has obviously its own limitations. The surfaces of the slit pore are assumed to be flat and uniformly charged; the ions are supposed to be fully hydrated and mobile; and the solvent is considered to be a dielectric continuum (typically the dielectric constant of bulk water would be used for clay or C–S–H). These hypotheses make the primitive model questionable in particular at small surface separations. In this case, the molecular structures of the solvent and the surfaces need to be taken into account to properly describe the emerging interactions. For weak electrostatic coupling conditions, corresponding, for instance, to sodium-based clays, it has been shown that the modification of PM to include the solvent explicitly (rather than through the bulk dielectric properties) leads to significant changes with respect to earlier findings, since the distribution of the ionic cloud of counterions cannot be properly described by the

Poisson-Boltzmann equation (Carrier 2013; Israelachvili 1992; Pellenq et al. 1997, 2008; Pellenq and Van Damme 2004). In the case of calcium-silicate-hydrates, to compensate the strong surface charge of C–S–H surfaces – on the order of one electron charge per 0.35 nm^2 for fully ionized C–S–H – Ca^{2+} counterions tend to accumulate even more in the vicinity of the surface. The two characteristics of strongly charged surfaces and divalent counterions make the interactions between C–S–H nanoscale units qualitatively different from clays and other weakly (or less strongly) interacting colloidal systems (Pellenq and Van Damme 2004). Overall, the basic picture that emerges from taking into account ionic positional correlations (with possibly an explicit description of the solvent if one wants to predict colloid cohesion quantitatively) is that the excess of counterions near a charged surface may be correlated with a deficit of counterions in the opposite same-charge surface, giving rise to net attractive ion-ion correlation forces. Such correlated density fluctuations of the ions lead to an attractive force in the same way as correlations between instantaneous electronic dipoles (due to the quantum fluctuations of the electron density around two argon atoms) give rise to London dispersion forces (Israelachvili 1992).

The possible attraction between two similarly charged surfaces immersed in an electrolyte has been explored within the theory of equilibrium electric double layer, where two regimes can be distinguished (Attard 1996): an attractive regime that occurs at small separation distances and in the presence of multivalent ions, where the forces change from repulsive to attractive as the surface charge and pH increase, and a second regime occurring at larger separations, where especially in case of multivalent ions the force can be alternately attractive or repulsive depending on the electrostatic coupling. The result is an effective interaction potential that can feature multiple minima separated by repulsive barriers. In the case of clays, the PMF can exhibit up to 3 minima corresponding to different water contents (Ebrahimi et al. 2012). When used in coarse-grained simulations of clay-platelets (Na-laponite) and parametrized with a Gay-Berne potential, the PMF obtained from atomistic calculations of the free energy of two Na-smectite clay-platelets in liquid water provided crucial information on the morphology of mesoscale aggregates and on mechanical properties of large particle assemblies, in good agreement with measurements in conventional 1D laboratory compression tests and nano-indentation measurements on natural clays and shale samples (Ebrahimi et al. 2014). In the case of C–S–H, the presence of a deep minimum and a repulsive barrier in the PMF, whose (respectively) depth and height change with increasing pH during cement hydration, was suggested by experiments (Garrault et al. 2005; Lesko et al. 2001) and supported by recent atomistic calculations (Bonnaud et al. 2016). Using such information from the PMF in a non-equilibrium molecular dynamics approach has allowed for novel insight into gelation and densification of cement hydrates (Ioannidou et al. 2014; Del Gado et al. 2014): the evolution of the PMF could be, in fact, responsible for the morphological changes of cement gels during hydration that are instrumental to reach the high density and mechanical strength needed for their technological function (Ioannidou et al. 2016a). Furthermore, the approach devised for C–S–H and based on PMF has proven able to provide quantitative

new understanding of mechanical and hygrothermal properties of cement gels, rationalizing experimental findings so far considered contradictory and opening the way to advanced material design (Ioannidou et al. 2016b, 2017b).

As stated above, the PMF-based approaches just described provided new opportunities to bridge the mesoscale features of those complex amorphous materials with the engineering-scale description. Recently, PMF-based lattice approaches have also been proposed, to capture the effective mechanical behavior of heterogeneous materials, using harmonic interactions inspired from beam-type stretch and bending interactions between a discrete set of mass points (Laubie et al. 2017). While traditional engineering approaches defined those interactions employing standard mechanical elements (Dormieux et al. 2006), such lattice approaches have clearly more to offer when set in the context of effective potentials. Closed-form calibration procedures for such interaction potentials can be derived for both the isotropic and the transverse isotropic elastic cases on cubic lattices, in the form of linear relations between effective elasticity properties and energy parameters that define the interactions. This framework provides a means to set methods based on lattice discretization of heterogeneous materials onto a path toward modeling the actual physics interactions between phases, within the limits defined by the geometric discretization of regular or irregular lattices.

3   Coarse-Grained Models and Microscopic Insight of Dynamical Processes

The effective interactions discussed above in terms of the PMF constitute the essential basic ingredients to investigate structure formation and kinetic arrest occurring in the amorphous solidification of materials on length scales much larger than the atomistic one. Those effective interactions can provide the basis to investigate the emerging mechanics and the rheology of such materials, using coarse-grained microscopic models and simulations that include fundamental features of specific materials (or class of materials) while simplifying the atomistic description. Such models and simulations allow for the large-scale numerical studies needed to unravel the connection between nanoscale structure and dynamics and the response on much larger length scales and timescales, typically relevant for engineering applications.

A coarse-grained microscopic model usually describes the material in terms of the nanoscale units (particles), their positions $\{\mathbf{r}_i\}$, and their collective interaction energy $U(\{\mathbf{r}_i\})$ (i.e., the result of the PMF). Such model can be used in molecular dynamics (MD) simulations, by integrating Newton, Brownian, or damped equations of motion to compute particle trajectories and velocities, forces and stresses, as well as the response to an externally imposed stress or deformation rate (Frenkel and Smit 2001; Allen and Tildesley 1987; Morriss and Evans 2013). Different types of degrees of freedom can be included in this type of description to account for the local anisotropy of the structural domains (or particle shapes) and of the interactions. They can also be added successively allowing for studies in which one can gradually investigate different degrees of complexity in the material description. All material-specific information are contained in the shape of $U(\{\mathbf{r}_i\})$ and in the
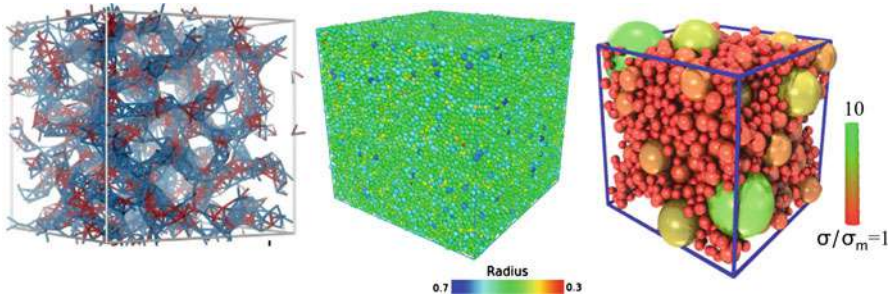
Fig. 2 Examples of typical microstructures in coarse-grained models for amorphous materials. Left: porous gel structure due to competing short-range attraction and long-range repulsion. Reproduced with permission from (Ioannidou et al. 2016a). Creative Commons license available at https://creativecommons.org/licenses/by/4.0/. Copyright 2017 Nature Publishing Group). Only bonds between particles are shown for clarity, and the color indicates the local coordination number. Center: soft repulsive particle assembly as a model for compressed emulsions, where the color indicates the particle size. Reproduced with permission from (Vasisht et al. 2018). Right: heterogeneous dense amorphous solid modeled in terms of repulsive soft particles, where the local strength of the effective interactions depends on the particle size. Reproduced with permission from (Masoero et al. 2012).

specific type of equations of motion that are chosen for a prescribed problem. The relative simplicity of such schemes allow for tackling the intrinsically many-body nature of the emerging mechanics and for using statistical mechanics to develop a bottom-up description of the macroscopic material properties.

Even when modeling the material simply as an assembly of point-like particles with U depending only on their relative distances, the variety of shapes of the effective interactions that emerge from surface forces and electrostatics between colloidal units in solution allows for a relatively wide variety of equilibrium and non-equilibrium microstructures. Hence, a variety of microstructural and mechanical heterogeneities in complex materials. Figure 2 shows some examples from different numerical studies, using effective interactions that can be derived from the PMF approach described in Sect. 2. The examples range from complex porous gel structures that result from competing short-range attraction and longer-range electrostatic repulsion (de Candia et al. 2006; Zhuang et al. 2016; Ioannidou et al. 2014, 2016a) to more spatially homogeneous amorphous assemblies of repulsive particles, where the size polydispersity allows for introducing structural disorder and the pair interaction strengths can be made size-dependent to account for local mechanical heterogeneities in the microstructure (Masoero et al. 2012, 2014).

3.1 Spatiotemporal Correlations and Non-affine Motion

From the single-particle trajectories obtained in the simulations just described, one can compute the spatiotemporal fluctuations of the microscopic dynamics, to quantify cooperative processes, and of the microscopic stresses, to analyze the role of related heterogeneities in the overall mechanical response. A space-
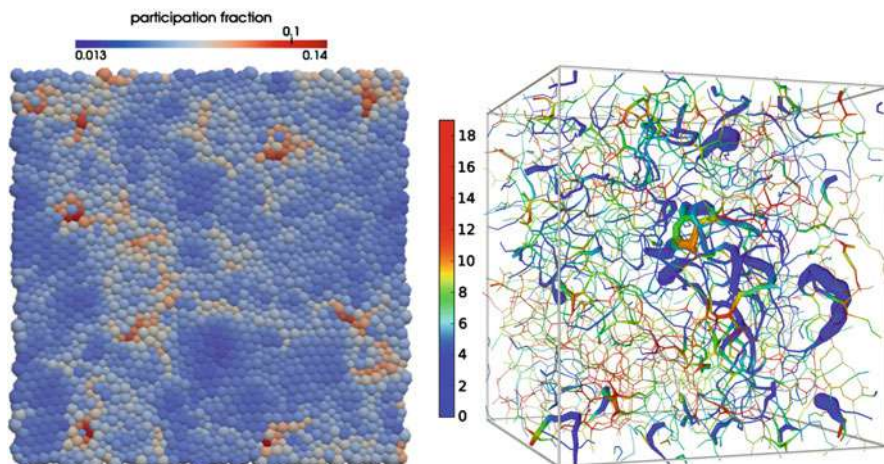
Fig. 3 Spatiotemporal correlations of microscopic dynamics. Left: maps of dynamical hetero-geneities in a 2D amorphous solid modeled using soft repulsive Lennard-Jones potential and obtained through an isoconfigurational analysis. Reproduced with permission from (Mosayebi et al. 2014). Right: Spatial map of the dynamical heterogeneities obtained through an isoconfigu-rational analysis suitably designed for a model colloidal gel. The segments represent interparticle bonds, and their thickness is proportional to the propensity for displacement of the bonded particles. The color code represents the local density of nodes in the gel. Reproduced with permission from (Colombo et al. 2013)

resolved analysis of the cooperativity of the microscopic dynamics can be obtained through an isoconfigurational analysis of particle trajectories (Widmer-Cooper and Harrowell 2007; Widmer-Cooper et al. 2009). Such approach, initially introduced to investigate dynamical heterogeneities in dense supercooled liquids, consists in generating an ensemble of MD trajectories from the same initial microscopic con-figuration and different initial velocities drawn from the same Maxwell-Boltzmann distribution. The distribution of particle displacements over this ensemble, at a fixed time, allows us to detect which particles and which parts of the structure are more likely to undergo the most significant displacements, and hence to be responsible for the overall structural relaxation. Figure 3 (left) shows an example of the spatial distribution of the dynamical heterogeneity in a 2D glassy alloy, modeled as a binary mixture of disks interacting through a purely repulsive soft effective potential and obtained with this method (Mosayebi et al. 2014; Widmer-Cooper and Harrowell 2006; Widmer-Cooper et al. 2008). Here, the dynamical heterogeneity is quantified by calculating the local Debye-Waller factors (DWF) of each particle. The DWF of the ith particle is defined as the isoconfigurational average of the variance of its position during a fixed time interval t^*, $DWF_i = \langle\langle(\mathbf{r}_i(t) - \langle\mathbf{r}_i\rangle_{t^*})^2\rangle_{t^*}\rangle_{\mathrm{iso}}$. Such analysis can be extended to low density, soft materials such as particle gels, but needs to be suitably modified. In the model colloidal gel shown in Fig. 3 in fact, one needs to filter out the large-amplitude particle vibrations for a given configuration

of fixed interparticle bonds, since they are far bigger than the amplitude of the cage rattling in a dense system. The problem can be solved by associating to any given network configuration a reference configuration, computed by constraining the interparticle bonds not to break and evaluating the average position of each particle during a simulation at constant temperature of the same model (Colombo et al. 2013; Colombo and Del Gado 2014). The distribution of intrinsic displacements over such isoconfigurational ensemble at a fixed time quantifies, for each particle, its tendency to undergo a significant displacement, thus contributing to the cooperative network restructuring. In Fig. 3 (right), the thickness of the gel strands indicates the parts of the network where the displacements responsible for the cooperative restructuring take place. These large displacements originate in regions where the stress accumulation favors the breaking of the gel structure and tend to propagate along the network. Gels may indeed contain weaker regions (where local stresses are higher and breaking is more likely to occur), and the simulations indicate that the consequences of bond breaking are nonlocal, with cooperative displacements involving parts of the structure that are relatively far away, in regions where non-affine displacements are more likely to concentrate under deformation. The mesoscale organization of the gel network has therefore a significant role, favoring stress build-up and bond breaking in densely connected regions. The occurrence and the spatial distribution of the non-affine motion in amorphous materials under deformation provides indeed useful insight into the mechanical response.

3.2   Micromechanical Consequences

The spatiotemporal analysis of the microscopic dynamical processes just described can be also used to unravel the elementary non-affine rearrangements that can be at the origin of the material response to an external perturbation. Figure 4 shows the results of such spatiotemporal analysis in a 3D soft solid model that was sheared at constant rate (Vasisht et al. 2018). The analysis, in this case, is based on the spatial correlations of elementary microscopic rearrangements (EMR), which can be obtained from the local non-affine motion (computed subtracting the overall imposed flow). Interestingly enough, the EMR strongly depend on the deformation rate imposed to shear the material: the flow at high shear rates originates from a highly coordinated motion of small-particle clusters that move collectively in the same direction. Conversely, at low shear rates, the bulk flow emerges from non-affine rearrangements that take place over much larger correlated domains and trigger events later in time, due to the long-range elastic strain field of the initially solid amorphous material. The results of the simulations, in this case, have been directly tested in experiments that combine bulk rheology with confocal microscopy, as shown in the figure, where similar data on the non-affine motion and its spatiotemporal correlations can be obtained. The microscopic insights gained through the simulations suggest that the bulk rheological flow curve of the material can be associated to changes in the nature of the EMR. A quantitative analysis of the EMR indicates that the change in their spatiotemporal correlations may be at the
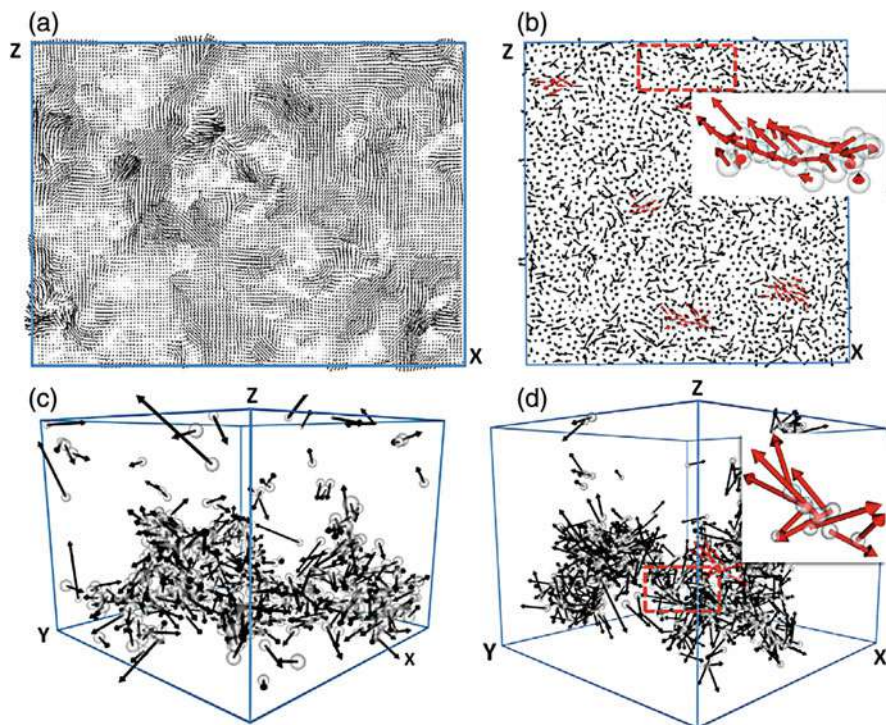
Fig. 4 Non-affine displacement maps for a compressed soft glassy material (experiments on the left and simulations using a coarse-grained model on the right in steady-state shear, respectively, at high rate (**a**) and low deformation rate (**b**). The arrows represent the non-affine displacement vectors. To highlight the localization of displacement vectors, the length of the vectors are amplified. Reproduced with permission from (Vasisht et al. 2018)

origin of the differences in the flow events when the material is sheared at different rates, in agreement with experiments. Recognizing the distinct and rate-dependent nature of the elementary excitations is a fundamental starting point for the development of microscopic theories and constitutive models. The microscopic quantitative information made available from coarse-grained simulations, on when and where local irreversible events are more likely to occur and non-affine displacements are localized, can then be incorporated into nonlocal constitutive models.

4   Toward Constitutive Models for Amorphous Materials

Connecting the different dynamical processes and their mechanical consequences to the emerging rheological properties is one of the crucial problems in the field. Figure 5 illustrates the challenge for the theoretical modeling posed by the largely different length and time scales. Is it possible to develop continuum
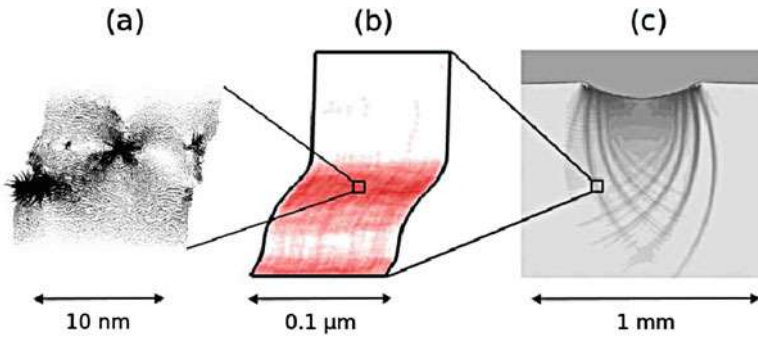
Fig. 5 Illustration of the multiscale problem of a sheared glass from the review of Rodney et al. (2011). (**a**) Localized plastic rearrangements from particle-based simulations (Tanguy et al. 2006). (**b**) Kinetic Monte Carlo simulations of a mesoscale model (Homer and Schuh 2009). (**c**) Constitutive law is used to simulate indentation with a finite element method (Su and Anand 2006)

equations starting from the underlying microscopic dynamics obtained from the approaches described in the previous sections? Usually when dealing with a large number of degrees of freedom with relatively simple local interactions, statistical mechanics is the tool of first choice to derive continuum descriptions by calculating ensemble averages. This technique usually allows to link directly all parameters in the continuum modeling to microscopic potential and dynamical parameters. However, in the case of driven glassy materials, we are dealing with strongly out-of-equilibrium dynamics in a complex potential energy landscape. Already the non-driven system might present out-of-equilibrium relaxation dynamics, also known as aging in glasses, and is as such nonstationary. The additional external forcing only adds to this complexity. Thus all standard tools from equilibrium statistical mechanics are not applicable, and the endeavor of starting from the whole set of microscopic degrees of freedom to derive macroscopic equations turns out to be a formidable if not impossible task.

4.1   Elastoplastic Models

To overcome these fundamental difficulties related to the complexity of the dynamics, a more phenomenological approach, known as elastoplastic modeling, has been proposed to bridge the gap between the different scales (for an extensive review, see Nicolas et al. 2018). Built upon ideas developed by Argon and Kuo 1979, who pioneered the understanding that macroscopic plastic deformation in dense disordered materials is induced by localized plastic events of a size of few particle diameters, models were proposed starting from the scale of the plastic regions (Bulatov and Argon 1994a, b; Baret et al. 2002), which implies an important coarse graining in space and in time. A dense particle assembly,

representing the amorphous solid, is divided into blocks of a size that is relevant to define a meaningful coarse-grained elastic modulus (Tsamados et al. 2009) together with fields of stress, elastic, and plastic deformation. This elastoplastic modeling approach depends crucially on the existence of localized plastic events, and the approach can thus only be valid in deeply jammed systems since close to unjamming dissipation mechanisms are known to become strongly nonlocal (Lerner et al. 2012).

An important ingredient of the elasto-plastic model is that an elastically deforming region in the material acquires a new reference configuration as a response to the external deformation, but also responding to the surrounding plastically deforming regions. This new configuration will differ from the original reference state by some *eigenstrain*. To simplify the problem further we will consider that the plastically deforming zones can be considered point-like at this coarse grained scale. In this limit only the far field response of the surrounding material to the locally deformed plastic inclusions can be considered. In the simplest formulation this response can be derived for a linear isotropic elastic medium, which is a good first order approximation of the complex problem. The mesoscopic model then involves the calculation of the elastic response of the surrounding medium to the forces that are exerted by the zones that have undergone a yield event. This calculation goes along the lines of a standard linear elasticity calculation, but we will omit the details of this technical point and we will give directly the final result that describes the response of the medium at point **r** to a yield event taking place at the origin using the propagator $\mathscr{G}$ given in Eq. 2.

Approximating the rearrangements corresponding to the local plastic events by a force quadrupole (Eshelby problem), one expects a four-fold quadrupolar symmetry for the inhomogeneous part of the stress propagator (Eshelby 1957) (see middle panel of Fig. 6). The form of this propagator for an infinite two dimensional medium reads in non-dimensional form

$$\mathscr{G}^{(\infty)}(r, \theta) = \frac{1}{\pi r^2} \cos(4\theta). \tag{2}$$

Elastic deformation   Plastic event   Stress redistribution
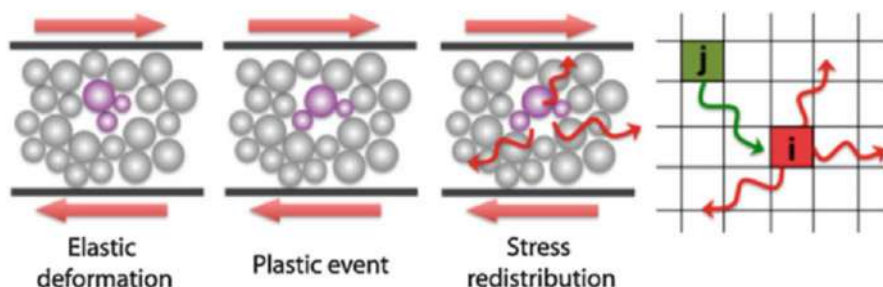
Fig. 6 Sketch of plastic deformation in amorphous media. Deformation occurs via elastic deformation, localized plastic events, and nonlocal redistribution of the elastic stress, potentially triggering other plastic events. (Reproduced with permission from Bocquet et al. 2009).

To model a yield stress material under steady shear with a fixed strain rate $\dot{\gamma}$, we start from a simple Maxwell-like description of the stress dynamics associated with a given element:

$$\partial_t \sigma(t) = \mu \dot{\gamma} - 2\mu \dot{\varepsilon}^{\text{pl}}(t) \tag{3}$$

where ∂_t denotes the partial time derivative, $\sigma(t)$ is the shear component of the stress tensor, μ is the shear modulus, and $\dot{\varepsilon}^{\text{pl}}(t)$ accounts for the change in the strain due to local yielding. To take into account the long-range effects of the plastic events, first predicted by Argon and Kuo (1979) and later evidenced in molecular dynamics simulations (Tanguy et al. 2006) and experiments on colloids (Schall et al. 2007), we change the above homogeneous equation to a field description with a stress propagator $G(\mathbf{r} - \mathbf{r}')$ accounting for the stress redistribution due to a plastic event:

$$\partial_t \sigma(\mathbf{r}, t) = \mu \dot{\gamma} + 2\mu \int d\mathbf{r}' \mathscr{G}(\mathbf{r} - \mathbf{r}') \dot{\varepsilon}^{\text{pl}}(\mathbf{r}', t) \tag{4}$$

To describe the deformation due to the plastic events, we model a Maxwellian viscoelastic-like relaxation of the material in the plastic state, i.e.,

$$\dot{\varepsilon}^{\text{pl}}(\mathbf{r}, t) = \frac{1}{2\mu\tau} n(\mathbf{r}, t) \sigma(\mathbf{r}, t), \tag{5}$$

where τ is the typical time for the stress release in the plastic phase and $n(\mathbf{r}, t)$ a local state variable. In the following, we refer to $n(\mathbf{r}, t)$ as local activity. We define $n(\mathbf{r}, t) = 0$ indicating the absence of a plastic event and $n(\mathbf{r}, t) = 1$ if the local region is in the plastic phase.

To incorporate the idea of a yield stress, we define the following stochastic dynamics for the state variable $n(\mathbf{r}, t)$ (see the left panel of Fig. 7). If the local stress exceeds a threshold value $\sigma(\mathbf{r}, t) > \sigma_y$, there is a finite probability to yield locally. The yielding rate is given by $1/\tau_{\text{pl}}$. Once yielded ($n(\mathbf{r}, t) = 1$), stress is redistributed using an appropriate stress propagator, and locally the stress is relaxing during a typical restructuring time τ_{el}. The rate to re-establish the elastic state ($n(\mathbf{r}, t) = 0$) is given by $1/\tau_{\text{el}}$. These dynamical rules are then discretized into a set of elastoplastic elements that occupy the nodes of a regular square lattice.

It has been shown that this very simplified cellular automata model for the mesoscopic elasto-plastic dynamics leads to many qualitative and semi-quantitative results, fitting the fluctuations and rheological measurements from experiments and microscopic simulations. Such cellular automata have been intensively studied in the framework of dynamical heterogeneities Martens et al. (2011) avalanche dynamics Talamali et al. (2011), shear-banding Martens et al. (2012), critical phenomena Lin et al. (2014) and transient creep dynamics Merabia and Detcheverry (2016), to name only few fields where they have been successfully used to understand the link between microscopic plasticity and the macroscopic mechanical response.
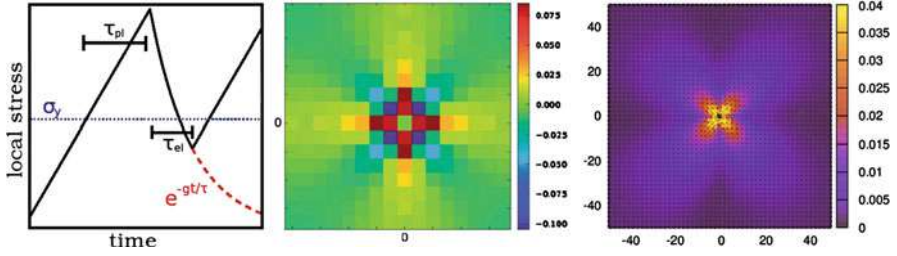
Fig. 7 Local stress dynamics. (**a**) Schematic representation of the local stress of an isolated site as a function of time. When the stress is larger than the thresholds, the site yields with a probability given by the plastic rate. While the site is active, the stress drops exponentially during a typical restructuring time until it becomes elastic again. The site is supposed to be isolated, to simplify the picture. In the completely connected model described in the main text, we would expect stress fluctuations due to the mechanical noise coming from the surrounding sites. (**b**) Stress change due to a single plastic event at the origin. The stress change at the origin is artificially set to zero to account for the visibility of the nearest neighbor changes. In reality the change at the origin is strongly negative. Reproduced with permission from (Martens et al. 2012). (**c**) Averaged displacement field caused by induced plastic events obtained from bidisperse Lennard-Jones particle simulations. Reproduced with permission from (Puosi et al. 2014)

4.2   Deriving Mean-Field Descriptions

The theoretical understanding of the yielding transition in athermally driven disordered systems is a highly challenging problem, and no consensus has been established even on the basic ingredients that should underlie coarse-grained descriptions of the nonlinear rheological response of yield stress materials (Sollich et al. 1997; Hébraud and Lequeux 1998; Falk and Langer 1998; Agoritsas et al. 2015; Lin and Wyart 2016).

A way to derive a phenomenological mean-field description from the above lattice model is to introduce the probability density $\mathscr{P}$ of local shear stresses $\sigma = \sigma_{xy}$ in regions of mesoscopic size W, while the material is sheared at rate $\dot{\gamma}$. The time evolution of $\mathscr{P}$ is given by

$$\partial_t \mathscr{P}(\sigma, t) = -G_0 \dot{\gamma}(t) \partial_\sigma \mathscr{P} - \frac{1}{\tau} \theta(|\sigma| - \sigma_c) \mathscr{P}$$
$$+ \Gamma(t)\delta(\sigma) + D(t)\partial_\sigma^2 \mathscr{P} \qquad (6)$$

where $\theta(x)$ and $\delta(x)$ denote, respectively, the usual Heaviside and delta distributions. The first term on the right-hand side proportional to the stress gradient of the probability density $\partial_\sigma \mathscr{P}$ accounts for the linear elastic response. The following term describes the loss in the probability density due to local yielding of overstressed regions above a critical stress σ_c at a rate given by $1/\tau$. It has been argued that the phenomenological parameter τ can be interpreted as the duration of a plastic event in the low shear rate limit (Agoritsas et al. 2015). The corresponding gain term is given in the third expression on the right-hand side, where the stress is set to zero after a yielding event with a rate given by the plastic activity rate

$$\Gamma(t) = \frac{1}{\tau} \int_{|\sigma|>\sigma_c} \mathscr{P}(\sigma, t)\mathrm{d}\sigma. \tag{7}$$

The last term encompasses the stress changes created through other yielding events in a mean-field manner, assuming that this mechanical noise can be approximated through a normal diffusion of the mesoscopic stresses. To describe this noise in a self-consistent manner, the HL approach (from Hébraud and Lequeux 1998) proposes that its amplitude should be related to the rate of plastic activity through a dimensionless coupling constant $\tilde{\alpha}$:

$$D(t) = \tilde{\alpha}\sigma_c^2 \Gamma(t). \tag{8}$$

This last relation introduces a nonlinearity into Eq. (6), since the rate of plastic activity itself depends on the probability density of the mesoscopic stresses. It is this coupling that renders the problem nontrivial and yields interesting results regarding the behavior of macroscopic quantities.

This model is known to exhibit a unique stationary state for a finite shear rate in the large time limit, where the probability density for the mesoscopic stresses becomes time independent. To determine the time-averaged macroscopic stress in the steady state, one averages over the mesoscopic stresses weighted by the corresponding steady-state probability density

$$\langle \sigma \rangle = \int \sigma \mathscr{P}(\sigma)\mathrm{d}\sigma. \tag{9}$$

Using appropriate units, we can write the equations in dimensionless quantities, expressing stress-related values in units of the local yield stress σ_c, time quantities in units of τ, the shear rate in units of $\sigma_c/(G_0\tau)$, and the stress diffusion coefficient in units of σ_c^2/τ, leaving only two independent dimensionless model parameters that determine the flow behavior, namely, the dimensionless shear rate and coupling constant $\tilde{\alpha}$.

The rheological results in the small shear rate limit for this model are well studied (Hébraud and Lequeux 1998; Olivier and Renardy 2011; Agoritsas and Martens 2017). For small enough coupling strength $\tilde{\alpha} < 1/2$, the HL model predicts a Herschel-Bulkley flow behavior of exponent $1/2$:

$$\langle \sigma \rangle \approx \sigma_Y + A\dot{\gamma}^{1/2}$$

for the time-averaged macroscopic stress in the steady state $\langle \sigma \rangle = \int \sigma \mathscr{P}(\sigma)\mathrm{d}\sigma$, with the two constants σ_Y (the dynamical yield stress) and A (the prefactor). Thus this very simplified model can already capture well the nonlinear behavior of the flow of yield stress materials. Moreover, it has also been shown that the typical transient dynamics during creep experiments can be qualitatively well reproduced using these modeling techniques (Liu et al. 2018).

5   Conclusions

The theoretical approach described in Sect. 4 captures many aspects of the complexity of the mechanics of amorphous solids and in particular their intrinsically nonlinear response. Particularly interesting to the discussion here is that, in principle, the more microscopic information obtained from the studies described in Sects. 2 and 3 can be included in such approach to bridge quantitatively the physics and physical chemistry of specific materials (Section 1) to the general theoretical formulation of their elastoplastic response (Section 4). While developments in this direction are still at a very early stage and far from being straightforward, a first remarkable example is shown in Fig. 7 (Puosi et al. 2014) which visualizes the averaged displacement field caused by induced plastic events obtained from a simple coarse-grained model for a glassy material, of the same type as the ones discussed in Sect. 3. The plot proves that the long-range strain field that is at the core of the elastoplastic modeling can be quantitatively obtained from a microscopic model. Along this line, various information on microscopic mechanical and dynamical heterogeneities, as described in Sects. 2 and 3, (and obtained through the approach described in Section 1), can be transferred into the parameters of the elastoplastic description (e.g., the distribution of local elastic moduli, local yield stresses, or characteristic relaxation times) to develop quantitative constitutive models that do not rely on ad hoc fitting parameters but truly bridge the relevant information over a dramatic range of length scales.

References

Agoritsas E, Martens K (2017) Non-trivial rheological exponents in sheared yield stress fluids. Soft Matter 13(26):4653–4660

Agoritsas E, Bertin E, Martens K, Barrat JL (2015) On the relevance of disorder in athermal amorphous materials under shear. Eur Phys J E 38(7):71

Allen MP, Tildesley DJ (1987) Computer simulations of liquids. Oxford University Press, New York

Argon A, Kuo H (1979) Plastic flow in a disordered bubble raft (an analog of a metallic glass). Mater Sci Eng 39(1):101–109

Attard P (1996) Electrolytes and the electric double layer. In: Prigogine I, Rice SA (eds) Advances in Chemical Physics 92. Wiley, pp 1–159. https://doi.org/10.1002/9780470141519.ch1

Baret JC, Vandembroucq D, Roux S (2002) Extremal model for amorphous media plasticity. Phys Rev Lett 89(19):195506

Bocquet L, Colin A, Ajdari A (2009) Kinetic theory of plastic flow in soft glassy materials. Phys Rev Lett 103(3):036001

Bonnaud PA, Labbez C, Miura R, Suzuki A, Miyamoto N, Hatakeyama N, Miyamoto A, Van Vliet KJ (2016) Interaction grand potential between calcium-silicate-hydrate nanoparticles at the molecular level. Nanoscale 8:4160–4172. https://doi.org/10.1039/C5NR08142D

Bulatov V, Argon A (1994a) A stochastic model for continuum elasto-plastic behavior. I. Numerical approach and strain localization. Model Simul Mater Sci Eng 2(2):167

Bulatov V, Argon A (1994b) A stochastic model for continuum elasto-plastic behavior. III. Plasticity in ordered versus disordered solids. Model Simul Mater Sci Eng 2(2):203

Carrier B (2013) PhD thesis. Ecole Nationale des Ponts et Chaussées, Marne-la-Vallée, France

Colombo J, Del Gado E (2014) Self-assembly and cooperative dynamics of a model colloidal gel network. Soft Matter 10(22):4003–4015

Colombo J, Widmer-Cooper A, Del Gado E (2013) Microscopic picture of cooperative processes in restructuring gel networks. Phys Rev Lett 110(19):198301

de Candia A, Del Gado E, Fierro A, Sator N, Tarzia M, Coniglio A (2006) Columnar and lamellar phases in attractive colloidal systems. Phys Rev E 74:010403. https://doi.org/10.1103/PhysRevE.74.010403

Del Gado E, Ioannidou K, Masoero E, Baronnet A, Pellenq RJ-M, Ulm FJ, Yip S (2014) A soft matter in construction – statistical physics approach to formation and mechanics of c–s–h gels in cement. Eur Phys J-Spec Top 223(11):2285–2295. https://doi.org/10.1140/epjst/e2014-02264-1

Doi M (2013) Soft matter physics. Oxford University Press, Oxford

Dormieux L, Kondo D, Ulm FJ (2006) Microporomechanics. Wiley, Chichester

Ebrahimi D, Pellenq RJM, Whittle AJ (2012) Nanoscale elastic properties of montmorillonite upon water adsorption. Langmuir 28(49):16855–16863. https://doi.org/10.1021/la302997g

Ebrahimi D, Whittle AJ, Pellenq RJ-M (2014) Mesoscale properties of clay aggregates from potential of mean force representation of interactions between nanoplatelets. J Chem Phys 140(15):154309

Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc R Soc Lond A 241(1226):376–396

Falk ML, Langer JS (1998) Dynamics of viscoplastic deformation in amorphous solids. Phys Rev E 57(6):7192

Frenkel D, Smit B (2001) Understanding molecular simulation: from algorithms to applications. Elsevier, Amsterdam

Garrault S, Finot E, Lesniewska E, Nonat A (2005) Study of C-S-H growth on C3S surface during its early hydration. Mater Struct 38(4):435–442. https://doi.org/10.1007/BF02482139

Hébraud P, Lequeux F (1998) Mode-coupling theory for the pasty rheology of soft glassy materials. Phys Rev Lett 81(14):2934

Homer ER, Schuh CA (2009) Mesoscale modeling of amorphous metals by shear transformation zone dynamics. Acta Materialia 57(9):2823–2833

Ioannidou K, Pellenq RJ-M, Del Gado E (2014) Controlling local packing and growth in calcium–silicate–hydrate gels. Soft Matter 10:1121–1133

Ioannidou K, Kanduc M, Li L, Frenkel D, Dobnikar J, Del Gado E (2016a) The crucial effect of early-stage gelation on the mechanical properties of cement hydrates. Nat Commun 7:12106

Ioannidou K, Krakowiak KJ, Bauchy M, Hoover CG, Masoero E, Yip S, Ulm FJ, Levitz P, Pellenq RJM, Del Gado E (2016b) Mesoscale texture of cement hydrates. Proc Natl Acad Sci 113(8):2029–2034. https://doi.org/10.1073/pnas.1520487113

Ioannidou K, Carrier B, Vandamme M, Pellenq R (2017a) The potential of mean force concept for bridging (length and time) scales in the modeling of complex porous materials. In: EPJ web of conferences, EDP sciences, vol 140, p 01009

Ioannidou K, Del Gado E, Ulm FJ, Pellenq RJ-M (2017b) Inhomogeneity in cement hydrates: linking local packing to local pressure. J Nanomech Micromech 7(2):04017003

Israelachvili JN (1992) Intermolecular and surface forces: with applications to colloidal and biological systems (Colloid Science), 2nd edn. Academic Press, Cambridge, MA

Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13(8):1011–1021

Laubie H, Radjai F, Pellenq R, Ulm FJ (2017) Stress transmission and failure in disordered porous media. Phys Rev Lett 119:075501. https://link.aps.org/doi/10.1103/PhysRevLett.119.075501

Lerner E, Düring G, Wyart M (2012) A unified framework for non-brownian suspension flows and soft amorphous solids. Proc Natl Acad Sci 109(13):4798–4803

Lesko S, Lesniewska E, Nonat A, Mutin JC, Goudonnet JP (2001) Investigation by atomic force microscopy of forces at the origin of cement cohesion. Ultramicroscopy 86(1–2):11–21. http://www.ncbi.nlm.nih.gov/pubmed/11215612

Lin J, Lerner E, Rosso A, Wyart M (2014) Scaling description of the yielding transition in soft amorphous solids at zero temperature. Proc Natl Acad Sci 111(40):14382–14387

Lin J, Wyart M (2016) Mean-field description of plastic flow in amorphous solids. Phys Rev X 6(1):011005

Liu C, Martens K, Barrat JL (2018) Mean-field scenario for the athermal creep dynamics of yield-stress fluids. Phys Rev Lett. APS 120(2):028004

Martens K, Bocquet L, Barrat JL (2011) Connecting diffusion and dynamical heterogeneities in actively deformed amorphous systems. Phys Rev Lett 106(15):156001. http://link.aps.org/doi/10.1103/PhysRevLett.106.156001

Martens K, Bocquet L, Barrat JL (2012) Spontaneous formation of permanent shear bands in a mesoscopic model of flowing disordered matter. Soft Matter 8(15):4197–4205

Masoero E, Del Gado E, Pellenq RJ-M, Ulm FJ, Yip S (2012) Nanostructure and nanomechanics of cement: polydisperse colloidal packing. Phys Rev Lett 109(15):155503

Masoero E, Del Gado E, Pellenq RJ-M, Yip S, Ulm FJ (2014) Nano-scale mechanics of colloidal C–S–H gels. Soft Matter 10:491–499. https://doi.org/10.1039/C3SM51815A

Merabia S, Detcheverry F (2016) Thermally activated creep and fluidization in flowing disordered materials. EPL (Europhysics Letters) 116(4):46003

Morriss GP, Evans DJ (2013) Statistical mechanics of nonequilbrium liquids. ANU Press, Canberra

Mosayebi M, Ilg P, Widmer-Cooper A, Del Gado E (2014) Soft modes and nonaffine rearrangements in the inherent structures of supercooled liquids. Phys Rev Lett 112(10):105503

Nicolas A, Ferrero EE, Martens K, Barrat JL (2018) Deformation and flow of amorphous solids: a review of mesoscale elastoplastic models. Rev Mod Phys 90:045001

Olivier J, Renardy M (2011) Glass transition seen through asymptotic expansions. SIAM J Appl Math 71(4):1144–1167

Pellenq RJ-M, Van Damme H (2004) Why does concrete set? The nature of cohesion forces in hardened cement-based materials. MRS Bull 29(5):319–323

Pellenq RJ-M, Caillol JM, Delville A (1997) Electrostatic attraction between two charged surfaces: a (n,v,t) monte carlo simulation. J Phys Chem B 101(42):8584–8594. https://doi.org/10.1021/jp971273s

Pellenq RJ-M, Lequeux N, van Damme H (2008) Engineering the bonding scheme in C-S-H: the iono-covalent framework. Cem Concr Res 38(2):159–174. https://doi.org/10.1016/j.cemconres.2007.09.026; http://www.sciencedirect.com/science/article/pii/S0008884607002372

Pellenq RJ-M, Kushima A, Shahsavari R, Van Vliet KJ, Buehler MJ, Yip S, Ulm FJ (2009) A realistic molecular model of cement hydrates. Proc Nat Acad Sci USA 106(38):16102–16107. https://doi.org/10.1073/pnas.0902180106; http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2739865&tool=pmcentrez&rendertype=abstract

Puosi F, Rottler J, Barrat JL (2014) Time-dependent elastic response to a local shear transformation in amorphous solids. Phys Rev E 89(4):042302

Rodney D, Tanguy A, Vandembroucq D (2011) Modeling the mechanics of amorphous solids at different length scale and time scale. Model Simul Mater Sci Eng 19(8):083001

Schall P, Weitz DA, Spaepen F (2007) Structural rearrangements that govern flow in colloidal glasses. Science 318(5858):1895–1899

Sollich P, Lequeux F, Hébraud P, Cates ME (1997) Rheology of soft glassy materials. Phys Rev Lett 78(10):2020

Su C, Anand L (2006) Plane strain indentation of a zr-based metallic glass: experiments and numerical simulation. Acta Materialia 54(1):179–189

Tanguy A, Leonforte F, Barrat JL (2006) Plastic response of a 2d Lennard-Jones amorphous solid: detailed analysis of the local rearrangements at very slow strain rate. Eur Phys J E 20(3): 355–364

Talamali M, Petäjä V, Vandembroucq D, Roux S (2011) Avalanches, precursors, and finite-size fluctuations in a mesoscopic model of amorphous plasticity. Phys Rev E 84:016115

Tsamados M, Tanguy A, Goldenberg C, Barrat JL (2009) Local elasticity map and plasticity in a model Lennard-Jones glass. Phys Rev E 80(2):026112

Vasisht VV, Dutta SK, Del Gado E, Blair DL (2018) Rate dependence of elementary rearrangements and spatiotemporal correlations in the 3d flow of soft solids. Phys Rev Lett 120(1):018001

Widmer-Cooper A, Harrowell P (2006) Predicting the long-time dynamic heterogeneity in a supercooled liquid on the basis of short-time heterogeneities. Phys Rev Lett 96:185701

Widmer-Cooper A, Harrowell P (2007) On the study of collective dynamics in supercooled liquids through the statistics of the isoconfigurational ensemble. J Chem Phys 126(15):154503

Widmer-Cooper A, Perry H, Harrowell P, Reichman DR (2008) Irreversible reorganization in a supercooled liquid originates from localized soft modes. Nat Phys 4:711–715

Widmer-Cooper A, Perry H, Harrowell P, Reichman DR (2009) Localized soft modes and the supercooled liquid's irreversible passage through its configuration space. J Chem Phys 131(19):194508

Zhuang Y, Zhang K, Charbonneau P (2016) Equilibrium phase behavior of a continuous-space microphase former. Phys Rev Lett 116:098301. https://link.aps.org/doi/10.1103/PhysRevLett.116.098301

Nanomechanics of Materials: Overview

74

Ting Zhu and Dengke Chen

Contents

Abstract

Recent experiments on nanoscale and nanostructured materials, including nanowires, nanopillars, nanoparticles, nanocrystalline, and two-dimensional (2D) materials, have revealed a host of "ultra-strength" phenomena, defined by stresses in the material generally rising up to a significant fraction of the ideal strength – the highest achievable strength of a defect-free crystal.

T. Zhu (✉)
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

School of Material Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA
e-mail: ting.zhu@me.gatech.edu

D. Chen
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA
e-mail: dengke.chen@me.gatech.edu

Understanding the "ultra-strength" phenomena requires an in-depth study that integrates the nanomechanical experiment and modeling. Here we present an overview on the strength-controlling deformation mechanisms as well as the nanomechanical modeling studies for ultra-strength materials. The general concepts and principles are described, with a particular emphasis on the size, temperature, and strain-rate dependence of the ultra-strength phenomena. The chapters in this Part of the handbook are reviewed to highlight recent progress in the nanomechanical modeling of ultra-strength materials. Perspectives on the future study of nanomechanics of materials are discussed.

1   Introduction

An engineering revolution is currently under way in that small structures and devices are now being fabricated at the micrometer to nanometer scales. The reliability concerns of such small systems call for an understanding of the mechanical properties of materials at small length scales (Freund and Suresh 2003). This overview describes the general concepts and principles underlying the ultra-strength phenomena in materials with very small length scales. The chapters in this Part of the handbook are reviewed to highlight recent progress in the nanomechanical modeling of ultra-strength materials.

Ultra-strength phenomena are characterized by sample-wide stress levels reaching a significant fraction of the material's ideal strength, defined as the highest stress a perfect crystal can sustain without undergoing immediate structural transformation (Ogata et al. 2002, 2004; Li et al. 2007). There is an intimate relation between observable strength and the timescale of observation. It is difficult to avoid "observing" ultra-strength in molecular dynamics simulations because of the short simulation timescales, typically nanoseconds, for stress ramp-up. But in recent years, more and more ultra-strength behaviors are revealed in laboratory experiments, typically on the seconds-to-minutes timescale. This is often enabled by the small sample sizes, small microstructural length scales, or small probes.

1.1   Ultra-Strength Experiments

Recent years have seen rapid growth of experimental studies on ultra-strength material systems and phenomena. This is largely enabled by the refinement of nanoscale mechanical experiments, which allow one to study the near-ideal strength behavior quantitatively in a controlled fashion. Representative data of experimentally measured ultra strengths have been collected previously (Zhu et al. 2009). These data are either yield or fracture strengths measured from tension, compression, or bending experiments. Evidently, ultra-strength behavior can be considered as a GPa-level phenomenon, in contrast to the relatively low strengths (tens to hundreds of MPa) in coarse-grained hard materials.

1.2   Length-Scale Effect

The attainment of ultra strength can result from a length-scale effect. For coarse-grained crystals, the experimentally measured strengths are far below the theoretical values. The low strength arises because the sample usually has a large number of grown-in defects such as dislocations or cracks. In well-annealed crystals, the dislocation density is initially low but nonzero. It increases rapidly once deformation starts, due to, for example, double cross slip and Frank-Read-type multiplication processes. The yield strength is then limited by the resistance to dislocation motion, typically a factor of 100 below the ideal strength, or even more in coarse-grained samples.

The strength of materials with preexisting defects can be increased by reducing the characteristic length scales in materials. Generally, the strength-controlling length scale can be the sample size, grain size, layer thickness, or contact loading zone size. In crystals, the strengthening mechanism is typically based on the blockage or elimination of dislocations. Various detailed size-strengthening mechanisms will be reviewed later in this chapter.

1.3   Timescale Effect

Ultra-strength phenomena are also closely related to the observation timescale, as manifested by the dependence of strength on loading rate and temperature. To understand the strength, temperature, and timescale relationships, one needs to know the controlling deformation mechanism, which is the focus of this review. Then the kinetic rate theory can be used to link time and strength.

1.4   Rate Theory of Strength

A unit process of strength-controlling structural transformation can be characterized by two types of quantitative measures: the athermal strength limit and activation parameters. The former measures the zero-temperature structural stability limit at which the structural transformation occurs instantaneously without the aid of thermal fluctuations. The activation parameters, including activation energy and activation volume, characterize the probabilistic nature of transformation by thermal fluctuations when the applied load is below the athermal limit. Consider, as an example, thermally activated shear transformation at a given temperature T and applied shear stress τ. According to transition state theory (Vineyard 1957), the transformation rate is

$$\nu = N\nu_0 \exp\left(-\frac{Q(\tau, T)}{k_B T}\right) \tag{1}$$

where ν_0 is the trial frequency, N is the number of equivalent sites of transformation, $k_\mathrm{B}T$ is the thermal energy, and Q is the activation free energy whose magnitude is controlled by the local shear stress τ along with the temperature. The athermal strength limit corresponds to the critical stress, giving zero activation energy.

The sensitivity of transformation rate to stress can be measured by the activation volume Ω, defined as

$$\Omega = -\frac{\partial Q}{\partial \tau} = k_\mathrm{B}T\frac{\partial \ln \nu}{\partial \tau}. \tag{2}$$

Physically, the activation volume is proportional to the number of atoms simultaneously involved in a thermally activated process, such that it measures the individual (single-particle) and collective nature of a transition. Different rate processes may have drastically different characteristic activation volumes, e.g., $\Omega \sim 0.1b^3$ for lattice diffusion and $\Omega \sim 100$–$1000b^3$ for Orowan looping through forest dislocation intersections, where b denotes the magnitude of the Burgers vector. As a result, the activation volume can serve as an effective kinetic signature of the deformation mechanism. Importantly, the activation volume can be determined by both atomistic modeling and experiments, thus providing a useful link between the two approaches for revealing the strength-controlling mechanisms. In atomistic modeling, the stress-dependent activation energy and activation volume can be calculated in a robust and efficient manner by using the reaction pathway exploration approach such as the nudged elastic band method (Henkelman et al. 2000). In experiments, the activation volume is typically determined by measuring the strain-rate sensitivity.

2   Strength-Controlling Deformation Mechanisms

Because of the interposition of small length scales, the deformation of ultra-strength materials often operates in a source-limited and nucleation-controlled regime. We next review representative deformation mechanisms in ultra-strength materials. The critical role of activation volume is emphasized in understanding the strength-controlling mechanisms, as well as the size, temperature, and strain-rate dependence of material responses.

2.1   Dislocation Nucleation

Ultra strength can be achieved in a small volume of perfect crystal. At low temperatures, the strength of a perfect crystal is limited by dislocation nucleation in the bulk or from the surface, depending on the sample geometry and loading mode. Homogenous dislocation nucleation in the bulk crystal has been experimentally observed in the atomic model systems of bubble raft (Gouldstone et al. 2001) and colloidal crystal (Schall et al. 2006) and also in atomistic simulations (Li et al.

2002). In addition, displacement bursts in nanoindentation experiments have been associated arguably with homogeneous dislocation nucleation in a small and perfect volume beneath the nano-sized indenter tip (Li et al. 2002).

Athermal dislocation nucleation in a perfect crystal results from elastic instability at large shear stress; equivalently, it occurs when the energy barrier of nucleation vanishes. Because of the elastic softening effect at large deformation, the athermal threshold of homogeneous dislocation nucleation should be determined by using the theory of hyperelastic (i.e., nonlinear elastic) instability (Hill 1975). An atomistics-based hyperelastic instability criterion has been developed that balances the elastically softening modulus and the GPa-level high stress at large shear (Li et al. 2002). Using this criterion, the critical nucleation conditions at zero temperature can be accurately predicted, including the nucleation site, critical stress, and activated slip system. The athermal nucleation from a surface can be similarly modeled in the hyperelasticity framework with an appropriate description of the surface stress effect (Park and Klein 2007).

Increasing temperature can lower the strength because of thermal fluctuations in assisting dislocation nucleation. To understand the temperature and strain-rate dependence of dislocation nucleation, Zhu et al. (2008) have developed an atomistic model that integrates reaction pathway calculations and theoretical analyses based on transition state theory. Consider, as an example, surface dislocation nucleation in a Cu nanopillar under an applied constant strain rate. Because of the probabilistic nature of the thermally activated nucleation process, the nucleation stress has a distribution even if identical samples are used. The most probable nucleation stress is defined by the peak of the nucleation event frequency distribution. Based on this nucleation statistics-based definition, Zhu et al. (2008) developed a nonlinear theory to determine the nucleation stress as a function of temperature and strain rate. A key result is that the activation volume associated with the surface dislocation source is within 1–$10b^3$, much lower than that of bulk dislocation processes, 100–$1000b^3$. The physical effect of such small activation volume can be clearly seen from a linear version of the theory, giving an analytic formula of the nucleation stress

$$\sigma = \sigma_a - \frac{k_B T}{\Omega} \ln \frac{k_B T N \nu_0}{E \dot{\varepsilon} \Omega} \tag{3}$$

where σ_a is the athermal stress of instantaneous nucleation and E is Young's modulus. Notice that the nucleation stress σ has a temperature scaling of $T\ln T$ and the activation volume Ω appears outside the logarithm, so that the small Ω associated with a surface source should lead to sensitive temperature and strain-rate dependence of nucleation stress, as quantitatively shown in atomistic simulations (Zhu et al. 2008). In very small volumes, surface dislocation nucleation is expected to dominate, as evidenced by the experiment (Bei et al. 2008). As schematically shown in Fig. 1, the strength mediated by surface nucleation should provide an upper bound to the size-strength relation in nanopillar compression experiments; this upper bound is strain-rate sensitive because of the small activation volume.
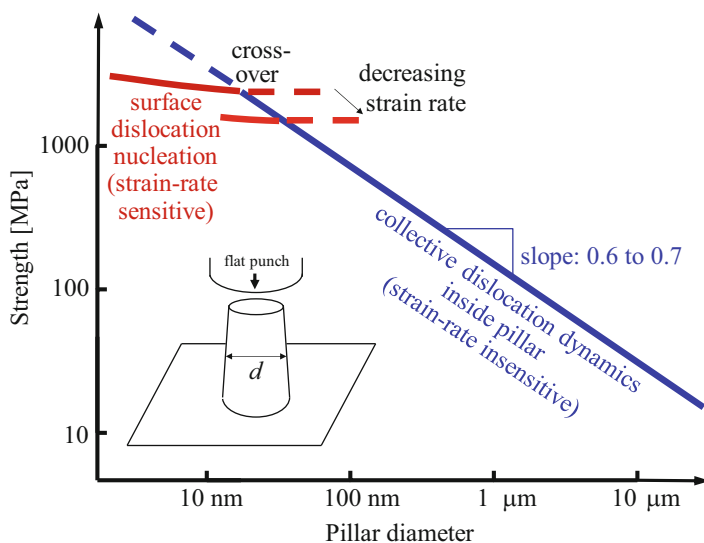
Fig. 1 Illustration of the surface effect on the rate-controlling process and the size dependence of yield strength in micro- and nanopillars of diameter d under compression (Zhu et al. 2008)

2.2   Dislocation Exhaustion

What if mobile dislocations preexist in small-volume materials? Under such conditions, ultra strength can be achieved by dislocation exhaustion. This has been experimentally measured in crystals of very small dimensions where the few dislocations present cannot multiply sufficiently before they are annihilated at free surfaces. Greer and Nix (Greer and Nix 2006) measured the strength of Au pillars under uniaxial compression. For a pillar with a diameter as small as 300 nm, the compressive strength is increased to about 800 MPa, about 50 times higher than for bulk Au. Similar high strengths have been measured for both the single crystal Au pillars and the nanometer-sized ligaments in nanoporous Au (Volkert and Lilleodden 2006; Volkert et al. 2006). All of these experimental values are extremely high for Au, close to the theoretical predictions of ideal shear strength from ab initio calculations (Ogata et al. 2004).

Greer and Nix (2006) have explained the ultra strength in Au nanopillars based on the concept of dislocation starvation. Unlike bulk samples, dislocations in nanopillars can travel only very small distances before annihilating at free surfaces, thereby reducing the overall dislocation multiplication rate. Gliding dislocations leave the crystal more rapidly than they multiply, decreasing the overall mobile dislocation density. Such processes would lead to a dislocation-starved state requiring very high stresses to nucleate new mobile dislocations.

Shan et al. (2008) recently performed in situ compression experiments of submicrometer nickel pillars inside a transmission electron microscope (TEM). The experiments showed that the pillar structures contain a high density of initial defects

after processing but can be made dislocation-free by applying purely mechanical stress. This phenomenon, termed mechanical annealing, provides evidence of source-limited deformation, where the hardening occurs through the progressive exhaustion of dislocation sources, causing dislocation starvation.

2.3   Dislocation-Interface Interaction

In thin films and layered materials, the strengthening arises from the constraints of the surrounding substrate and layers (Kraft et al. 2002). Achieving ultra strength requires that the film and layer thickness be in the range of a few to a few tens of nanometers. Under such conditions, the dislocation-interface interaction is expected to play an important role in controlling the strength.

The effect of layer thickness on flow strength has been summarized by Misra et al. (2005) for metallic multilayers. They showed that a Hall-Petch-type model based on dislocation pileup is applicable at the submicrometer length scales. At the few to a few tens of nanometer length scales, confined layer slip of single dislocations is the operative mechanism; the strength is influenced by interface stress and interface dislocation arrays. As layer thicknesses are further reduced to a few nanometers or less, the strength reaches a peak in ultra strength. This peak strength is set by the interface resistance to single dislocation transmission.

The dislocation-interface interactions may govern not only strength but also ductility. For example, Lu et al. (2004) have reported experiments of an unusual combination of ultrahigh yield strength ($\sim$1 GPa) and high ductility ($\sim$14% elongation to failure) in ultrafine-grained copper (i.e., with a grain size of several hundred nanometers) containing layered nano-twins, typically tens of nanometers in thickness. While the ultrahigh strength can be attributed to the size-strengthening effect of thin twin lamellas, the mechanistic connection between the retained high ductility and twinned nanostructures is not fully understood.

In bulk nanocrystalline metals, tensile ductility is often limited by the onset of necking instability. This has been well studied at the continuum level. The central result is that the high strain-rate sensitivity and the high rate of strain hardening can both help delay the onset and development of necking, thereby promoting tensile ductility (Hart 1967). To connect the phenomenological theories with nanostructure-mediated plastic deformation, one needs to first determine the rate-controlling mechanisms. The experiment by Lu et al. (2005) showed that nano-twinned copper increases the rate sensitivity ($m \approx 0.02$) by up to an order of magnitude relative to microcrystalline metals with grain sizes in the micrometer range and a concomitant decrease in the activation volume by two orders of magnitude (e.g., down to $\Omega \approx 20b^3$ when twin lamellas are approximately 20 nm thick).

The mechanistic origin of decreased activation volume and increased rate sensitivity has been studied for the nano-twinned system (Zhu et al. 2007). Slip transfer reactions between the lattice dislocation and coherent twin boundary (TB) were analyzed. Using the free-end nudged elastic band method, minimum energy paths were determined for reactions of absorption, desorption, and slip transmission.

Predictions of yield stress, activation volume, and rate sensitivity were consistent with the experimental measurement. This agreement demonstrates that the slip transfer reactions at TBs are the rate-controlling mechanisms in nano-twinned copper.

Besides increasing the rate sensitivity, the dislocation-TB interaction causes an increase in the strain-hardening rate, which may play a key role in preserving the tensile ductility. The modeling study (Zhu et al. 2007) indicated that in contrast to the general high-angle grain boundaries in nanocrystals, which have a limited capacity of hardening, the coherent TB is quite hardenable as it gradually loses coherence during the initial stage of plastic deformation. This effective hardenability facilitates the accommodation of incompatible deformation between adjacent twins; it raises the resistance of plastic flow, promotes the strain hardening, and helps retain tensile ductility. Also, the coherent TB is more resistant to tensile decohesion compared to random boundaries and, therefore, can store larger amounts of deformation-induced incompatibility (interfacial dislocation content) before fracture occurs.

2.4   Deformation Twinning

In ultra-strength materials, the formation of deformation twins may dominate over dislocation slip. This has been observed by recent TEM experiments of nanocrystals (Dao et al. 2007). The preference for deformation twinning in nanocrystals can be understood by comparing the critical resolved shear stress (CRSS) needed to emit a perfect dislocation, τ_f, with the CRSS, τ_p, required to initiate the partial twinning dislocation to generate stacking faults and deformation twins (Chen et al. 2003). These CRSSs can be estimated by considering the energy balance between the elastic energy increase in expanding a boundary dislocation loop and the potential energy decrease associated with the work done by the CRSS. Assuming the source size is approximately equal to the grain size, d, the CRSSs for the emission of a perfect dislocation and twinning partials are given by $\tau_f/\mu = b/d$ and $\tau_p/\mu = b_p/d + \Gamma_{sf}$, respectively. Here, $\Gamma_{sf} = \gamma_{sf}/(\mu b_p)$, where γ_{sf} is the stacking fault energy and b_p is the magnitude of the Burgers vectors of a partial dislocation. Equating τ_f and τ_p gives the critical grain size $d_c = (b - b_p)/\Gamma_{sf}$, where the transition will occur from the plastic response dominated by perfect dislocations for large grains to the twinning and stacking faults controlled by plastic behavior for small grains. The estimated d_c is about 10 nm for aluminum nanocrystals.

It should be emphasized that the foregoing analysis of glide dislocations and deformation twins is based on the energy balance of expanding the existing boundary dislocations. As a result, their CRSSs only involve the material energetic parameter of stack fault energy, γ_{sf}, which characterizes the metastable stacking state of the lattice. Van Swygenhoven et al. (2004) have emphasized the importance of examining the generalized stacking fault energy curve, which additionally provides activation-related parameters such as unstable stacking energy γ_{us} and unstable twinning energy γ_{ut}. Along this line, Asaro and Suresh (2005) have performed the mechanics analysis of athermal emission of perfect dislocations

and deformation twins from the grain boundary. The athermal stresses derived incorporate naturally γ_{us} and γ_{ut}. Finally, it is noteworthy that the thermal activation of deformation twins has not yet been studied for understanding the temperature and strain-rate dependence of nucleation stresses at small length scales.

2.5   Fracture

Plastic flow and brittle fracture are competing deformation modes in ultra-strength materials. Bulk silicon is brittle, but silicon nanowires in tension can undergo large plastic deformation at room temperature (Han et al. 2007). A revealing analysis of brittle-ductile competition has been given by Dumitrica et al. (2006) for a perfect single-walled carbon nanotube. By molecular dynamics simulations at various temperatures T and at short timescale (about 0.1 ns), they identified two primary mechanisms: the "cool" mechanism of brittle fracture at low T and the "hot" mechanism of formation of Stone-Wales defects (i.e., dislocation dipoles) at high T. To explore the competition at long timescales, they have studied the energy landscapes governing the two processes. They found the brittle mode of bond breaking requires little thermal agitation; the fracture stress is controlled by the ideal tensile strength of the C-C bond. In contrast, the Stone-Wales defect formation by single-bond rotation sensitively depends on temperature, implying a small activation volume associated with this process. According to transition state theory, the activation energy under constant stress needs to be reduced to $Q = k_B T \ln(N \nu t)$ in order to observe the formation of Stone-Wales dipole over the test duration t (probability ~ 1). Dumitrica et al. (2006) have further calculated Q as a function of stress. Then, the stress needed to form the Stone-Wales dislocation dipole at different temperatures and times can be backed out from the formula of Q given previously. A breaking-strain map (Dumitrica et al. 2006) was predicted to provide useful guidance for the experimental study of temperature and strain-rate dependence of mechanical responses of carbon nanotubes.

The foregoing analysis has considered an idealized situation of competition between the brittle and ductile response in a perfect system (Dumitrica et al. 2006; Kang and Cai 2007; Belytschko et al. 2002). In an alternative scenario, such competition could occur at the crack tip in a system with a preexisting crack (Warner et al. 2007). In this case, bond breaking at the crack tip may have different temperature and strain-rate dependences. The analysis along the lines of Dumitrica et al. (2006) needs to be performed to better understand the brittle-to-ductile transition in ultra-strength materials.

3   Nanomechanical Modeling

Recent years have seen a dramatic increase in the nanomechanical modeling study for a variety of ultra-strength materials, including nanowires, nanopillars, nanopar-ticles, nanocrystalline, and 2D materials. The chapters in this Part of the Handbook

highlight recent progress in the computational modeling of nanomechanical behavior of ultra-strength materials. For example, Ogata (▶ Chap. 75, "First-Principles Modeling of Intrinsic Materials Strength") presents a chapter on the first-principles modeling of intrinsic material strength. He shows that the density functional theory calculations can provide a deep understanding of the effects of electronic structures and chemical bonding on the intrinsic strength of materials, which represents the upper bound of ultra strength. Li et al. (▶ Chap. 76, "Atomistic Simulations of Fracture and Fatigue in Nanotwinned and Amorphous Materials") present a chapter on the atomistic simulations of fracture and fatigue in nano-twinned and amorphous materials. They show that large-scale atomistic simulations can provide unique atomic-level insights into the controlling deformation mechanisms in nano-twinned and amorphous materials. Notably, nanomechanics of 2D materials has attracted considerable attention in recent years. In this section, Sorkin et al. (▶ Chap. 77, "Modelling of Defects and Failure in 2D Materials: Graphene and Beyond") present a chapter on modeling of defects and failure in 2D materials. They show that point defects and line defects can play a critical role in the mechanical behavior of graphene and other 2D materials. Park (▶ Chap. 78, "Mechanics and Electromechanics of Two-Dimensional Atomic Membranes") presents a chapter on the mechanics and electromechanics of 2D atomic membranes. He shows that atomistic simulations can effectively reveal the auxetic behavior in 2D materials. He also describes the new computational techniques that couple molecular dynamics, tight binding, and quantum transport for understanding how mechanical strain can impact the electronic properties of graphene kirigami. Overall, these chapters highlight the power of materials modeling for understanding and predicting the mechanical behavior of nanoscale and nanostructured materials.

4   Conclusions and Perspectives

In the past decade, we have seen a dramatic increase in experimental studies of ultra-strength materials and phenomena, driven by technological demands and projections. The high strength achieved often is caused by a scarcity of grown-in defects and the exhaustion of mobile defect population destabilized by close proximity to interfaces or surfaces, which consequently cause the strength to be defect nucleation controlled rather than propagation controlled. The plastic deformation in the ultra-strength regime is generally characterized by enhanced strain-rate sensitivity and temperature sensitivity. Mechanisms of strain hardening (or the lack thereof), interfacial structural evolution and decohesion, and tensile ductility are some of the issues that need to be further considered for comprehensive understanding of ultra-strength materials at low temperatures. Although temperature-controlled nanomechanical testing is still quite rare today, it is a direction for future growth. The chapters in this Part of the Handbook highlight recent progress in the computational modeling of nanomechanical behavior of ultra-strength materials.

Given the rapid development of both experimental and computational modeling capabilities in recent years, we expect a great advancement in the nanomechanics study of materials in the coming years. These studies will have a significant impact on our understanding of deformation mechanisms in ultra-strength materials, including diffusive and hybrid displacive-diffusive processes. The deformation mechanism maps (Frost and Ashby 1982) in the parametric space of stress, temperature, and strain rate will be obtained through coupled experiment and modeling. Such maps will enable the design of better ultra-strength materials to realize their latent mechanical properties to the full (Ma and Zhu 2017).

References

Asaro RJ, Suresh S (2005) Mechanistic models for the activation volume and rate sensitivity in metals with nanocrystalline grains and nano-scale twins. Acta Mater 53:3369–3382

Bei H, Gao YF, Shim S, George EP, Pharr GM (2008) Strength differences arising from homogeneous versus heterogeneous dislocation nucleation. Phys Rev B 77:060103

Belytschko T, Xiao SP, Schatz GC, Ruoff RS (2002) Atomistic simulations of nanotube fracture. Phys Rev B 65:235430

Chen MW, Ma E, Hemker KJ, Sheng HW, Wang YM, Cheng XM (2003) Deformation twinning in nanocrystalline aluminum. Science 300:1275–1277

Dao M, Lu L, Asaro RJ, De Hosson JTM, Ma E (2007) Toward a quantitative understanding of mechanical behavior of nanocrystalline metals. Acta Mater 55:4041–4065

Dumitrica T, Hua M, Yakobson BI (2006) Symmetry-, time-, and temperature-dependent strength of carbon nanotubes. Proc Natl Acad Sci U S A 103:6105–6109

Freund LB, Suresh S (2003) Thin film materials: stress, defect formation and surface evolution. Cambridge University Press, Cambridge

Frost HJ, Ashby MF (1982) Deformation-mechanism maps. Pergamon Press, New York

Gouldstone A, Van Vliet KJ, Suresh S (2001) Simulation of defect nucleation in a crystal. Nature 411:656

Greer JR, Nix WD (2006) Nanoscale gold pillars strengthened through dislocation starvation. Phys Rev B 73:245410

Han XD, Zheng K, Zhang YF, Zhang XN, Zhang Z, Wang ZL (2007) Low-temperature in situ large-strain plasticity of silicon nanowires. Adv Mater 19:2112–2118

Hart EW (1967) Theory of the tensile test. Acta Metall 15:351–355

Henkelman G, Uberuaga BP, Jonsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113:9901–9904

Hill R (1975) On the elasticity and stability of perfect crystals at finite strain. Math Proc Camb Philos Soc 77:225–240

Kang K, Cai W (2007) Brittle and ductile fracture of semiconductor nanowires – molecular dynamics simulations. Philos Mag 87:2169–2189

Kraft O, Freund LB, Phillips R, Arzt E (2002) Dislocation plasticity in thin metal films. MRS Bull 27:30–37

Li J, Van Vliet KJ, Zhu T, Yip S, Suresh S (2002) Atomistic mechanisms governing elastic limit and incipient plasticity in crystals. Nature 418:307–310

Li T, Morris JW, Nagasako N, Kuramoto S, Chrzan DC (2007) "Ideal" engineering alloys. Phys Rev Lett 98:105503

Lu L, Shen YF, Chen XH, Qian LH, Lu K (2004) Ultrahigh strength and high electrical conductivity in copper. Science 304:422–426

Lu L, Schwaiger R, Shan ZW, Dao M, Lu K, Suresh S (2005) Nano-sized twins induce high rate sensitivity of flow stress in pure copper. Acta Mater 53:2169–2179

Ma E, Zhu T (2017) Towards strength–ductility synergy through the design of heterogeneous nanostructures in metals. Mater Today 20:323–331

Misra A, Hirth JP, Hoagland RG (2005) Length-scale-dependent deformation mechanisms in incoherent metallic multilayered composites. Acta Mater 53:4817–4824

Ogata S, Li J, Yip S (2002) Ideal pure shear strength of aluminum and copper. Science 298: 807–811

Ogata S, Li J, Hirosaki N, Shibutani Y, Yip S (2004) Ideal shear strain of metals and ceramics. Phys Rev B 70:104104

Park HS, Klein PA (2007) Surface Cauchy-Born analysis of surface stress effects on metallic nanowires. Phys Rev B 75:085408

Schall P, Cohen I, Weitz D, Spaepen F (2006) Visualizing dislocation nucleation by indenting colloidal crystals. Nature 440:319–323

Shan ZW, Mishra RK, Asif SAS, Warren OL, Minor AM (2008) Mechanical annealing and source-limited deformation in submicrometre-diameter Ni crystals. Nat Mater 7:115–119

Van Swygenhoven H, Derlet PM, Froseth AG (2004) Stacking fault energies and slip in nanocrystalline metals. Nat Mater 3:399–403

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3:121–127

Volkert CA, Lilleodden ET (2006) Size effects in the deformation of sub-micron Au columns. Philos Mag 86:5567–5579

Volkert CA, Lilleodden ET, Kramer D, Weissmuller J (2006) Approaching the theoretical strength in nanoporous Au. Appl Phys Lett 89:061920

Warner DH, Curtin WA, Qu S (2007) Rate dependence of crack-tip processes predicts twinning trends in f.c.c. metals. Nat Mater 6:876–881

Zhu T, Li J, Samanta A, Kim HG, Suresh S (2007) Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. Proc Natl Acad Sci U S A 104:3031–3036

Zhu T, Li J, Samanta A, Leach A, Gall K (2008) Temperature and strain-rate dependence of surface dislocation nucleation. Phys Rev Lett 100:025502

Zhu T, Li J, Ogata S, Yip S (2009) Mechanics of ultra-strength materials. MRS Bull 34:167–172

First-Principles Modeling of Intrinsic Materials Strength

75

Shigenobu Ogata

Contents

Abstract

Intrinsic materials strength, which is determined by the mechanical nature of the chemical bonds in materials, is essential and key information for developing ultra-strength and, moreover, ultra-toughness materials because a thorough control of the nucleation and migration of defects is necessary to develop these materials. Well-established first-principles approaches enable us to predict the intrinsic strength of materials, which should lead to the high-throughput design of ultra-strength and ultra-toughness materials. In this article, first-principles modeling of the mechanical nature of chemical bonding is discussed, followed by a discussion of first-principles modeling of the strength of a crystal lattice, which is useful for describing not only the intrinsic strength but also the deformation mechanism map of a crystal lattice.

S. Ogata (✉)
Department of Mechanical Science and Bioengineering, Graduate School of Engineering Science, Osaka University, Osaka, Japan
e-mail: ogata@me.es.osaka-u.ac.jp

1   Introduction

The strength of materials is predominantly determined by the mechanical nature of the materials' chemical bonds, although extrinsic factors such as the materials' size, shape, dimensions, microstructure, type, and density of defects, surface, interface, temperature, and strain rate may change their "measurable" strength. However, the upper limit of a material's strength, sometimes referred to as its ideal strength, is always limited by the intrinsic strength of its interatomic chemical bonds, which is uniquely determined by the atomic arrangement and atomic species within the Born-Oppenheimer approximation. The recent challenge in the development of ultra-strength materials (Zhu et al. 2009; Zhu and Li 2010) is how to approach the upper limit by controlling extrinsic factors. Therefore, first extracting and understanding the mechanical nature of chemical bonding and then modeling materials strength while taking into account the extrinsic factors is an intuitive approach for the predictive modeling and high-throughput design of materials strength from first principles. Currently, first-principles density functional theory (DFT) and related first-principles methods enable us to directly examine and study the mechanical nature of chemical bonding in materials and their lattice strength without empirical information, whereas direct experimental measurement is still difficult. In this article, I first discuss the mechanical nature of chemical bonding in materials based on bond stretchability and shearability concepts. These concepts lead to a concept of the intrinsic ductility of materials. I then discuss the strength of crystal lattices from the viewpoint of the Gibbs free energy landscape at zero Kelvin in a joint space of deformation/strain and atomic shuffling, which can be used to study the strength against lattice instabilities that leads to the initiation and nucleation of cracks and dislocations and the occurrence of deformation twinning and solid-solid transformations such as martensitic and Bainitic transformations.

2   Mechanical Strength of Chemical Bonding

The strain limit that chemical bonding can withstand can be defined as the mechanical strength of chemical bonding. In particular, two independent strain limits – specifically, the shear and tensile strain limits – are important for materials strength and plasticity (Foreman et al. 1951; Xu et al. 1995) because the shear-induced bond switching and bond angle-change and tension-induced bond breaking in perfect crystal and in lattice defects, such as dislocations, cracks, and boundaries, in defect-containing crystals actually determine a material's strength and plasticity. Therefore, shearability and tensibility are conveniently defined by the maximum shear and tensile strains at which a perfect crystal becomes unstable (Ogata et al. 2004), which are computable using first-principles DFT. Notably, because chemical bonding in the lattice defects may differ from that in perfect lattice, direct simulation of these lattice defects is necessary to precisely determine the mechanical nature of chemical bonding in defects. However, analysis using perfect crystal lattice, which is more convenient in first-principles computations, should still provide some essential information about the bonding response in defects to the strains.

More precisely, the shearability can be defined as

$$s_m \equiv \arg\max \sigma(s), \tag{1}$$

where $\sigma(s)$ is the resolved shear stress and s is the engineering shear strain in a specified slip system of a crystal. "arg max" is the arguments of the maxima. The tensibility is defined as

$$t_m \equiv \arg\max -P((1+t)V_0), \tag{2}$$

where $P(V)$ is the pressure-volume relation and V_0 is the equilibrium volume, $P(V_0) = 0$. The shearability and tensibility have been computed using first-principles DFT for various metals and ceramics (Ogata et al. 2004), including face-centered cubic (FCC) Ag, Cu, Au, ferromagnetic (FM) and paramagnetic (NM) Ni, and Al; body-centered cubic (BCC) W, Mo, and Fe (FM); hexagonal close-packed (HCP) Mg, Ti, and Zn; L1$_0$ TiAl and D0$_{19}$ Ti$_3$Al; diamond cubic C and Si; β-SiC, α-, and β-Si$_3$N$_4$; and B1 NaCl, MgO, KBr, and CaO. In these computations, incremental affine shear strains were imposed on each crystal along experimentally determined common slip systems to obtain the corresponding unrelaxed and relaxed energies and stresses, defined, respectively, by the conditions $\varepsilon_{ij} = 0$ except $s \equiv x/d_0$, with d_0 being the interplanar separation and x taken along the Burgers vector, and $\sigma_{ij} = 0$ except for the resolved shear stress. For α- and β-Si$_3$N$_4$, the common slip systems are unknown experimentally. Therefore, six systems for each phase are calculated, and the one with the lowest shear strength is used (Ogata et al. 2003). When the relaxed ideal shear stress σ_m^r normalized by G_r and the shearability s_m^r for different materials are plotted together (Fig. 1 (Ogata et al. 2004)), gaps are observed in the distributions of s_m^r between metals and covalent solids. Such gaps can also be observed when comparing the results of different works for elastic shear instability of metals (Krenn et al. 2001; Roundy and Cohen 2001; Ogata et al. 2002) and covalent solids (Roundy et al. 2001; Ogata et al. 2003). In addition, among the metals, the noble metals Au, Ag, and Cu and the more directionally bonded Al and BCC Mo, W, and Fe (FM) are at opposite sides of the distributions, which suggest that directional bonding allows for longer-range shear distortion of the bonds before peak resistance is attained. This behavior can be rationalized by observing that greater bond covalency results in a greater concentration of valence charge in nonnuclear-centered regions (Silvi and Gatti 2000; Mori-Sanchez et al. 2002), e.g., bond centers and other high-symmetry interstices, which can be verified through examination of the charge-density isosurface plots (see Fig. 3). These localized charge pockets would require a certain spatial arrangement among them for the total energy to be well-minimized. By contrast, if the valence charge density in the interstices is completely delocalized, then no such constraints would exist, and the energy barrier to shear would arise mainly from a misfit-volume effect. In FCC Cu and Ag, when the local interstice volumes completely recover their equilibrium values at the intrinsic stacking fault, the energy penalties are very low despite the "wrong" bond angles (Ogata et al. 2002). Consequently, a rather general interpretation of our results is that, for the rearrangement of charge
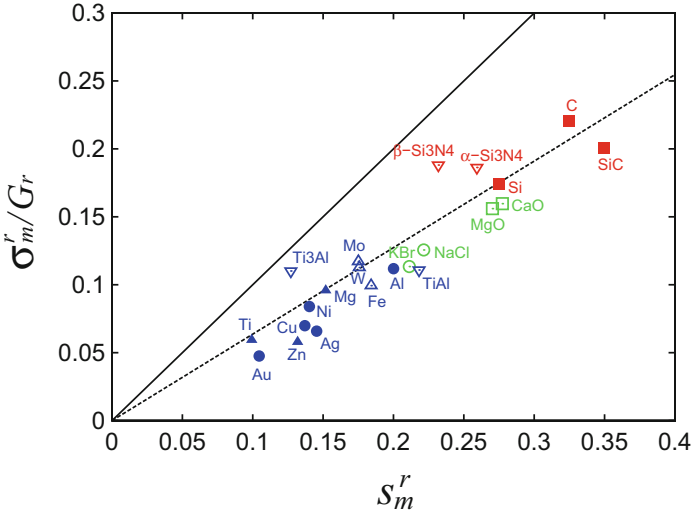
Fig. 1 Shearability (relaxed) (Ogata et al. 2004). The solid line indicates a unit slope, whereas the dashed line corresponds to a slope of $2/\pi$

density in response to mechanical deformation, bond-angle dependence brings about geometric constraints on the atomic configurations above and beyond the volumetric constraints. In the relaxed s_m^r, σ_m^r distributions (Fig. 1), ionic ceramics lie midway between directionally bonded metals and covalent solids; α- and β-Si$_3$N$_4$, being more ionic than SiC, also lie within this range (Ogata et al. 2003). However, in unrelaxed shear (Fig. 2), these solids manifest abnormally large ideal shear strains s_m^u and stresses σ_m^u, which are attributed to the bare Coulomb repulsion between like-charge ions, as in a simple Madelung sum model. In an atomic environment such as a crack tip, the surrounding medium would not allow for either fully relaxed or fully unrelaxed local shear. This result implies that ionic materials could be either less or much more brittle than covalent materials, depending on the subsidiary deformation constraints; this situation is analogous to the distinction of plane-stress vs. plane-strain loading conditions in the fracture of metals. Another noteworthy feature of Figs. 1 and 2 is the approximately universal linear scaling between s_m and σ_m/G across a range of crystal structures, bond types, and slip systems. The original Frenkel model (Frenkel 1926), which contains a single parameter G, is well known and widely used (Peierls 1940; Nabarro 1947; Rice 1992; Kysar 2003) in an empirical fashion. Various researchers have noted that, on the basis of physical intuition, a more realistic description is to treat the peak position of $\sigma(s)$ as an adjustable parameter (Foreman et al. 1951; Xu et al. 1995) rather than as a fixed value at $b/4d_0$. As shown here, on the basis of ab initio first-principles computation results, a two-parameter representation

$$\sigma = \frac{2Gs_m}{\pi} \sin\left(\frac{\pi s}{2s_m}\right), 0 < s < s_m, \quad \sigma_m = \frac{2Gs_m}{\pi} \tag{3}$$
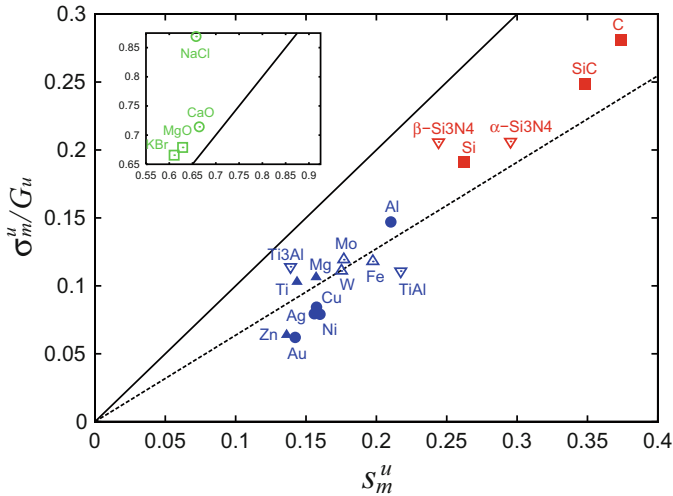
Fig. 2 Shearability (unrelaxed) (Ogata et al. 2004). The solid line indicates a unit slope, whereas the dashed line corresponds to a slope of $2/\pi$

with the shear modulus G and the shearability s_m as fundamental materials parameters provides an accurate description of simple metals and ceramics. As shown in Figs. 1 and 2, the slope of $2/\pi$ implied by our proposed extension of the Frenkel model (Eq. 3) indeed represents the data well. In addition, the fundamental constitutive behavior for shear deformation can be captured in a master curve in terms of normalized stress $\tilde{\sigma} \equiv \sigma/Gs_m$ and strain $\tilde{s} \equiv s/s_m$, as shown in Fig. 4. In this rescaling, all curves have an initial unit slope and reach a maximum at $\tilde{s} = 1$. The behavior labeled as "Frenkel (renormalized)" reflects a universal shear-softening response for $s < s_m$. Notably, this new, "renormalized Frenkel" model is the shear counterpart to the universal binding energy relation (Rose et al. 1983), which also has two parameters and has been quantitatively checked against ab initio computation results. Taken together, the renormalized Frenkel model and the universal binding energy relation enable the formulation of materials design and performance criteria in which tensile and shear dissipation modes compete (Rice 1992; Xu et al. 1995; Li et al. 2002; Vitos et al. 2003; Kysar 2003). For example, the "brittleness parameter" of Rice (1992), which compares the unstable stacking energy γ_{us} to the surface energy γ_s, can be very crudely estimated as

$$\beta \equiv \frac{\gamma_{us}}{\gamma_s} \propto \frac{Gs_m^2}{Bt_m^2} = \left(\frac{G}{B}\right) \cdot \left(s_m^2\right) / \left(t_m^2\right) \tag{4}$$

by scaling arguments (Ogata et al. 2004; Ogata and Li 2009). Parameter G/B is accessible experimentally and has been used as a brittleness predictor. However, s_m and t_m, although easy to obtain in ab initio first-principles computations (Černy et al. 2003), are unavailable experimentally despite having been theoretically established
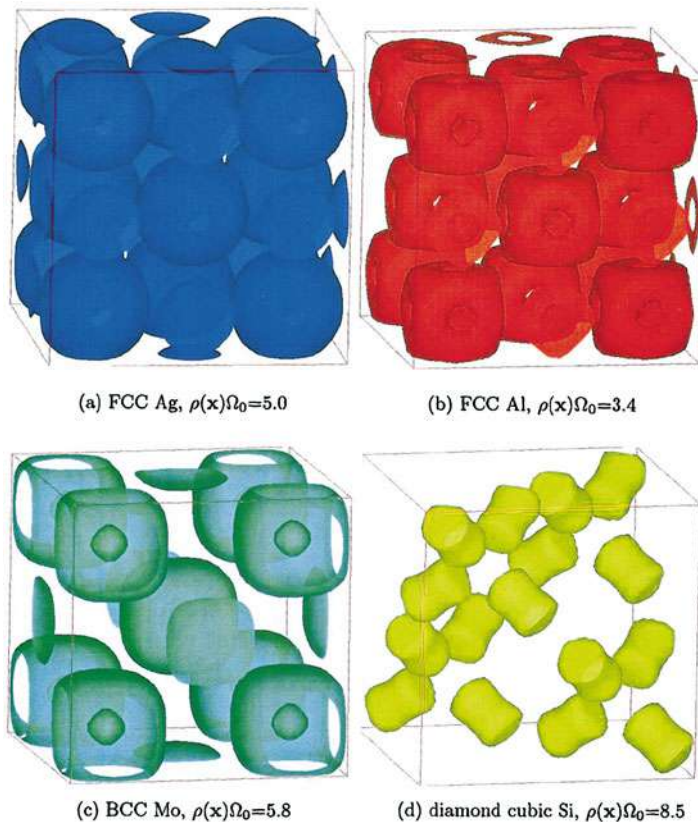
Fig. 3 Isosurface plot of valence charge density of elements at their stress-free states (Ogata et al. 2004): (**a**) FCC Ag, (**b**) FCC Al, (**c**) BCC Mo, and (**d**) diamond cubic Si. Ω_0 is the atomic volume

to be important (Foreman et al. 1951; Xu et al. 1995). The results indicate that a wide gap in β exists between metals and ceramics because G/B and s_m are governed not only by the crystal structure but also by the nature of chemical bonding (e.g., $s_m^r = 0.105$ in FCC Au vs. $s_m^r = 0.200$ in FCC Al) and by the loading condition (e.g., $s_m^r = 0.221$ vs. $s_m^u = 0.658$ in B1 NaCl). By contrast, the relative variation of t_m is less sensitive than that of s_m, and it has no spectral gap (see Fig. 5). The results suggest that the brittleness is governed by the shearability (Ogata and Li 2009).

3 Intrinsic Strength of Crystal Lattice

In the previous section, lattice strain, such as the shear in a common slip system, and triaxial tension were applied to primitive cells to examine the shearability and tensibility of chemical bonding. Here, we more generally consider the intrinsic
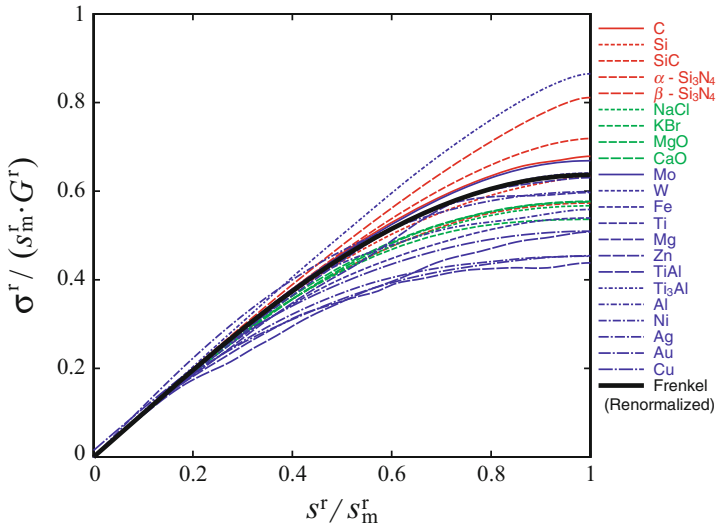
Fig. 4 Relaxed shear stress-strain curves for 22 materials and a renormalized Frenken model (Ogata et al. 2004)
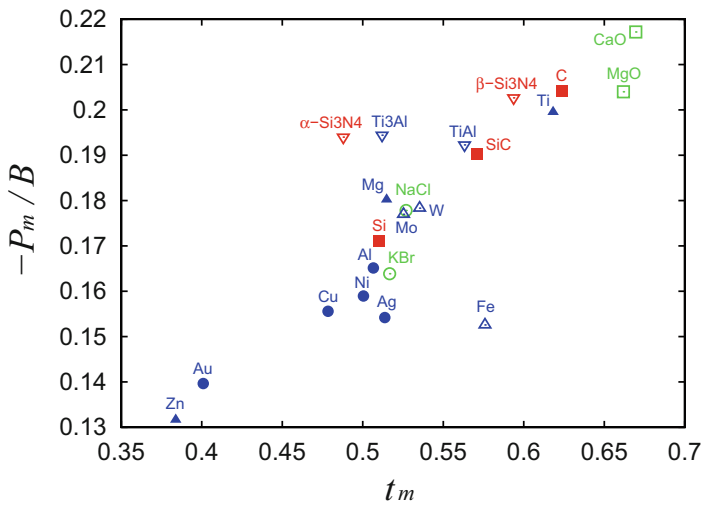
Fig. 5 Tensibility (Ogata et al. 2004)

strength of the crystal lattice, which is defined as the critical stress that induces lattice instability on a minimum deformation unit. Lattice instability sometimes leads to shear deformation (e.g., dislocation or deformation twinning), sometimes to decohesion (e.g., cracks), and sometimes to a solid-solid transformation (e.g., martensitic or Bainitic transformation). Because the deformation, decohesion, and

transformation instability modes, hereafter referred to simply as "deformation," have different critical stresses, a first-principles prediction of the critical stress enables the construction of a deformation mechanism map of the crystal lattice as a function of stress. Unlike the analysis of strength of chemical bonding, we cannot always use a primitive cell because the deformation has different minimum deformation units (irreducible lattice correspondence pattern), such as in the case of deformation twinning of HCP Mg, which is discussed later. Thus, generally, the minimum deformation unit differs from and is larger than the primitive cell. That is, the deformation unit may have more internal degrees of freedom independent of lattice strain. To compute the critical stress at zero temperature from first principles, one can actually apply various stresses to the supercell of the minimum deformation unit and detect the lattice instability using, for example, soft mode analysis (Qi and Chrzan 2014). In this case, the task is no longer technically difficult. Importantly, the internal degrees of freedom, so-called non-affine atomic "shuffling," cannot be controlled independently from the lattice strain induced by stress because the internal degrees of freedom are "slaves" of the affine lattice strain. However, in the meantime, actual deformation occurs at finite temperatures. The thermal energy induces thermal fluctuations that can independently disturb the internal degrees of freedom from the lattice strain, that is, free from slavery and excite the internal degrees of freedom. Moreover, the thermal energy paves the way for untouched instability points (saddle points) on the Gibbs free energy landscape that cannot be taken only by stress, where the soft mode is mostly related to corresponding non-affine atomic shuffling. The usual method of applying stress to the lattice at zero temperature may overlook these instability points. Drawing the Gibbs free energy landscape in a space spanned by lattice strain and shuffling degrees of freedom under different stresses is the best approach for observing all of the saddle points without omission. After the Gibbs free energy landscape has been computed, one can obtain not only the critical stress at zero temperature (ideal strength of lattice) but also the minimum energy pathway (MEP) and the activation Gibbs free energy of each deformation mode. Note that the critical stress is generally a temperature- and strain-rate-dependent quantity.

To determine the Gibbs free energy landscape with the non-affine atomic shuffling, one can separate the non-affine atomic shuffling with the instantaneous affine strain at the smallest lattice correspondence pattern level as the "shuffling" and "strain" variables and compute the Gibbs free energy landscape and MEP (Jónsson et al. 1998) of possible strain modes in this joint "shuffling"-"strain" space (Ishii et al. 2016). Let us define a scalar to represent the shuffling degree of freedom (Ishii et al. 2016):

$$I \equiv (s - s^{\text{ini}})^T \mathbf{H}_0^T \mathbf{H}_0 (s^{\text{fin}} - s^{\text{ini}})/M \qquad (5)$$

where s^{ini} is the internal coordinates of labeled atoms before deformation ($\lambda = 0$) and s^{fin} is the internal coordinates of labeled atoms after deformation ($\lambda = 1$) for a given σ. Notably, the s differences in (5) are the changes in internal coordinates without periodic boundary condition (PBC) wraparound. I has the unit of Å^2 and

takes the meaning of the mean-square non-affine displacements (MSDs) when the deformation is complete. The deformation transforms the crystal from a reference configuration to another configuration with strain $\boldsymbol{\varepsilon}_{\text{final}}$. The supercell describing the deformation can be taken as the irreducible lattice correspondence pattern, which can be larger than the primitive cell of the host lattice. The deformation can be represented atomistically by

$$\mathbf{x}_m(\lambda) = \mathbf{H}(\lambda)\mathbf{s}_m(\lambda), \quad \mathbf{H}(\lambda) = \mathbf{R}(\lambda)(\mathbf{I} + 2\boldsymbol{\varepsilon}(\lambda))^{1/2}\mathbf{H}_0, \tag{6}$$

where λ is the reaction progress variable (scalar), $\mathbf{x}_m$ is the Cartesian position, and $\mathbf{s}_m = [s_{m1}; s_{m2}; s_{m3}] \in [0, 1)$ is the reduced coordinate vector of atom m under PBC; $\mathbf{H} = [\mathbf{h}_1\mathbf{h}_2\mathbf{h}_3]$ is a 3×3 matrix, where $\mathbf{h}_1$, $\mathbf{h}_2$, and $\mathbf{h}_3$ are three edge vectors of the supercell; and $m = 1..M$ is the index of atoms in the supercell. Equation components $(\mathbf{I} + 2\boldsymbol{\varepsilon}(\lambda))^{1/2}$ and $\mathbf{R}(\lambda)$ are the irrotational and rotational parts of the deformation gradient, respectively, where $\boldsymbol{\varepsilon}(\lambda)$ is the Lagrangian strain with respect to the initial configuration. The representation (6) is general in the sense that it can be used to represent liquid or solid geometries in molecular dynamics simulations that have general displacive and diffusive motions. Here, however, the integer M would be taken to be as small as possible. Although deformation in reality must be a heterogeneous and asynchronous transformation, the supercell of the irreducible M should still provide essential and key physics about the process (Ogata et al. 2005) and the strength of the lattice. We can then compute the MEP with the least M using parameter-free ab initio first-principles computation at a constant external stress $\boldsymbol{\sigma}$, which gives the activation Gibbs energy $G(\lambda, \boldsymbol{\sigma})$ vs. reaction coordinate λ, parametrized by $\boldsymbol{\sigma}$. An algorithm is applied to compute the MEP and to find the saddle point of deformation:

$$Q(\boldsymbol{\sigma}) \equiv G(\lambda^*, \boldsymbol{\sigma}) - G(0, \boldsymbol{\sigma}) \tag{7}$$

on the joint $\boldsymbol{\varepsilon} \otimes \mathbf{s}$ space (Sheppard et al. 2012), where $\lambda = 0$ is used to denote the state before deformation, $\lambda = 1$ to denote the state after deformation, and λ^* for a saddle point, at constant external stress $\boldsymbol{\sigma}$.

3.1 Strain-Controlled vs. Shuffling-Controlled

An interesting but very important question is whether the MEP or reaction coordinate is dominated by $\Delta\mathbf{s}$ or $\Delta\boldsymbol{\varepsilon}$. Two opposite physical transformation limits can be identified. In the diffusive limit (annealing transformation), $\boldsymbol{\varepsilon}_{\text{final}} \approx 0$; in addition, because of the randomized short-range diffusion that destroys labeled-atom registry (if identical atoms have labels), the lattice correspondence pattern is random, and M can be taken to be infinity. Thus, this deformation would be a pure "shuffling-dominated" deformation with a huge but unchanged (fixed-shaped) supercell. By contrast, many simple metals have one-atom primitive cells (e.g., FCC Cu, Au, and Al and BCC Mo and Fe) that deform at low temperatures with $M = 1$,

where the irreducible lattice correspondence pattern (before and after deformation) involves merely one atom, as evidenced by direct molecular dynamics simulations; thus, there is no shuffling degree of freedom, and the metal is purely strain-dominated. The former case is classified as civilian/diffusive transformation, the latter case is classified as military/displacive transformation (Christian 2002), and the two cases tend to manifest different sensitivities to temperature and strain rate (Suresh and Li 2008; Li 2007; Zhu and Li 2010). The civilian/diffusive deformation tends to be more temperature and strain-rate sensitive but less stress sensitive (less nonlinear-thresholding behavior) than military/displacive transformation because of the smaller activation volume (Li 2007; Zhu and Li 2010). The origin of this can, in turn, be traced to the lack of long-range elastic interactions, which tends to reduce the activation volume. Activation processes with long-range elastic interactions (Shen et al. 2008) tend to have much larger activation volume, reflecting greater collective coordination. If the same MEP and saddle-point energy can be obtained by continuously increasing just the shuffling and instantaneously relaxing the strain by steepest descent, we refer to the path as being "shuffling-controlled." Vice versa, if the same MEP and saddle-point energy can be obtained by continuously increasing the strain and relaxing the internal non-affine displacement by steepest descent, we refer to the path as being "strain-controlled."

Let us give a precise and general mathematical definition of the *strain-controlled* vs. *shuffling-controlled* concept for deformation as follows (Ishii et al. 2016). One can choose dominant strain components $\boldsymbol{\varepsilon} = \{\varepsilon_1, \cdots, \varepsilon_n; n \leq 6\}$, which can be taken to be strain components induced by the deformation. One can then increase $\boldsymbol{\varepsilon}$ slightly, fix it as a constraint, and then relax the remaining $3M + 6 - n$ degrees of freedom in ab initio first-principles supercell computation by the steepest descent method. The process is then repeated. If one can reach the same activation energy $G(\lambda^*)$ as the full-$\boldsymbol{\varepsilon} \otimes \mathbf{s}$ MEP search by trotting along $\boldsymbol{\varepsilon}$, this activation path can be called strain-controlled in the sense that the reaction coordinate can be well represented by $\boldsymbol{\varepsilon}$. By definition, $I = 0$ when $\boldsymbol{\varepsilon} = \mathbf{0}$. In an alternative calculation, one increases I slightly, fixes it as a constraint variable, relaxes the remaining $3M + 6 - n$ degrees of freedom by steepest descent, and then repeats trotting along I. In this case, all six strain tensor components become slaved variables of I: the supercell shape and size respond instantaneously to the tweaking of this internal shuffling motion. If one can reach the same saddle point $G(\lambda^*)$ as the full-$\boldsymbol{\varepsilon} \otimes \mathbf{s}$ path search, this activation path can be called shuffling-controlled in the sense that the reaction coordinate can be well represented by I. If neither $\boldsymbol{\varepsilon}$-control or I-control give the correct saddle energy, or if both do, we refer to such a transformation path as being under *mixed*-control.

3.2 A Case of Non-basal $\{10\bar{1}2\}\langle10\bar{1}\bar{1}\rangle$ Deformation Twinning in HCP Magnesium (Ishii et al. 2016)

Deformation twinning (DT) (Christian and Mahajan 1995) transforms a crystal into a mirrored configuration with transformation strain $\boldsymbol{\varepsilon}_{\text{final}}$. Numerous experimental
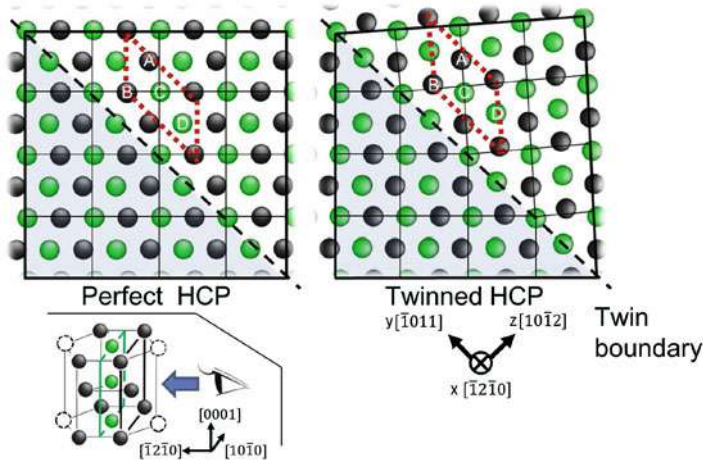
Fig. 6 Atomic arrangements and four-atom supercell shape of perfect HCP and twinned HCP configurations viewed from $[1\bar{2}10]$ (Ishii et al. 2016)

studies and theoretical analyses on the HCP $\{10\bar{1}2\}\langle10\bar{1}\bar{1}\rangle$ twinning system have been reported. Nevertheless, information relating to the required mechanical conditions and atomistic twinning processes remains controversial; in addition, certain aspects of and mechanisms involved have not been fully clarified, especially the issue of whether the twinning processes are atomic-shuffling dominant (Li and Ma 2009; Li and Zhang 2014) or shear-strain induced. The reason for this controversy appears to be a lack of fundamental energetics. The idea discussed in the previous section can be naturally applied to this issue, thus shedding light on a DFT-based Gibbs free energy landscape that can clarify the deformation mechanism. Figure 6 shows perfect ($\lambda = 0$) and twinned ($\lambda = 1$) HCP atomic structures viewed from $\langle1\bar{2}10\rangle$, at $\sigma = 0$. The four-atom parallelepiped supercell used in this study, which includes atoms A, B, C, and D, is indicated as a broken red line. The four-atom supercell ($M = 4$) is the minimum lattice correspondence pattern unit necessary to render the atomic arrangements during $\{10\bar{1}2\}\langle10\bar{1}\bar{1}\rangle$ twinning shear deformation (Li and Ma 2009; Wang et al. 2013). Notably, even though the Bravais primitive cell of HCP metals contains only $N = 1$ atom, because of the anisotropy (a_0 and c_0 are required to specify an HCP lattice, in contrast just a_0 for cubic metals) and low symmetry of the twinning system, M can be greater than 2. The internal energy U and Cauchy shear stress of the supercell with respect to the uniform twinning shear deformation $\gamma = \gamma_{yz} = 2\varepsilon_{yz}$;

$$U(\gamma) = \min_{s, \varepsilon \in \gamma} U(\varepsilon, s) \qquad (8)$$

is shown in Fig. 7a, b. The shear deformation has been applied to two different fully relaxed HCP supercells with HCP and twinned HCP configurations; these supercells are labeled as "original" (red plots in Fig. 7a, b) and "twinned" (green
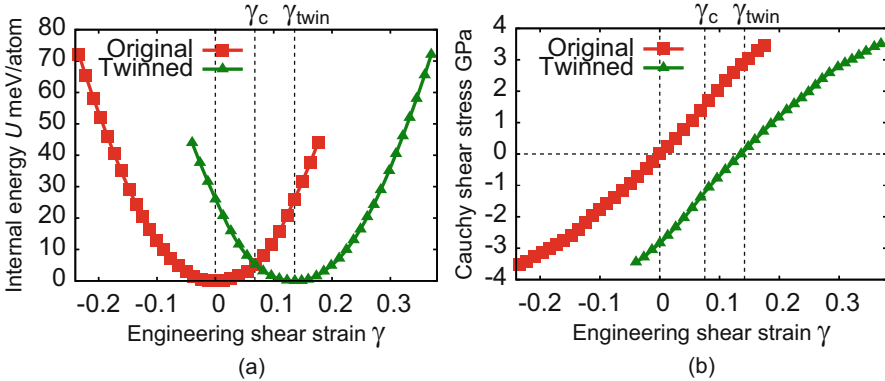
Fig. 7 Supercell internal energy (**a**) and Cauchy shear stress (**b**) with respect to engineering shear strain γ. The shear deformation is applied to two different initial supercell atomic configurations: "original" (red plots) and "twinned" (green plots). An energy crossover occurs at $\gamma = \gamma_c (= 0.068)$ (Ishii et al. 2016)

plots in Fig. 7a, b), respectively. Both of these supercells have the same perfect HCP lattice structure; however, the twinned supercell has the sheared supercell frame from the original supercell of $\gamma = \gamma_{yz} = 0.136$ engineering shear strain. Note that the internal atomic configurations in these supercells can be different even though they have the same supercell shape. The internal atomic configuration **s** at a certain engineering shear strain is dependent upon whether it has been sheared from the "original" or the "twinned" supercell. This dependence causes internal energy and stress differences at each engineering strain γ, shown as the red and green plots in Fig. 7a, b. In Fig. 7a, an energy crossover occurs at a critical $\gamma = \gamma_c (= 0.068)$ between these plots. At strains above the critical strain, an atomic configuration sheared from a twinned supercell becomes energetically favorable compared with that of the original supercell with the same cell shape, and thermodynamics would start to allow DT to occur at $T = 0$ K. However, even at a strain only slightly above the critical strain, a transition from the red to the green curve in Fig. 7a, b still appears to involve an energy barrier. That is, spontaneous twinning in the absence of thermal fluctuations and waiting times is not observed in the DFT calculation. This observation already hints that the twinning shear strain γ may not be the dominant reaction coordinate. Something internal "got stuck," preventing the red→green transition from occurring spontaneously, even at strains substantially greater than the equi-energy strain γ_c.

To examine this interesting behavior, we evaluated the Gibbs free energy landscape numerically using ab initio first-principles supercell energetics.

$$G(\boldsymbol{\varepsilon}, \mathbf{s}, \boldsymbol{\sigma}) \equiv U(\boldsymbol{\varepsilon}, \mathbf{s}) - W(\boldsymbol{\varepsilon}, \boldsymbol{\sigma}), \tag{9}$$

where $W(\boldsymbol{\varepsilon}, \boldsymbol{\sigma})$ is the work done by constant external Cauchy stress $\boldsymbol{\sigma}$ (Wang et al. 1995):

$$W(\boldsymbol{\varepsilon}, \boldsymbol{\sigma}) \equiv \int_0^1 dl \, \det |\mathbf{J}(\boldsymbol{\eta} = l\boldsymbol{\varepsilon})\mathbf{H}_0| \times \mathrm{Tr}\left[\mathbf{J}^{-1}(\boldsymbol{\eta} = l\boldsymbol{\varepsilon})\boldsymbol{\sigma}\mathbf{J}^{-T}(\boldsymbol{\eta} = l\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\right], \quad (10)$$

where $\boldsymbol{\eta} = l\boldsymbol{\varepsilon} = 1/2(\mathbf{J}^T\mathbf{J} - \mathbf{I})$ is the Lagrangian strain tensor, $\mathbf{J}$ is the corresponding deformation gradient tensor,

$$\mathbf{J} = \mathbf{R}(\mathbf{I} + 2\boldsymbol{\eta})^{1/2} \quad (11)$$

and $\mathbf{R}$ is an additional rotation matrix $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ that is completely defined when the transformation coordinate frame convention is chosen. Even though $G(\boldsymbol{\varepsilon}, s, \boldsymbol{\sigma})$ is now defined by (9), it is difficult to visualize directly because $\boldsymbol{\varepsilon} \otimes s$ space is $3M + 6$ dimensional. Thus, (5) can be used to aid visualization. By applying energy minimization to all degrees of freedom of the supercell system other than γ and I, we can uniquely compute $G(\gamma, I, \boldsymbol{\sigma})$:

$$G(\gamma, I, \boldsymbol{\sigma}) \equiv \min_{\boldsymbol{\varepsilon} \in \gamma, s \in I} G(\boldsymbol{\varepsilon}, s, \boldsymbol{\sigma}) \quad (12)$$

The computed Gibbs free energy landscapes $\Delta G(\gamma, I)$ at different external Cauchy shear stresses ($\sigma_{yz} = \sigma_{zy} = 0, 1$ and 2 GPa along the twinning direction) are shown in Fig. 8. The other Cauchy stress components were set to zero. The red curves on the Gibbs free energy landscapes represent the MEPs from the original to the twinned configuration under these external shear stresses, which were estimated by applying the nudged elastic band (NEB) method (Jónsson et al. 1998) to obtain the two-dimensional Gibbs free energy landscapes. The Gibbs free energy change along the MEP under different external shear stresses is shown in Fig. 10a, b shows the change in the Gibbs free energy barrier with respect to the external shear stress. The following two NEB calculations were independently performed: (1) an NEB calculation with respect to the internal atomic configuration I, in which the supercell frame is relaxed under the predefined external stress (I-control NEB), and (2) an NEB calculation with respect to γ, in which internal atomic configuration s is relaxed for each supercell frame shape (γ-control NEB). A comparison between their Gibbs free energy barriers is presented in Fig. 10b; the Gibbs free energy barrier determined by I-control NEB agrees well with the Gibbs free energy barrier based on the two-dimensional Gibbs free energy landscape, whereas the Gibbs free energy barrier obtained by the γ-control NEB is substantially higher. This result indicates that the DT is an I-dominant (non-affine displacement dominant) deformation. Thus, if an assist of thermal energy were to exist, the twinned structure could be generated without generating local shear strain because the phonons could toggle the "internal cog." After the "internal cog" is flipped, the voxel γ will later spontaneously relax toward the twinned configuration. The MEP analysis based on the Gibbs free energy landscape also clearly indicates that a uniform supercell shape change and an atomic position change simultaneously occur during the DT process. Figure 9 represents the change in atomic position along the MEP at a shear stress of 0 GPa. A "dumbbell flipping" or "shuffling" behavior – that is, staggered rotations of
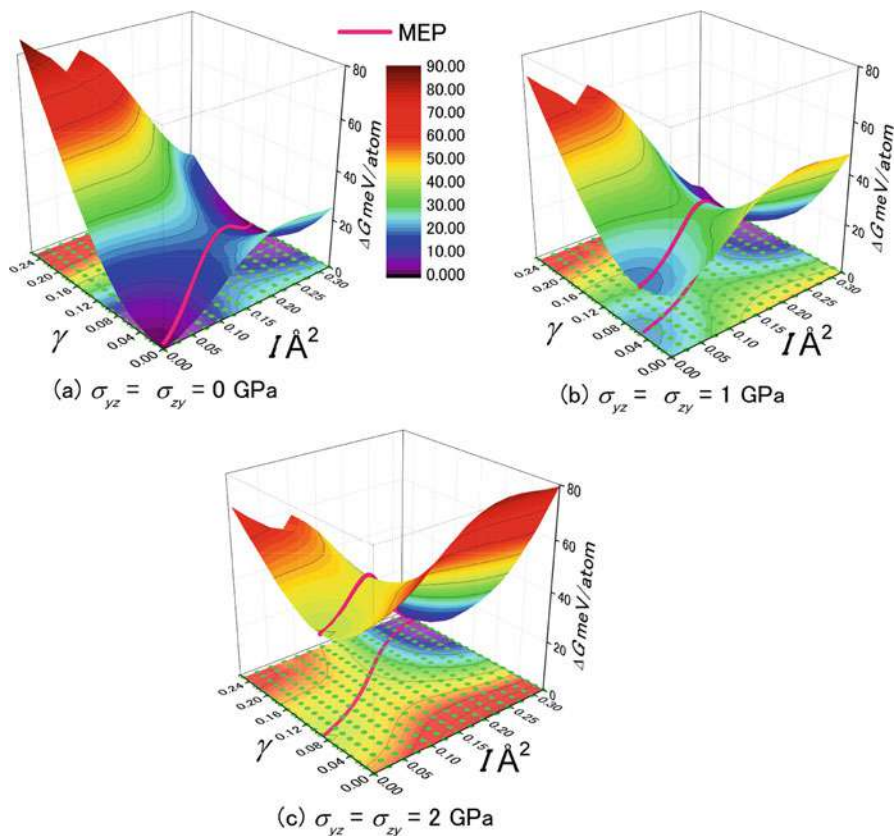
Fig. 8 Gibbs free energy distribution $\Delta G(\gamma, I)$ at different levels of external shear stress $\sigma_{yz} = \sigma_{zy} = 0$ GPa (**a**), 1 GPa (**b**), and 2 GPa (**c**). The red curves represent the MEPs of the deformation twinning processes (Ishii et al. 2016)
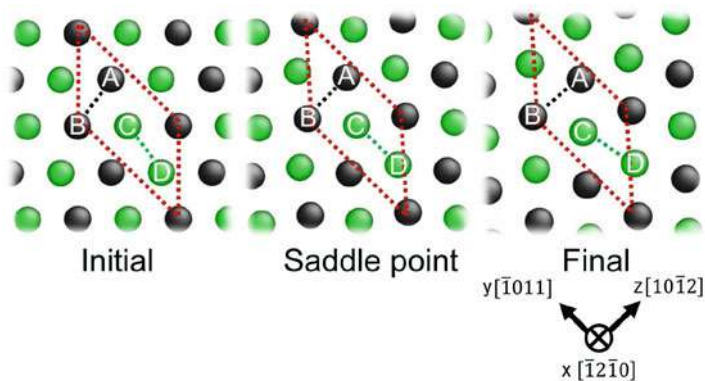
Fig. 9 Atomic configuration change along the MEP (Ishii et al. 2016)
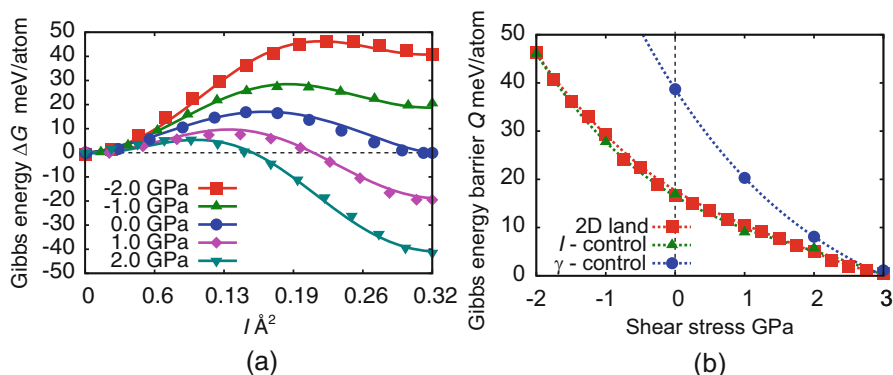
Fig. 10 Gibbs free energy variation along the MEP (**a**) and the energy-barrier change with respect to external shear stress (**b**). In (**a**), the solid curves and dots represent results obtained from the Gibbs free energy landscape and direct NEB analysis, respectively. In (**b**), the square, triangular, and circular symbols represent the Gibbs free energy barriers obtained from the two-dimensional Gibbs free energy landscape (2D land), I-control NEB, and γ-control NEB, respectively (Ishii et al. 2016)

the A-B and C-D atomic bonds – is observed together with uniform supercell shear deformation. Figure 10b also shows that the Gibbs free energy barrier decreases with increasing applied external shear stress and that an athermal condition is eventually achieved at a critical external shear stress of $\sim$3.0 GPa. Thus, applying an external shear stress of 3.0 GPa can induce the DT regardless of the internal atomic configuration I. Thus, the DT becomes γ-dominant (shear-strain induced) under such an extremely high shear stress.

4   Conclusions

The first-principles computations of primitive crystal lattices revealed the fundamental mechanical properties of chemical bonding in crystals from the standpoint of shearability and tensibility. The simplified computational framework enabled us to perform vast numbers of computations for different crystal lattices. Thus, the computational approach may lead to a base of high-throughput ultra-strength and ultra-toughness structural alloys designed from first-principles, although extrinsic factors will need to be considered for a more accurate quantitative analysis. The first-principles computation of intrinsic strength of crystal lattice based on Gibbs free energy landscape in the shuffling-strain joint space redefines the strength of a crystal lattice, which is not only stress but also temperature dependent; moreover, it depicts untouched saddle points that cannot be accessed only by applying stress or strain. The computation framework judges whether the lattice deformation is shuffling-controlled or strain-controlled. The former is more temperature sensitive, and the latter is more stress sensitive. This computational approach may be very useful for

lattice deformations with a more shuffling-contorted aspect, such as deformation twinning in HCP and solid-solid transformations. Although these studies provide valuable fundamental information about the strength of materials, which could be useful in high-throughput materials design, numerous tasks remain to enable the consideration of extrinsic factors for more precise predictions of the mechanical properties of actual structural materials. A sophisticated spatial- and temporal-scale coarse-grained framework (Zhao et al. 2013) based on the given first-principles intrinsic mechanical properties is desired because, in reality, deformation must be heterogeneous and asynchronous.

References

Černy M, Pokluda J, Šob M, Friák M, Šandera P (2003) Ab initio calculations of elastic and magnetic properties of Fe, Co, Ni, and Cr crystals under isotropic deformation. Phys Rev B 67:035116

Christian JW (2002) The theory of transformations in metals and alloys. Elsevier, Amsterdam

Christian JW, Mahajan S (1995) Deformation twining. Prog Mater Sci 39:1–157

Foreman AJ, Jaswon MA, Wood JK (1951) Factors controlling dislocation widths. Proc Phys Soc London Sect A 64:156

Frenkel J (1926) Zur Theorie der Elastizitatsgrenze und der Festigkeit kristallinischer Körper. Z Phys 37:572–609

Ishii A, Li J, Ogata S (2016) Shuffling-controlled versus strain-controlled deformation twinning: the case for hcp Mg twin nucleation. Int J Plas 82:32–43

Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Berne BJ, Ciccotti G, Coker DF (eds) Classical and quantum dynamics in condensed phase simulations. World Scientific, Singapore, pp 385–404

Krenn CR, Roundy D, Morris JWJ, Cohen ML (2001) Ideal strengths of bcc metals. Mater Sci Eng A 111:319–321

Kysar JW (2003) Energy dissipation mechanisms in ductile fracture. J Mech Phys Solids 51:795–824

Li B, Ma E (2009) Atomic shuffling dominated mechanism for deformation twinning in magnesium. Phys Rev Lett 103:035503

Li B, Zhang X (2014) Global strain generated by shuffling-dominated twinning. Scr Mater 71:45–48

Li J, Van Vilet KJ, Zhu T, Yip S, Suresh S (2002) Atomistic mechanisms governing elastic limit and incipient plasticity in crystals. Nature 418:307–310

Li J (2007) The mechanics and physics of defect nucleation. MRS Bull 32:151–159

Mori-Sanchez P, Pendas AM, Luana V (2002) A classification of covalent, ionic, and metallic solids based on the electron density. J Am Chem Soc 124:14721–14723

Nabarro FRN (1947) Dislocations in a simple cubic lattice. Proc Phys Soc London 59:256–272

Ogata S, Li J (2009) Toughness scale from first-principles. J Appl Phys 106:113534

Ogata S, Li J, Yip S (2002) Ideal pure shear strength of aluminum and copper. Science 298:807–811

Ogata S, Hirosaki N, Kocer C, Shibutani Y (2003) An ab initio study of the ideal tensile and shear strength of single-crystal β-Si_3N_4. J Mater Res 18:1168–1172

Ogata S, Li J, Hirosaki N, Shibutani Y, Yip S (2004) Ideal shear strain of metals and ceramics. Phys Rev B 70:104104

Ogata S, Li J, Yip S (2005) Energy landscape of deformation twinning in bcc and fcc metals. Phys Rev B 71:224102

Peierls R (1940) The size of a dislocation. Proc Phys Soc London 52:34–37

Qi L, Chrzan DC (2014) Tuning ideal tensile strength and intrinsic ductility of bcc refractory alloys. Phys Rev Lett 112:115503

Rice JR (1992) Dislocation nucleation from a crack tip: an analysis based on the Peierls concept. J Mech Phys Solids 40:239–271

Rose JH, Smith JR, Ferrante J (1983) Universal features of bonding in metals. Phys Rev B 28:1835–1845

Roundy D, Cohen ML (2001) Ideal strength of diamond, Si, and Ge. Phys Rev B 64:212103

Roundy D, Krenn CR, Cohen ML, Morris JWJ (2001) The ideal strength of tungsten. Philos Mag A 81:1725–1747

Shen C, Li J, Wang Y (2008) Finding critical nucleus in solid-state transformations. Metall Mater Trans 39:976–983

Sheppard D, Xiao P, Chemelewski W, Johnson DD, Henkelman G (2012) A generalized solid-state nudged elastic band method. J Chem Phys 136:074103

Silvi B, Gatti C (2000) Direct space representation of the metallic bond. J Phys Chem A 104: 947–953

Suresh S, Li J (2008) Deformation of the ultra-strong. Nature 456:716–717

Vitos L, Korzhavyi PA, Johansson B (2003) Stainless steel optimization from quantum mechanical calculations. Nat Mater 2:25–28

Wang J, Li J, Yip S, Phillpot S, Wolf D (1995) Mechanical instabilities of homogeneous crystals. Phys Rev B 52:12627–12635

Wang J, Yadav SK, Hirth JP, Tomé CN, Beyerlein IJ (2013) Pure-shuffle nucleation of deformation twins in hexagonal-close-packed metals. Mater Res Lett 1:126–132

Xu G, Argon AS, Ortiz M (1995) Nucleation of dislocations from crack tips under mixed modes of loading: implications for brittle against ductile behaviour of crystals. Philos Mag A 72:415–451

Zhao P, Li J, Wang Y (2013) Heterogeneously randomized STZ model of metallic glasses: softening and extreme value statistics during deformation. Int J Plast 40:1–22

Zhu T, Li J (2010) Ultra-strength materials. Prog Mater Sci 55:710–757

Zhu T, Li J, Ogata S, Yip S (2009) Mechanics of ultra-strength materials. MRS Bull 34:167–172

Atomistic Simulations of Fracture and Fatigue in Nanotwinned and Amorphous Materials

76

Xiaoyan Li, Xuan Zhang, and Huajian Gao

Contents

Abstract

Atomistic simulation has been recognized as a powerful tool to investigate and predict the mechanical properties and behaviors of nanostructured and amorphous materials. In this chapter, we first introduce some basic methodologies, including the setups for simulations of fracture and fatigue, and how to characterize the resistance to fracture and fatigue based on the information from atomistic simulations. Then we highlight some recent studies about large-scale atomistic simulations for fracture and fatigue of nanotwinned and amorphous

X. Li · X. Zhang
Center for Advanced Mechanics and Materials, Applied Mechanics Laboratory, Department of Engineering Mechanics, Tsinghua University, Beijing, China

H. Gao (✉)
School of Engineering, Brown University, Providence, RI, USA
e-mail: huajian_gao@brown.edu

materials (such as metallic glasses and lithiated silicon). At the same time, we further address the mechanistic insights to damage initiation around crack tip, interaction between crack and various defects, and origins of brittle and ductile behaviors. These results from atomistic simulations help provide a fundamental understanding of fracture and fatigue behaviors of nanotwinned and amorphous materials, which facilitates the design and fabrication of new materials with excellent mechanical properties and performance.

1   Introduction

Fracture is one of the most common failure modes of materials with or without pre-existing cracks under static and dynamic strain, and fatigue usually refers to the deformation and failure of materials under cyclic loads. Investigating the fracture and fatigue behaviors of materials is crucial for the reliability and safety of materials and structures in many practical applications.

Over the past decades, a number of novel nanostructured and amorphous materials (such as nanocrystalline and nanotwinned materials, metallic glasses, and lithiated silicon) have emerged as a result of rapid advances in fabrication and processing technologies. In contrast to conventional crystalline materials with characteristic sizes above tens of micrometers, nanostructured and amorphous materials can often achieve extraordinary mechanical properties owing to their unique microstructures, including higher strength, better ductility, higher fracture toughness, and more remarkable resistance to fatigue, corrosion, and wear. Several comprehensive reviews of experimental studies on the fracture and fatigue behaviors of these new materials can be found in the literature (Kumar et al. 2003; Ovid'ko 2007; Sun and Wang 2015; Zhu and Li 2010; Li et al. 2016; Pineau et al. 2016a, b, c; McDowell et al. 2016). However, due to various limitations in testing and visualization techniques, it is still difficult or even impossible to determine the deformation processes or mechanisms of materials at the smallest characteristic scales (usually the nanoscale) through experiments alone.

In recent years, atomistic simulations have become increasingly recognized as a powerful tool to investigate the deformation and failure of materials (Yip 2005; Li and Gao 2015), including both crystalline and amorphous materials. Due to their intrinsic atomic-scale resolution, atomistic simulations are often used to complement experiments in probing the prevailing deformation and failure mechanisms in nanostructured and amorphous materials. On the other hand, an amount of caution is needed to interpret the simulation results in view of the large disparities in sample sizes and loading rates between simulations and experiments, even though continuous advances in supercomputers and the development of efficient algorithms (multiscale modeling, machine learning methods, etc.) have to some extent reduced such gaps in recent years. Inspiringly, the deformation and failure mechanisms revealed by atomistic simulations can be validated at overlapping length scales through direct in situ observations using a scanning/transmission electron microscope (SEM/TEM) (Li and Gao 2015). Currently, atomistic

simulations are broadly applied to explore deformation and failure behaviors and the associated mechanisms of nanostructured and amorphous materials (Li and Gao 2015). Atomistic simulations not only provide new mechanistic insights into the microstructural evolution and interaction between various defects but also shed light on the structure-property-processing relationships of new materials. The results and insights from atomistic simulations have profound implications on the design and fabrication of new materials with excellent mechanical properties and performance.

In this section, we first introduce some fundamental methodologies for simulations of fracture and fatigue and for the extraction of mechanical information or quantities (characterizing the resistance to fracture and fatigue) from atomistic simulations. We highlight some recent studies using atomistic simulations to examine the fracture and fatigue of nanotwinned and amorphous materials (such as metallic glasses and lithiated silicon) with a number of illustrative examples. These atomistic simulations can offer an atomically resolved view of complex fracture and fatigue processes, as well as the underlying failure mechanisms. Moreover, we offer some mechanistic insight into damage initiated around the tip of a crack, the interaction between a crack and various defects, and the origins of brittle and ductile behaviors, which improves the fundamental understanding of the fracture and fatigue of new nanostructured and amorphous materials.

2   Methodology

Studies on the fracture of materials have mainly focused on crack propagation or growth under an applied mechanical loading. Therefore, when one constructs fracture simulations, a pre-existing crack should be first introduced in the simulated samples. Appropriate loading and boundary conditions are then applied. As the applied loading varies, mechanical information (such as atomic stress, atomic strain, crack advance, and applied stress intensity factor) related to crack propagation is recorded and used to quantify the relationships among crack propagation, microstructure, and applied load. At the same time, some mechanistic details (such as the interaction between the crack and other defects) of crack propagation are also captured. In this part, we briefly introduce the basic procedure and setup for simulations of fracture and fatigue. For fracture simulations, we also introduce several methods to calculate the fracture energy or toughness of materials according to information extracted from atomistic simulations.

Based on the relaxation method, atomistic simulations can be generally classified into two groups: molecular statics (MS) simulations and molecular dynamics (MD) simulations. During MS simulations, the system is relaxed based on an energy minimization algorithm (such as conjugate gradient or second-order Newton-Raphson). MS simulations can reflect the quasi-static evolution of the system under a slowly varying strain but are not able to account for the effects of finite temperature. In contrast, during MD simulations, the system is relaxed by involving the actual motion of the atoms, which is determined by numerically solving Newton's equations of motion within a given time step. This allows MD simulations

to study the evolution of the dynamics and thermodynamics and the response of the system at finite temperatures. However, MD simulations have the disadvantage that the time step is around one femtosecond; thus, the total simulation time is generally limited to several nanoseconds, leading to an extremely high strain rate which is 7–11 orders of magnitude higher than those used in typical experiments. Fortunately, in many cases such high strain rate does not have a significant effect on the underlying deformation and failure mechanisms of materials. In parts 3 and 4, we introduce some recent MD simulations on the fracture and fatigue of nanotwinned and amorphous materials to highlight the success of atomistic simulations. When introducing fatigue in nanotwinned metals, we will also compare the simulation results between the MS and MD techniques.

2.1   Simulation Setup for Fracture

During fracture simulations, we are mainly interested in the failure and damage from crack propagation and advancement. Therefore, a pre-existing crack is introduced in the initial configurations of the simulated samples. Generally, there are two methods to introduce a pre-existing crack. The first method is to create a crack or notch by removing a wedge of atoms, while the second method is to generate a crack by preventing interatomic interaction across the crack surface. The crack generated by the second method has an atomically sharp tip. The initial crack is generally located at the center or edge of a simulated sample to achieve the center or edge cracked configuration.

   During fracture simulations, applying a loading on the crack is a crucial step for the overall simulation. Mode I is the dominant fracture mode in practical fracture cases; thus, most simulations primarily model mode I crack propagation, where the applied tensile loading is normal to the crack surface. There are three methods to achieve mode I loading. The first method is that during each loading step, a uniform tensile strain is applied, and then the overall system is relaxed with a fixed boundary along a direction normal to the tensile direction. Such loading step is repeated until the simulation is terminated. In the second method, one edge along the tensile direction of the sample is fixed, while another edge is stretched at a given constant velocity. To avoid stress waves induced at the onset of the constant velocity loading at one edge, a linear velocity profile is imposed on all atoms at the beginning of the simulation. Throughout the simulation, atoms on both edges are subjected to a controlled velocity, while all other atoms are free to move according to Newton's equations of motion. The third method is to apply a K-field to the simulated sample (hence often referred to as the K-controlled method). Figure 1 shows the schematic illustration of mode I loading on a simulated sample with a pre-existing crack. During the simulation, mode I loading is accomplished by increasing the stress intensity factor K in a very small increment ΔK. Such increment is usually $K_{IC}/100$ or less to ensure the iteration accuracy and convergence, where K_{IC} is the critical stress intensity factor for crack initiation. During each loading step, all the atoms in the fixed region in Fig. 1 are first positioned according to the continuum
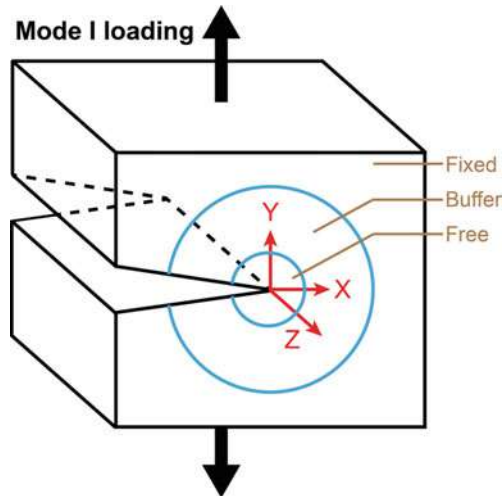
elasticity solution (i.e., asymptotic solution) of the crack tip displacement field with
the specified K, while all interior atoms in the free region are relaxed via the MS or
MD technique. The applied K-field thus serves as the boundary conditions for the
free region. To avoid the influence of the K-field on the crack tip plastic behavior,
the fixed region is required to be far from the crack tip, which indicates small-scale
yielding conditions, i.e., the size of the simulated sample in Fig. 1 is large enough
to accommodate the plastic zone around the crack tip. If the simulated sample is
relatively small, a buffer region needs to be introduced between the fixed and free
regions (Yip 2005), as shown in Fig. 1. Atoms in the buffer region are controlled
according to the forces from atoms in the free region (Yip 2005), which to some
extent avoid the influence of the fixed region of the crack tip (Yip 2005). The first
and second methods are simple and can be realized easily during fracture simulation.
In contrast to the first and second methods, the third method is more complex but
could be more effective in determining the fracture resistance curve and the fracture
toughness K_{IC}.

Atomistic simulations not only provide mechanistic details of the fracture
process but can also be used to estimate the fracture toughness. Here, we introduce
four methods to calculate the fracture toughness based on the results of atomistic
simulations. The first method is to calculate the energy release rate and is usually
used for the brittle materials. The critical energy release rate G_c can be calculated
by comparing the initial and final cracked configurations as follows:

$$G_c = -\frac{\Delta U - \Delta W}{t \Delta a} \tag{1}$$

where ΔU is the change in the strain energy and ΔW the change in the external
work associated with crack extension Δa, and t is the thickness of the simulated
sample; ΔU can be estimated based on the elastic stress and strain states of

the initial and final configurations, while ΔW can be calculated by integrating the force-displacement curves obtained from simulations. This method has been used to calculate the G_c of nanocrystalline graphene, with result consistent with experimental measurements (Jung et al. 2015).

The second method is to estimate the fracture energy by integrating the stress-strain curves for crack propagation. For a strip with an edge crack, the fracture energy for crack propagation can be calculated by the following expression (Freund 1990):

$$G = 2H \int_0^{\varepsilon_c} \sigma \, d\varepsilon \tag{2}$$

where H is the half width of the strip and ε_c is the critical strain at which the crack starts to propagate. The integration in Eq. (2) represents the area under the stress-strain curve, which can be obtained from the simulations. For brittle materials, Eq. (2) can be reduced to a linear elastic stress-strain relationship,

$$G = HE\varepsilon_c^2 \tag{3}$$

where E is the Young's modulus.

The third method is to obtain the critical stress intensity factor K_{IC}. During the fracture simulation, if the applied loading is achieved by imposing a K-field, K_{IC} can be easily taken as the stress intensity factor at crack initiation. This method has been used to predict the fracture toughness of lithiated silicon (Khosrownejad and Curtin 2017). The predictions for amorphous $Li_{0.5}Si$ and $LiSi$ from atomistic simulations are comparable to experimental measurements (Khosrownejad and Curtin 2017).

The fourth method is to extract the interfacial fracture energy based on a cohesive-zone-volume-element (CZVE) model (Yamakov et al. 2006). This method is only for interfacial cracks. The CZVE model is very similar to the continuum cohesive zone model and can provide the traction-separation relationship through a statistical average of the atomic information (such as atomic stress and displacement) around the interfacial crack. Before the simulations, the regions on both sides of the interface need to be sliced into a large number of square CZVEs. The dimension of each CZVE is usually set to 1 nm. During the simulations, the traction of each element is the average of the atomic stress of all atoms in the element, while the separation (i.e., opening displacement of the crack) is estimated by calculating the displacement of the mass point of the element. In the process of crack opening and propagation, the traction and separation of each CZVE are averaged at each given time interval to avoid thermal fluctuations. A traction-separation curve is obtained through the statistics of all data on the traction and separation of each CZVE. Finally, the interfacial energy is calculated by integrating over the traction-separation curve. This method has been used to capture the brittle and ductile fracture of an interfacial crack along a high-angle grain boundary in aluminum (Yamakov et al. 2006), as well as to calculate the interfacial fracture energy between lithiated silicon and amorphous silicon (Lang et al. 2017). These results

from atomistic simulations are consistent with those from the relevant continuum cohesive zone modeling (Yamakov et al. 2006) and experiments (Lang et al. 2017).

To calculate the fracture toughness through atomistic simulations for three-dimensional block samples, the plane strain condition must be imposed in the direction parallel to the crack front to ensure the validity of the calculated fracture toughness. Currently, there is still a gap in quantitative predictions of fracture toughness between atomistic simulations and experimental measurements, which may be attributed to distinct differences in the length scale and microstructure of the samples.

2.2   Simulation Setup for Fatigue

For fatigue simulations, the focus of the study is to assess the damage evolution of samples with or without pre-existing cracks under cyclic loading. For samples without an initial crack, cyclic loading is generally achieved by applying periodic alternating strain within a given timeframe. In each strain-controlled cycle, the maximum applied strain is generally smaller than the yielding strain of the simulated samples. This type of simulation is used to study damage initiation and development as well as the associated microstructural evolution of simulated samples with the accumulation of plastic strain.

For samples with a pre-existing crack, cyclic loading is accomplished by imposing a triangle-shaped alternating stress intensity factor. In each loading/unloading step, a specific K-field is imposed around the crack tip according to the K-controlled method mentioned above. The stress intensity factor range $\Delta K = K_{\max} - K_{\min}$, where $K_{\max}$ and $K_{\min}$ are the maximum and minimum stress intensity factors, respectively, should be smaller than the fracture toughness (i.e., K_{IC}) of the simulation samples. This type of simulation is used to investigate fatigue crack growth behaviors and the associated mechanisms. Furthermore, when the applied stress intensity factor range ΔK varies, the relationship between the fatigue crack growth rate da/dN and ΔK can be obtained and compared with the experimental results and relevant analytical models.

During fatigue simulations of samples with a pre-existing crack, tracking the fatigue crack growth (or determining the crack extension) is crucial for calculating the crack growth rate da/dN. Figure 2 shows two typical methods to determine the crack extension: one is by locating the maximum coordinate of the crack tip, referred to as the maximum coordinate method (Fig. 2a), while the other is by taking the average value over a characteristic point distributed on the crack front profile, referred to as the average coordinate method (Fig. 2b). In real experiments, the position of the crack tip is usually indirectly estimated from the measurement of the crack tip stress and/or displacement fields. In atomistic simulations, it is relatively easy to locate the exact position of the crack tip. However, such atomic-level resolution may induce a large error in the fatigue crack growth rate if one uses the maximum coordinate method to locate the position of the crack tip. For fatigue crack growth in ductile materials, the error can reach 10% (Zhou et al. 2015).

a

n th cycle $\quad x_{max}^{n}$

$\Delta x = x_{max}^{n+1} - x_{max}^{n}$

n+1 th cycle $\quad x_{max}^{n+1}$

b

x_1^n
x_2^n
x_3^n
x_4^n
x_5^n
x_6^n
x_7^n

n th cycle

$\Delta x = \dfrac{1}{m}\sum_{i=1}^{m}(x_{max}^{n+1} - x_i^n)$

x_1^{n+1}
x_2^{n+1}
x_3^{n+1}
x_4^{n+1}
x_5^{n+1}
x_6^{n+1}
x_7^{n+1}

n +1 th cycle
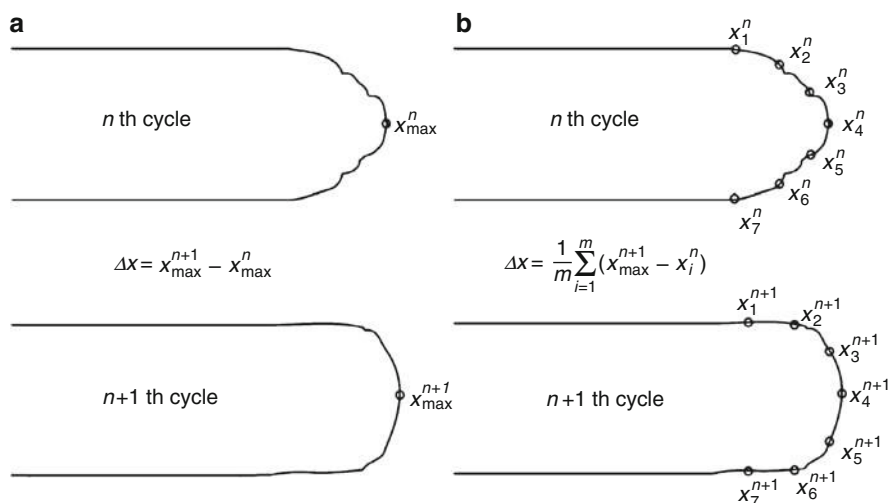
Fig. 2 Schematic illustrations for determining the fatigue crack growth during atomistic simulations (Zhou et al. 2015). (**a**) The maximum coordinate method and (**b**) the average coordinate method. (Reproduced with permission)

As indicated in Fig. 2b, the average coordinate method to some extent reflects the blunting effect of cracks. Therefore, the average coordinate method can be more suitable for estimating fatigue crack growth in ductile materials.

Since MD simulations have inherent limitations in terms of the time scale, fatigue simulations based on the MD technique only last for several hundred cycles, leading to low-cycle fatigue. Based on current computational power, it is possible to use the MS technique to study the high-cycle fatigue (i.e., cycle numbers above 10^4) of materials. However, the related computational cost is extremely high, and the length scale of the simulated sample is very limited.

3 MD Simulations for the Fracture of Nanotwinned and Amorphous Materials

In this section, we review recent atomistic simulations into the fracture behaviors of nanotwinned metals and ceramics and two representative amorphous materials (including metallic glasses and lithiated silicon). For nanotwinned metals and ceramics, the atomistic simulations reveal the complex interactions between a crack and various defects (such as dislocations and twin boundaries (TBs)). For the metallic glasses and lithiated silicon, atomistic simulations are used to investigate the microstructural evolution and deformation behaviors ahead of the crack tip.

3.1   Fracture of Nanotwinned Metals and Ceramics

Nanotwinned materials have a hierarchical microstructure, with a high density of nanoscale twins embedded in the micro- or submicron-sized grains. Therefore, nanotwinned materials exhibit higher strength and larger fracture toughness than their twin-free counterparts. So far, a large number of experimental and computational studies have investigated the plastic deformation of nanotwinned materials and revealed various strengthening mechanisms. There have also been an increasing number of studies on the fracture and fatigue behaviors of nanotwinned materials (Li et al. 2016). However, in contrast to plastic deformation mechanisms, the underlying fracture and fatigue mechanisms of nanotwinned materials are much less understood.

To reveal the underlying fracture mechanisms, we performed large-scale MD simulations of crack propagation in nanotwinned metals and ceramics. We simulated crack propagation in nanotwinned copper thin films with thicknesses varying from several nanometers to approximately 100 nm. Under mode I loading, an edge crack tip in a 100-nm-thick film emits numerous dislocations, which impact the TB ahead of the crack tip, as shown in Fig. 3a. These dislocations react with the TB, leading to the formation of a domain with a high density of dislocations on both sides of the TB, as shown in Fig. 3b. Such domain corresponds to the TB-affected dislocation zones observed experimentally and spans over approximately 5–7 nm, which is consistent with experimental observations (Kim et al. 2012). Such domain serves as a dislocation wall, which blocks further dislocation activities, resists crack propagation, and provides a reasonable explanation for the experimental observation of crack arrest and bridging by nanoscale twins (Kim et al. 2012). When the thickness of the nanotwinned film is only several nanometers, dislocations emitted from the crack tip easily slip across the TB, so that no dislocation wall forms (Kim et al. 2012). Figure 3c shows the typical crack propagation route along a zigzag path in the 1.7-nm-thick film (Zeng et al. 2015). Such zigzag cracking is induced by the persistent slip of dislocations and screw dislocation-mediated local thinning ahead of the crack (Zeng et al. 2015). The zigzag crack path corresponds to the cross-slip direction (i.e., <110> direction) of dislocations in the alternating matrix and twin domains (Zeng et al. 2015). These results (formation of dislocation wall and zigzag cracking) from MD simulations are consistent with previous experimental observations from in situ TEM (Kim et al. 2012; Zeng et al. 2015). Our MD simulations not only reveal the toughening mechanisms (i.e., the interaction between the crack and TBs) but also indicate the influence of film thickness on the fracture mechanisms of nanotwinned metallic thin films. These results have significant implications for understanding the mechanical failure of nanotwinned metallic thin films.

Due to the presence of hierarchical structures, conch shells possess remarkable strength and fracture toughness. Notably, as a basic structural building block, the third-order aragonite lamellae of a conch shell contain a large number of nanoscale {110} growth twins with an average TB spacing of 8 nm (Shin et al. 2016). However, the roles and functions of nanoscale growth twins in the mechanical behaviors of
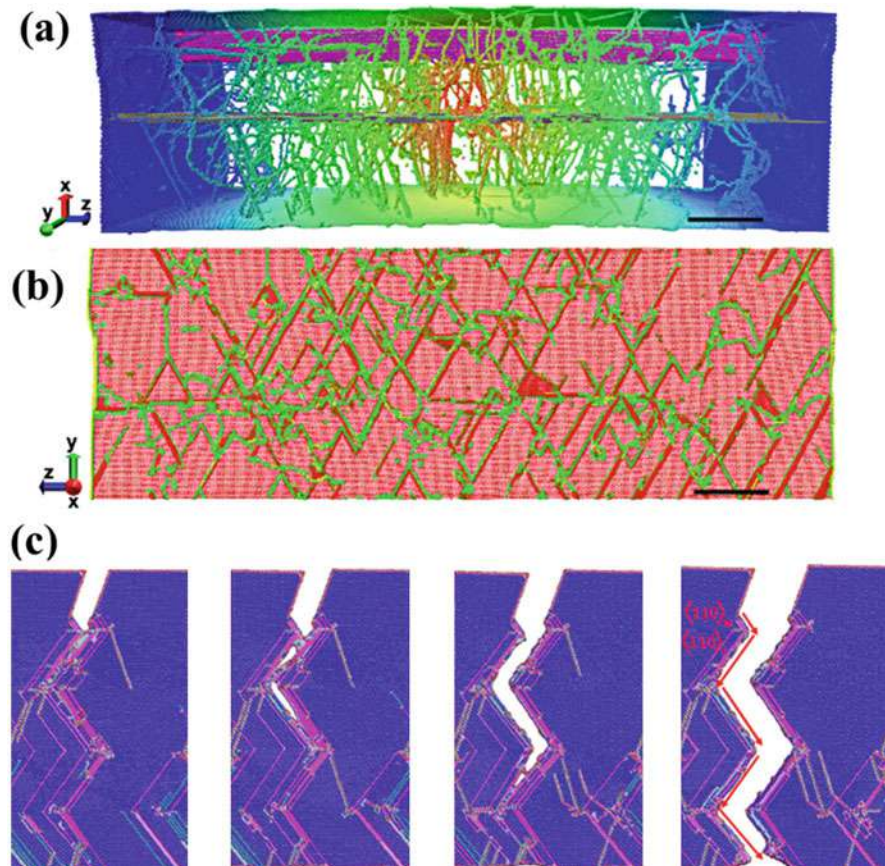
Fig. 3 Fracture mechanisms in nanotwinned copper thin films (Kim et al. 2012; Zeng et al. 2015). (**a**) Interaction between a TB and dislocations emitted from a crack tip ahead of the TB in the 100-nm-thick film (Kim et al. 2012). This leads to the TB transforming into a dislocation wall, which further blocks crack propagation. (**b**) Patterns of the dislocation structures along the TB (Kim et al. 2012). (**c**) Snapshots from the MD simulation showing crack growth along a zigzag path (Zeng et al. 2015). The edges of a growing crack are along the <110>$_M$ and < 110>$_T$ directions in the alternating matrix (M) and twin (T) domains, respectively. (Reproduced with permission)

the conch shell remain mysterious. We performed large-scale MD simulations of crack propagation in nanotwinned aragonite to explore the influence of nanoscale twins on the fracture properties and behaviors (Shin et al. 2016). Twin-free aragonite acts as the reference material for comparison. Figure 4a, b show snapshots of crack propagation in the twin-free aragonite single crystal with twin or matrix orientations, respectively. In the twin-free aragonite, the crack propagates fast and smoothly by cleavage, indicating brittle fracture. However, in nanotwinned aragonite, the crack tip is trapped by the TB and becomes blunt from sliding along the TB, as evidenced by Fig. 4c, d. It is noted that in twin-free aragonite, the stress/strain is

severely localized around the crack tip (Fig. 4a, b), while there exists an apparent stress/strain delocalization in nanotwinned aragonite (Fig. 4a, b). Such deformation delocalization is attributed to substantial irreversible phase transformation ahead of the crack tip. When the crack is trapped by the TB, a phase transformation occurs and spreads through the nanoscale twin domains ahead of the crack tip. Such phase transformation is driven by high crack tip stress (Shin et al. 2016). In twin-free aragonite, a phase transformation is also observed, but the transformed region is highly limited (Fig. 4e). In contrast, the transformed region in nanotwinned aragonite spans over the nanoscale twin lamella (Fig. 4f) and is much larger than that in the twin-free sample. The phase transformation involves a coordinated rotation of carbonate groups (Fig. 4g–i) and cannot spontaneously recover upon unloading. Figure 4g–i show the relevant details for the change in atomic configuration during the phase transformation. Our MD simulations reveal the underlying fracture mechanisms in nanotwinned aragonite, including the TB blocking crack propagation, phase transformation ahead of the crack tip, and crack tip blunting. These mechanisms facilitate energy dissipation and further delay catastrophic crack propagation, leading to an enhance fracture energy of the nanotwinned sample. This result is further evidenced by the calculation of the fracture energy based on our MD simulations. From the simulated stress-strain curves for crack propagation in the twin-free and nanotwinned samples, according to the second method mentioned in part 2, we calculated the fracture energy by integrating the stress-strain curve. It was found that the fracture energy of nanotwinned aragonite is one order of magnitude higher than that of twin-free aragonite. This prediction is in good agreement with experiment measurements (Shin et al. 2016).

3.2   Fracture of Metallic Glasses and Lithiated Silicon

Metallic glasses are a novel type of engineering alloy with metastable glassy states and are usually fabricated through rapid cooling of melting alloys. Due to their amorphous atomic structures, metallic glasses exhibit unique mechanical properties, such as high strength and high resistance to corrosion and oxidation. Notably, the fracture toughness of metallic glasses has a pronounced variation from 1 to 80 MPa·m$^{1/2}$ (Murali et al. 2011). Thus, metallic glasses are generally classified into brittle and ductile types. To reveal the fracture behaviors and associated mechanisms of these two types of metallic glasses, we performed MD simulations of crack propagation in FeP (brittle) and CuZr (ductile) samples (Murali et al. 2011). During the simulations for the FeP sample, some nanoscale voids nucleated and coalesced ahead of the crack tip, and subsequently, the crack merged with the nanoscale voids, leading to rapid growth of an atomically sharp crack, as shown in Fig. 5a. In contrast, in the CuZr sample, the crack tip becomes blunted through extensive shear banding ahead of the crack tip, as shown in Fig. 5b. These two distinct fracture mechanisms reflect the intrinsic features of brittle and ductile fracture, respectively. Further detailed analyses based on atomic stress ahead of the crack tip and critical hydrostatic stress for cavitation show that the observed nanoscale
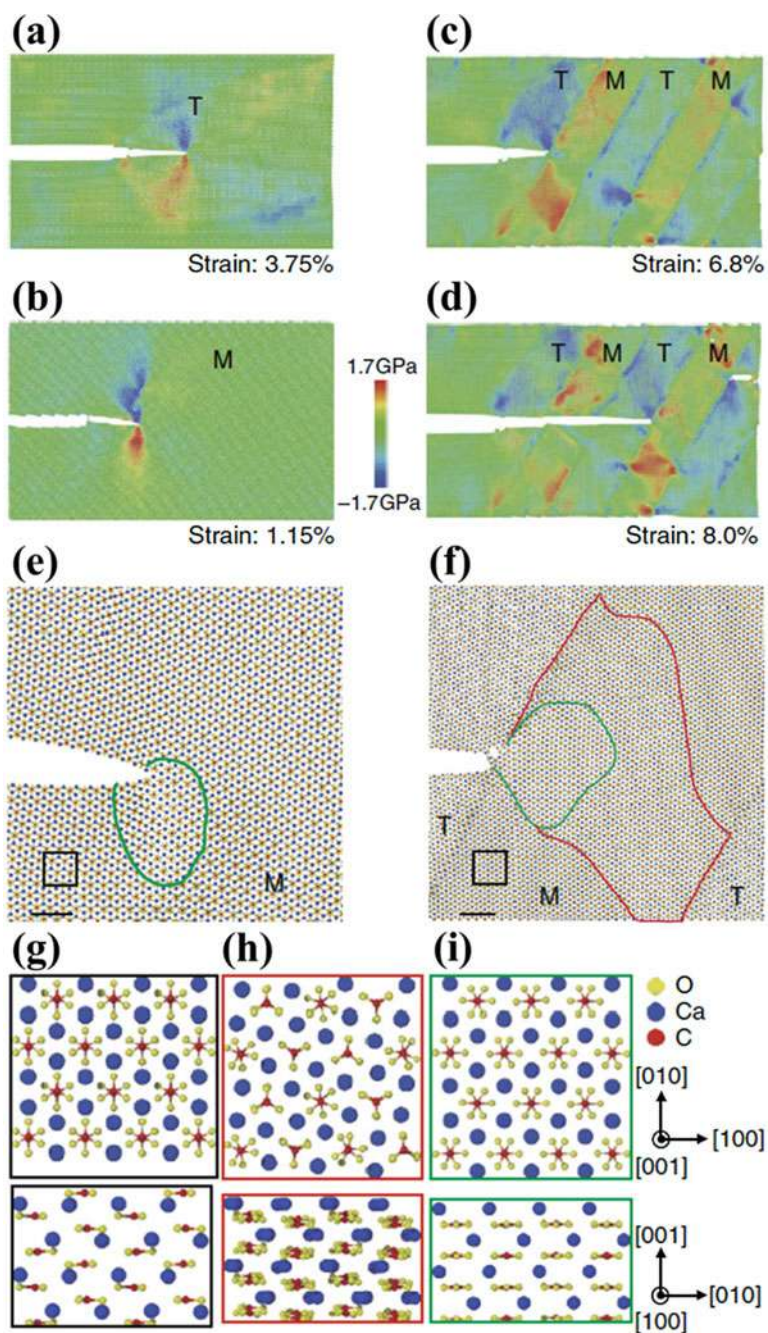
Fig. 4 (continued)

cavitation in the FeP sample can be attributed to pronounce spatial fluctuations in the local cavitation stress (Murali et al. 2011). Moreover, our MD simulations show that for samples with an initial nanoscale void under equi-biaxial plane strain loading, a single, isolated void grows in the CuZr sample during deformation, while multiple voids nucleate near the initial void and then merge with the initial void in the FeP sample. The critical stress for void growth is insensitive to the void size in the FeP sample but decreases with increasing void size in the CuZr sample.

A recent experimental study (Pan et al. 2015) reported an abnormal inverse notch effect in Zr-based metallic glasses, i.e., when a notch is introduced into the material, its tensile fracture strength is dramatically enhanced rather than decreased. To reveal the underlying mechanisms behind such inverse notch effect, we carried out large-scale MD simulations to investigate the fracture behaviors of notched CuZr samples. Similar to the experimental samples, we constructed two notched samples (ND1 and ND6), as shown in the inset in Fig. 5c. The ND6 sample has a narrower and sharper notch than the ND1 sample. Thus, based on the stress concentration, the ND6 sample is expected to have a lower fracture strength than the ND1 sample. The MD simulations show that the ND1 sample exhibits a higher fracture strength than the ND6 sample, which is consistent with the experimental results, as shown in Fig. 5c. Our MD simulations further reveal that this abnormal phenomenon is related to a transition in the fracture mechanism from shear banding to cavitation. In the notched ND1 sample, a large number of shear transformation zones (STZ) are activated and merge with one another, leading to the formation of an inclined shear band through the sample (Fig. 5d). However, in the notched ND6 sample, due to the sharper notch, there exists a triaxial stress state ahead of notch tip which is large enough to drive the nucleation and growth of nanoscale voids rather than shear banding (Fig. 5e). These fracture simulations provide a fundamental understanding of failure mechanisms in brittle and ductile metallic glasses, as well as mechanistic insights into the design of the macrostructure of metallic glasses to achieve enhanced capability.

Due to its high theoretical charge capacity, silicon has been identified as a promising cathode material for lithium-ion batteries. During the cyclic charging and discharging of batteries, crystalline silicon gradually becomes amorphous. In particular, during charging, numerous lithium ions are inserted into the silicon cathode, leading to a large volume change and the formation of amorphous lithiated

Fig. 4 Fracture mechanisms in nanotwinned aragonite (Shin et al. 2016). (**a**, **b**) Snapshots of crack propagation in twin-free aragonite single crystals with different orientations under strains of 3.75% and 1.15%, respectively. The labels "M" and "T" represent the matrix and twin orientations, respectively. (**c**, **d**) Snapshots of crack propagation in nanotwinned aragonite under strains of 6.8% and 8.0%, respectively. (**e**, **f**) Structural transformation ahead of the crack tip in single-crystalline and twinned aragonite. The perfect matrix regions are outlined by black lines, while the transformed regions are outlined by green and/or red lines. Scale bar, 2 nm. (**g–i**) Typical atomic configurations of aragonite near the crack tip before and after the structural transformation. (Reproduced with permission)
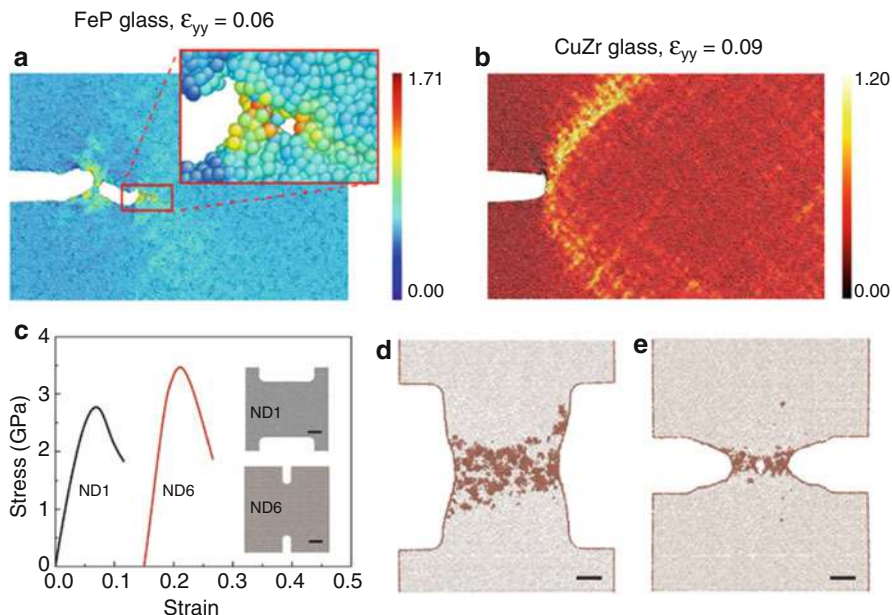
Fig. 5 Fracture mechanisms and an inverse notch effect in metallic glasses (Murali et al. 2011; Pan et al. 2015). (**a, b**) Snapshots of crack propagation in FeP and CuZr glasses (Murali et al. 2011). Atoms are colored according to their local von Mises strains. (**c**) Stress-strain curves of two notched CuZr samples (ND1 and ND6) with different notch dimensions (Pan et al. 2015). The peak stress of ND6 is 1.25 times than that of ND1. The insets show the initial configurations of two simulated samples. Scale bars, 10 nm. (**d, e**) Cross-sectional view of deformed samples at 15% strain (Pan et al. 2015). Scale bars, 5 nm. In (**d**), plastic deformation localizes into an inclined shear band, whereas in (**e**), a nanoscale void appears near the center of the sample as it fails by mode I fracture (Pan et al. 2015). (Reproduced with permission)

silicon (a-Li_xSi, where x represents the lithium concentration). Fracture of lithiated silicon is believed to be a major cause of capacity fading in the battery. Therefore, it is essential for the design and development of high-capacity batteries to study the fracture behaviors and mechanisms of lithiated silicon. Recent experimental studies show that as the lithium concentration increases, a brittle-to-ductile transition occurs in lithiated silicon (Wang et al. 2015). This implies that the fracture toughness of lithiated silicon is dependent on the lithium concentration. To explain the experimentally observed brittle-to-ductile transition, we performed a series of large-scale MD simulations to mimic crack propagation in a-Li_xSi with x varying from 0 to 4.4. During the simulations, a modified embedded-atom method (MEAM) potential was used to describe the interatomic interactions. Figure 6a–d show typical snapshots of crack propagation in a-$Li_{0.33}Si$ and a-$Li_{4.4}Si$ obtained from
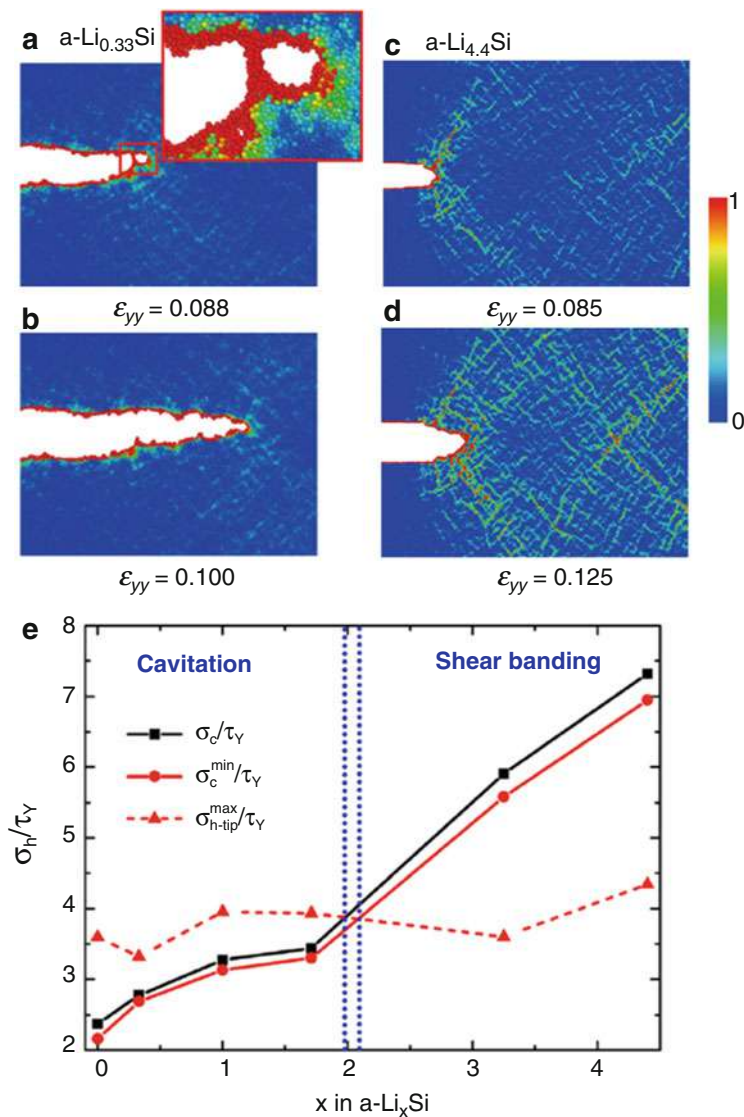
Fig. 6 Brittle-to-ductile fracture transition in lithiated silicon (Ding et al. 2015). (**a, b**) Snapshots of crack propagation in a-Li$_{0.33}$Si, showing a brittle fracture mode, i.e., rapid propagation of an atomically sharp crack through the nucleation and coalescence of nanovoids ahead of the crack tip. (**c, d**) Snapshots of crack propagation in a-Li$_{4.4}$Si, showing a ductile fracture mode, i.e., crack blunting through extensive shear banding. Atoms are colored according to their von Mises strains. (**e**) Ratios of hydrostatic (σ_h) and cavitation (σ_c) stresses to shear yield stress (τ_y) as functions of lithium concentration x in a-Li$_x$Si. (Reproduced with permission)

our MD simulations. In the samples with relatively low lithium concentrations (i.e., x ≤ 1.71), nanoscale voids nucleate ahead of the crack tip and then coalesce with the main crack, leading to rapid propagation of an atomically sharp crack, similar to the mechanism of brittle fracture in metallic glasses, as discussed in part 3.2. In contrast, in the samples with high lithium concentrations (i.e., x ≥ 3.25), shear bands form ahead of the crack tip. As the applied strain increases, the crack tip becomes blunted, which is similar to the fracture response of ductile metallic glasses discussed in part 3.2. Our simulation results are consistent with those from recent MD simulations for crack propagation in a-Li$_{0.5}$Si and a-Li$_{2.5}$Si (Wang et al. 2015), where the reactive force field (ReaxFF) was used. Most recently, the K-controlled loading method (introduced in part 2.1) and MEAM potential were used to simulate edge crack propagation in as-quenched a-Li$_x$Si with $x = 0.5, 1.0, 1.5$ and discharged samples with $x = 0.5, 1.5$ (Khosrownejad and Curtin 2017). During crack propagation, void nucleation occurs in these samples, which is consistent with our simulations. In these MD simulations, the fracture toughnesses were estimated by determining the critical K at crack initiation. The results indicated that the fracture toughness of lithiated silicon depends not only on the lithium concentration but also on the lithiation/delithiation history (Khosrownejad and Curtin 2017).

Our MD simulations provide important insight into the fracture mechanisms of lithiated silicon. According to the elastic-plastic fracture theory, we established the following criterion (Ding et al. 2015) for cavitation ahead of the crack tip:

$$\frac{\sigma_c}{\tau_y} \leq \alpha \left(\nu, \varepsilon_y, N \right) \tag{4}$$

where σ_c is the critical cavitation stress, τ_y is the shear yield stress, and α is a dimensionless ratio $\sigma_{h\text{-tip}}^{\max}/\tau_y$ dependent on Poisson's ratio ν, tensile yield strain ε_y, and hardening index N. To validate the above criterion, we determined the critical cavitation stresses σ_c and the shear yield stresses τ_y of lithiated silicon and computed the maximum hydrostatic stress $\sigma_{h\text{-tip}}^{\max}$ ahead of the crack tip and the minimum cavitation stress $\sigma_c^{\min}$ arising from atomic-scale fluctuations through MD simulations. Figure 6e shows the variations in σ_c/τ_y, $\sigma_c^{\min}/\tau_y$, and $\sigma_{h\text{-tip}}^{\max}/\tau_y$ with the lithium concentration x. As shown in Fig. 6e, $\sigma_{h\text{-tip}}^{\max}/\tau_y$ is nearly constant close to the predicted value of $1 + \pi$ from the Prandtl slip-line theory (Ding et al. 2015). For x ≤ ~2, σ_c/τ_y is less than $\sigma_{h\text{-tip}}^{\max}/\tau_y$, while for x ≥ ~2, σ_c/τ_y is greater than $\sigma_{h\text{-tip}}^{\max}/\tau_y$. According to the criterion in Eq. (4), it is predicted that nanoscale cavitation occurs in samples with x ≤ ~2, while shear banding dominates in samples with x ≥ ~2. Such prediction is in good agreement with our MD simulations, confirming a brittle-to-ductile transition in the fracture mechanism from intrinsic nanoscale cavitation to extensive shear banding ahead of the crack tip. Our MD simulations address a crucial effect of lithium concentration on the fracture properties of lithiated silicon, which may have profound implications for the design of damage-tolerant electrodes.

4   MD Simulations of the Fatigue of Nanotwinned and Amorphous Materials

As introduced in part 3, the fracture behaviors of nanotwinned and amorphous materials under monotonic loading were studied. However, their failure behaviors (i.e., fatigue) under cyclic loading and the associated microscopic mechanisms are much less known. In this part, we introduce some recent studies on low-cycle fatigue of nanotwinned metals and metallic glasses through large-scale atomistic simulations, with an emphasis on the underlying fatigue mechanisms as well as the relationship between the fatigue properties and microstructures.

4.1   Fatigue of Nanotwinned Metals

A few experimental studies have shown that nanotwinned metals have a noticeably improved resistance to fatigue crack growth compared with their twin-free counterparts (Li et al. 2016). However, the underlying mechanism responsible for the enhanced fatigue resistance is still poorly understood. We therefore conducted a series of large-scale MS and MD simulations to investigate the fatigue behaviors and mechanisms of nanotwinned copper with a mean grain size d from 10 nm to 20 nm and TB spacing λ from 0.83 nm to 5.01 nm. The ratios of d/λ in the simulated samples are between 5 and 18, close to those of experimental samples. During the simulations, the simulated samples are subject to a cyclic loading through the K-controlled method. The applied stress intensity factor amplitude ΔK varies from 2.40 MPa m$^{1/2}$ to 4.32 MPa m$^{1/2}$. All simulated samples are loaded for 30 cycles under a given ΔK. Figure 7a, b show snapshots of the crack tip configurations under maximum and minimum loading in a typical cycle, respectively. Many preexisting TBs ahead of the crack tip disappear during fatigue (Fig. 7a), indicating that detwinning is activated by TB migration through consecutive slip of partial dislocations along the TBs (Zhou et al. 2015). The crack tip is blunted due to dislocation emission and slip. Notably, double striations form at the crack tip (Fig. 7b), which are attributed to alternating crack tip blunting and sharpening during loading and unloading cycles (Zhou et al. 2015). Such double striation is commonly observed during the fatigue of ductile coarse-grained metals and alloys and is essentially distinct from the nucleation of nanoscale voids ahead of the crack tip during fatiguing of twin-free nanocrystalline metals (Zhou et al. 2015). For nanotwinned metals, detwinning and crack tip blunting are two dominant mechanisms during fatigue, which induces large energy dissipation and enhanced fatigue resistance. During the fatigue simulations, we obtained the crack growth curve by tracking the crack tip via the average coordinate method (introduced in part 2.2). We then characterized the fatigue crack growth rate da/dN by fitting the linear slope of the crack growth curve (Fig. 7c). Figure 7d shows a linear relationship between da/dN and ΔK on the log-log scale, indicating that the fatigue crack growth from atomistic simulations conforms to Paris' law (Zhou et al. 2015). It is noted
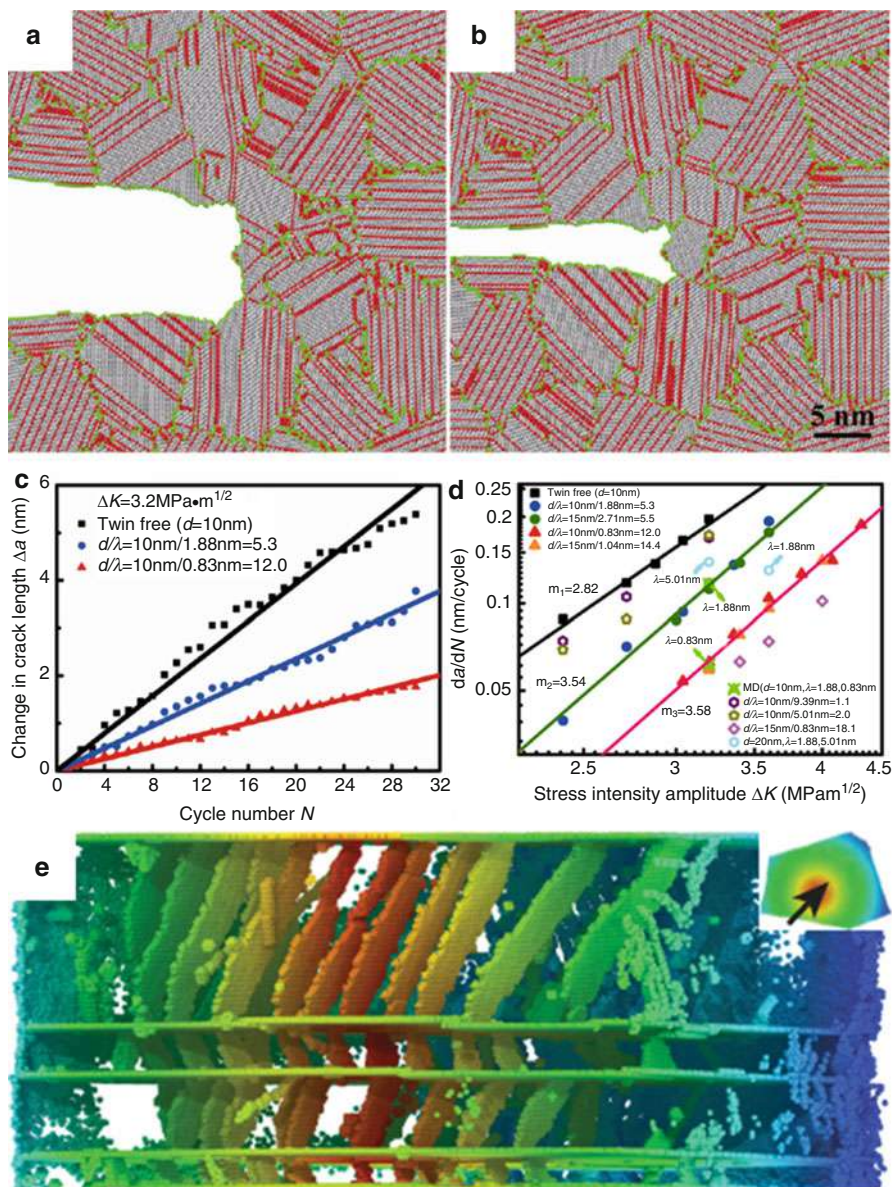
Fig. 7 (continued)
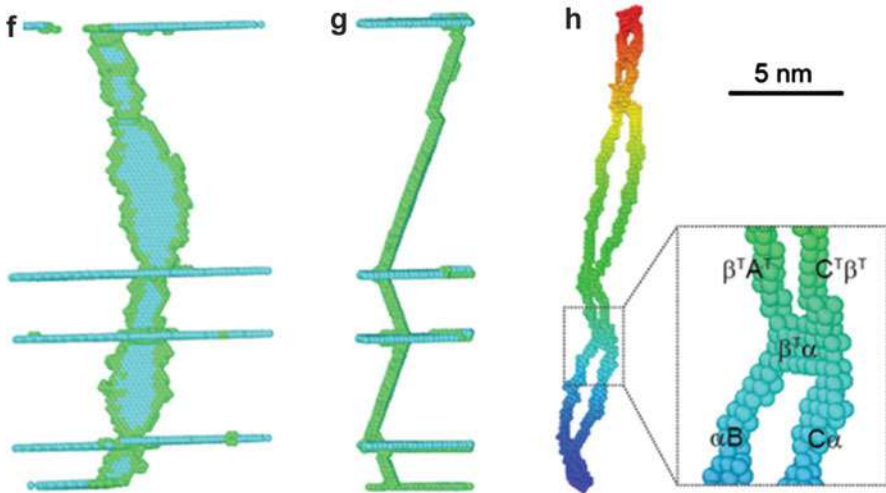
Fig. 7 Fatigue mechanisms and characterization of nanotwinned metals from atomistic simulations (Zhou et al. 2015; Pan et al. 2017). (**a, b**) Crack tip configurations of nanotwinned Cu under maximum and minimum loadings in a typical cycle (Zhou et al. 2015). (**c**) Crack growth as a function of cycles from our atomistic simulations (Zhou et al. 2015). (**d**) Crack growth rate versus stress intensity amplitude from our atomistic simulations (Zhou et al. 2015). (**e**) CND structures in a representative grain during fatigue (Pan et al. 2017). The arrow in the inset indicates viewing angle. The atoms are colored according to their spatial coordinates. (**f–h**) Atomic configurations of a typical CND (Pan et al. 2017). The Burgers vectors of the dislocation segments are based on the double Thompson tetrahedron representation. (Reproduced with permission)

that as the ratio of d/λ increases, the da/dN-ΔK curve moves down in the log-log diagram, suggesting that the fatigue resistance increases with increasing d/λ (Zhou et al. 2015). This result is consistent with experimental measurements. By fitting the simulation data in Fig. 7d, the exponent m in Paris' law is obtained as $\sim$3.5, which is close to that ($\sim$3.9) obtained from experimental studies (Zhou et al. 2015). Based on the damage accumulation model, we derived the following da/dN-ΔK relationship in the form of Paris' law (Zhou et al. 2015),

$$\frac{da}{dN} = \frac{5\pi}{96\alpha K_{IC}^2 \sigma_y^2}(\Delta K)^4 \tag{5}$$

where K_{IC} is the fracture toughness, σ_y is the yield strength, and α is a material constant (approximately 3000 for nanotwinned copper). The predictions from Eq. (5) agree well with the results from both atomistic simulations and experiments (Zhou et al. 2015).

During fatigue, the cyclic stress responses of many metals and alloys are unstable (hardening or softening) due to the accumulation of irreversible damage to the microstructures and exhibit an apparent dependence on the loading history (Pan et al. 2017). Remarkably, a recent fatigue experiment on nanotwinned copper

with columnar grains showed that when the cyclic loading axis is approximately parallel to the majority of the twin planes, the cyclic response becomes stable and history independent, irrespective of the imposed plastic strain amplitudes and cycle numbers (Pan et al. 2017). To explore the underlying mechanisms behind the history-independent and stable cyclic response, we carried out a series of MD simulations for the cyclic deformation of nanotwinned copper with $d = 40$ nm and $\lambda = 2$ nm, 3 nm, 5 nm, and 12 nm. The simulated samples contain columnar grains with <111> texture and have an initial dislocation density of 1.2×10^{17} m^{-2}, similar to the experimental samples. During the simulations, the simulated samples were subjected to uniaxial symmetric tension-compression cyclic loading through a strain-controlled method. All simulated samples were loaded for 20 cycles. Our MD simulations showed similar history-independent cyclic behaviors as the experimental observations. More importantly, a new type of dislocation structure, referred to as a correlated necklace dislocation (CND) (Fig. 7e), was discovered. Such CND is completely distinct from the typical dislocations observed in conventional metals and alloys during fatigue, as well as various dislocations found in nanotwinned metals under monotonic loading (Pan et al. 2017). A CND consists of a series of threading dislocation segments on inclined slip planes between neighboring TBs and stair-rod dislocations at the intersection between the inclined slip planes and TBs (Pan et al. 2017), as illustrated in Fig. 7f–h. The CND structures were verified by TEM observations (Pan et al. 2017). They are formed by linking neighboring threading dislocation tails on the TBs, and collectively they move back and forth along the TBs under cyclic deformation (Fig. 7e), which not only carries relatively homogeneous plastic strain in the neighboring twin/matrix lamellae but also preserves the coherency and stability of the TBs. Throughout the simulations, the collective motion of CNDs is identified as a dominant deformation mechanism responsible for the observed stable and history-independent cyclic behaviors. Our MD simulations also indicated that as long as the highly oriented twins are titled within approximately 15° of the loading axis, CNDs are formed during cyclic deformation, thereby leading to a history-independent cyclic response. Our MD simulations thus revealed a novel fatigue mechanism at the atomic scale, which can be used to account for the observed history-independent behavior and may benefit the fabrication and design of structured materials with enhanced fatigue resistance.

4.2 Fatigue of Metallic Glasses

Fatigue experiments on various metallic glasses have revealed some characteristic deformation behaviors under cyclic loading (Sun and Wang 2015), e.g., shear banding significantly affects the fatigue of metallic glasses. The experimental results also indicated that the fatigue properties of metallic glasses are sensitive to defects (such as micro-cracks, pores, and surface flaws) as well as environmental and geometrical effects. Currently, there is little or no atomic-level understanding of the formation of shear bands in metallic glasses during fatigue. Thus, we conducted a series of MD simulations on the cyclic deformation of a CuZr slab. During the

Fig. 8 Fatigue mechanisms in metallic glasses (Sha et al. 2015). (**a**) Aggregation of STZs marked by the red line in the 13th cycle, indicating the formation of a shear band. (**b**) Penetration of the shear band through the entire sample occurred at the 20th cycle. (**c**) Times for shear band initiation propagation as functions of applied strain amplitude. (Reproduced with permission)
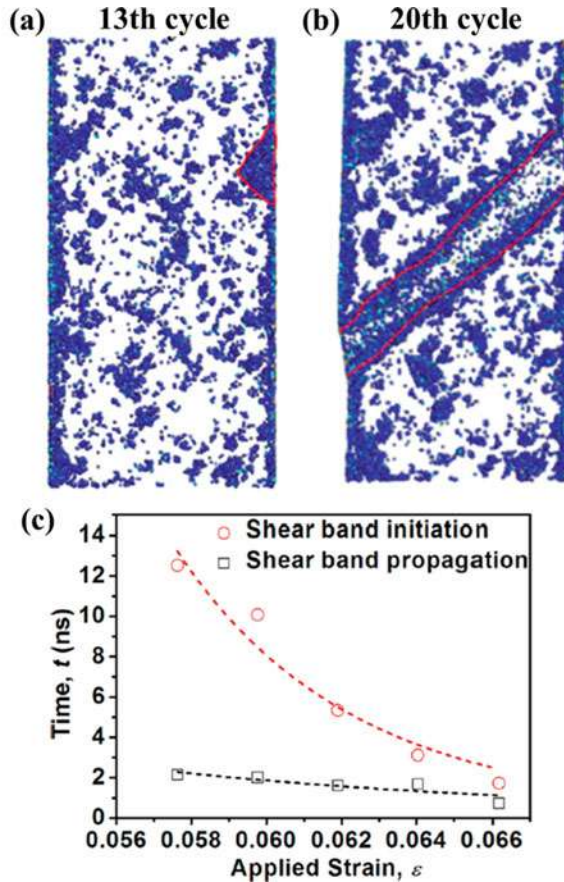
simulations, the simulated samples were subjected to tension-compression loading through a strain-controlled method. The maximum applied strain varied from 1.71% to 6.40%, which ranges over the elastic and plastic regimes of the simulated sample under monotonic loading (Sha et al. 2015). The simulated samples were loaded for 175 cycles. Figure 8a, b show two typical snapshots of the simulated sample during fatigue with an applied strain of 6.40%. In the initial cycles, a large number of STZs formed at the surface or inside of the sample. With increasing cycle number, some STZs near the surface aggregated in an embryonic region of the shear band (Fig. 8a). When the region size exceeded a critical value, the shear band initiated and then quickly propagated across the entire sample (Fig. 8b). Figure 8c shows the time instants for shear band initiation and propagation as functions of applied strain. It is seen that as the applied strain increases, the time needed for shear band initiation slowly decreases, whereas the time for shear band propagation quickly decreases (Sha et al. 2015). Notably, shear band initiation dominates most of the fatigue life of metallic glasses. This phenomenon is somewhat similar to

that in crystalline materials, where the fatigue life is also dominated by the damage initiation stage (Sha et al. 2015). Moreover, during the MD simulations, we changed the cycling frequency to investigate its influence in the fatigue life of metallic glasses. It was found that increasing the cycling frequency prolonged the fatigue life of metallic glasses (Sha et al. 2015). This phenomenon is attributed to the fact that the accumulation rate of STZs decreases with increasing cycling frequency (Sha et al. 2015). Our MD simulations not only provide some atomistic insights into the fatigue process of metallic glasses but may also suggest a strategy to enhance the fatigue life of metallic glasses by suppressing shear band initiation.

5   Summary

Atomistic simulations were applied to explore the fracture and fatigue behaviors and associated mechanisms of nanotwinned and amorphous materials. The results from atomistic simulations shed light on the fundamental failure mechanisms of these materials and have been used to explain or interpret experimental results. The phenomena observed in the atomistic simulations are consistent with those from related experiments, especially those involving in situ TEM observations. Furthermore, atomistic simulations can supply some quantitative predictions, such as the fracture toughness, the fatigue crack growth rate, and the da/dN-ΔK diagram, which are found to be comparable to experimental measurements.

Despite the rapid development of atomistic simulations, there still exists an apparent gap between atomistic simulations and experiments due to the inherent limitations in the length scales and time scales of atomistic simulations, as well as the limited accuracy of interatomic potentials. For example, the strain rate of MD simulations is several orders of magnitude higher than the loading rate used in experiments, and atomistic simulations have been limited to modeling low-cycle fatigue. Such gap can be reduced with the emergence of increasingly accurate interatomic potentials and multiscale modeling techniques, as well as the continuous improvement of supercomputing capacity. More powerful supercomputers allow atomistic simulations to mimic the failure of materials at slower and more realistic strain rates and at larger length scales. In recent years, some multiscale methods have been proposed by coupling quantum calculations, atomistic simulations, and continuum-level modeling. These multiscale methods bridge different length scales and allow us to comprehensively capture complex fracture processes of materials, from bond breakage, to crack growth, to the interplay of crack tip elastic and plastic fields. The interatomic potentials used for atomistic simulations are becoming increasingly more accurate, reliable, and sophisticated. In particular, reactive fore fields based on the bond order have been developed to describe bond formation and breakage (Li and Gao 2015). Such force fields should be more accurate than empirical interatomic potentials. Atomistic simulations with reactive force fields can treat systems with length scales on the order of tens of nanometers. Most recently, with the advances of artificial intelligence, machine learning methods (including artificial neural networks, Gaussian processes, and support vector machines) have

been used to develop interatomic potentials (Behler 2016). Such machine learning potentials can provide very accurate energies close to those of quantum methods but are much more efficient than quantum methods (Behler 2016). Currently, atomistic simulations with machine learning potentials can address systems with length scales of approximately 10 nm. The rapid development of supercomputers, multiscale methods, and interatomic potentials will undoubtedly make atomistic simulations more realistic, predictive, and useful for research into the processing, properties, microstructure, deformation, and failure of next generations of advanced materials systems.

References

Behler MJ (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145:170901

Ding B, Li X, Zhang X, Wu H, Xu Z, Gao H (2015) Brittle versus ductile fracture mechanism transition in amorphous lithiated silicon: from intrinsic nanoscale cavitation to shear banding. Nano Energy 18:89–96

Freund LB (1990) Dynamic fracture mechanics. Cambridge University Press, Cambridge

Jung G, Qin Z, Buehler MJ (2015) Molecular mechanics of polycrystalline graphene with enhanced fracture toughness. Extreme Mech Lett 2:52–59

Khosrownejad SM, Curtin WA (2017) Crack growth and fracture toughness of amorphous Li-Si anodes: mechanisms and role of charging/discharging studied by atomistic simulations. J Mech Phys Solids 107:542–559

Kim S, Li X, Gao H, Kumar S (2012) In situ observations of crack arrest and bridging by nanoscale twins in copper thin films. Acta Mater 60:2959–2972

Kumar KS, Van Swygenhoven H, Suresh S (2003) Mechanical behavior of nanocrystalline metals and alloys. Acta Mater 51:5743–5774

Lang J, Ding B, Zhang S, Su H, Ge B, Qi L, Gao H, Li X, Li Q, Wu H (2017) Scalable synthesis of 2D Si nanosheets. Adv Mater 29:1701777

Li X, Gao H (2015) Atomistic modelling of deformation and failure mechanisms in nanostructured materials. Natl Sci Rev 2:133–136

Li X, Dao M, Eberl C, Hodge A, Gao H (2016) Fracture, fatigue, and creep of nanotwinned metals. MRS Bull 41:298–304

McDowell MT, Xia S, Zhu T (2016) The mechanics of large-volume-change transformations in high-capacity battery materials. Extreme Mech Lett 9:480–494

Murali P, Guo T, Zhang Y, Narasimhan R, Li Y, Gao H (2011) Atomic scale fluctuations govern brittle fracture and cavitation behavior in metallic glasses. Phys Rev Lett 107:215501

Ovid'ko IA (2007) Review on the fracture processes in nanocrystalline materials. J Mater Sci 42:1694–1708

Pan J, Zhou H, Wang Z, Li Y, Gao H (2015) Origin of anomalous inverse notch effect in bulk metallic glasses. J Mech Phys Solids 84:85–94

Pan Q, Zhou H, Lu Q, Gao H, Lu L (2017) History-independent cyclic response of nanotwinned metals. Nature 551:214–217

Pineau A, Benzerga AA, Pardoen T (2016a) Failure of metals I: brittle and ductile fracture. Acta Mater 107:424–483

Pineau A, McDowell DL, Busso EP, Antolovich SD (2016b) Failure of metals II: fatigue. Acta Mater 107:484–507

Pineau A, Benzerga AA, Pardoen T (2016c) Failure of metals III: fracture and fatigue of nanostructured metallic materials. Acta Mater 107:508–544

Sha Z, Qu S, Liu Z, Wang T, Gao H (2015) Cyclic deformation in metallic glasses. Nano Lett 15:7010–7015

Shin Y, Yin S, Li X, Lee S, Moon S, Jeong J, Kwon M, Yoo S, Kim Y, Zhang T, Gao H, Oh S (2016) Nanotwin-governed toughening mechanism in hierarchically structured biological materials. Nat Commun 7:10772

Sun BA, Wang WH (2015) The fracture of bulk metallic glasses. Prog Mater Sci 74:211–307

Wang X, Fan F, Wang J, Wang H, Tao S, Yang A, Liu Y, Chew HB, Mao SX, Zhu T, Xia S (2015) High damage tolerance of electrochemically lithiated silicon. Nat Commun 6:8417

Yamakov V, Saether E, Phillips DR, Glaessgen EH (2006) Molecular-dynamics simulation-based cohesive zone representation of intergranular fracture processes in aluminum. J Mech Phys Solids 54:1899–1928

Yip S (ed) (2005) Handbook of materials modeling. Springer, Dordrecht

Zeng Z, Li X, Lu L, Zhu T (2015) Fracture in a thin film of nanotwinned copper. Acta Mater 98:313–317

Zhu T, Li J (2010) Ultra-strength materials. Prog Mater Sci 55:710–757

Zhou X, Li X, Chen C (2015) Atomistic mechanisms of fatigue in nanotwinned metals. Acta Mater 99:77–86

Modelling of Defects and Failure in 2D Materials: Graphene and Beyond

77

V. Sorkin, Q. X. Pei, and Y. W. Zhang

Contents

Abstract

Defects in two-dimensional materials can have profound effect on their mechanical properties. In thermodynamic equilibrium, the second law of thermodynamics predicts the inevitable existence of defects. In non-equilibrium state, defects

V. Sorkin · Q. X. Pei · Y. W. Zhang (✉)
Institute of High Performance Computing, A*STAR, Singapore, Singapore
e-mail: zhangyw@ihpc.a-star.edu.sg

can be hosted unintentionally or intentionally into two-dimensional materials to modify their mechanical properties. In general, there are two types of intrinsic defects in two-dimensional materials: point defects (vacancies, self-interstitials, dislocations, and topological defects) and line defects (grain boundaries). Due to the difference in lattice structures and bonding energies, the configurations of these defects may take difference forms. Here we outline the main computational techniques (ab initio, density functional-based tight binding and molecular dynamics) most commonly used to study defects in two-dimensional materials. The basic modelling approaches to simulate the defect forms and their effects on the mechanical properties of 2D materials are reviewed. Although the family of two-dimensional materials is expanding fast, we mostly focus on four members of this family: graphene, boronitrene, MoS_2, and phosphorene, with emphasis on typical types of defects: vacancies, Stone-Wales defects, dislocations, and grain boundaries.

1   Introduction

The family of two-dimensional (2D) materials is expanding rapidly. Due to the difference in chemical compositions, the lattice structure and bonding energy for each family member are different. As a result, they exhibit various distinctive mechanical, physical, and chemical properties. Here, we briefly discuss four prime representative 2D materials: graphene, boronitrene or hexagonal boron nitride (h-BN), molybdenum disulfide (MoS_2), and phosphorene. Graphene, a single layer of carbon atoms in a hexagonal honeycomb lattice, is one of the most prominent examples of 2D materials. It has attracted considerable attention due to its unique electronic structure, outstanding electrical conductivity, and exceptional strength. Boronitrene is a 2D insulator with noteworthy mechanical properties. Structurally, boronitrene, also called "white graphene," is similar to graphene with comparable bond lengths and layer thicknesses. However, these two materials display different mechanical behavior, predominantly due to the different type of bonding (covalent bonding between C-C atoms vs. mixed ionic-covalent bonding between B-N atoms) (Pakdel et al. 2012). Semiconducting MoS_2 has a triple-layer structure with the central hexagonal plane of molybdenum sandwiched between two hexagonal sulfur planes. Its crystal structure is effectively a 2D honeycomb lattice with the two sub-lattices populated by Mo- and S-atoms, where each Mo-atom is sixfold coordinated, while each S-atom is threefold coordinated. Its unusual mechanical properties arise from the significant ionic character of Mo-S bonds (Sorkin et al. 2014). Semiconducting phosphorene, a monolayer of black phosphorus, has provoked a significant interest due to its unusual puckered structure, direct bandgap, and strong anisotropy of the mechanical properties (Sorkin et al. 2016).

Defects in 2D materials can have profound effects on their mechanical properties. In thermodynamic equilibrium, the second law of thermodynamics predicts the inevitable existence of defects. In non-equilibrium state, defects can be intentionally

introduced into 2D materials during fabrication to realize certain functionalities that are absent in the defect-free ones. In general, there are two classes of intrinsic defects in 2D materials: point defects (vacancies, self-interstitials, topological defects, and dislocations) and line defects (grain boundaries). Due to the difference in lattice structures and bonding energies, the configurations of these defects may take different forms in different 2D materials.

In this book chapter, we will first briefly discuss several major computational techniques (density functional theory, tight binding, and molecular dynamics) that have been widely used in the study of defects in 2D materials, and also basic modelling methods that are often used to examine the defect geometries, the formation energies, and their effects on the mechanical properties. We next discuss in detail the specific structural features and properties of vacancies, topological defects, dislocations, and grain boundaries in the four selected 2D materials, mainly focusing on their effects on the ultimate strength, fracture strain, and failure mechanism. Finally, we will discuss the opportunities and challenges of defect engineering in 2D materials.

2   Computational Methods and Techniques

2.1   Computational Methods

Three main computational techniques: density functional theory, tight binding, and molecular dynamics, are commonly used to investigate defects in 2D materials. Among them, density functional theory (DFT) is one of the most efficient and accurate first-principles computational techniques in 2D materials research. Since DFT is a quantum mechanical method, it does not require any specific prior knowledge of system under the examination. Given that DFT calculations are generally rather computationally expensive, only relatively small samples (reaching a few hundred atoms) can be studied by this technique. In dealing with 2D materials, plane-wave DFT calculations are usually carried out with the exchange correlation potential, approximated by generalized gradient approximation (GGA) or a variety of other functionals. In these calculations, all the structures are treated by using the periodic boundary conditions. Commonly used DFT calculation software packages are VASP or CASTEP.

The behavior of larger systems with thousands to billions of atoms can be studied using classical molecular dynamics (MD) simulations. In MD simulations, the atoms in 2D materials obey Newton's law of motion driven by interatomic forces derived from semiempirical interatomic potentials. These interatomic potentials are constructed and fitted to model the particular behavior of the system obtained by first-principles calculations and/or experimental measurements. For example, in order to simulate graphene, adaptive intermolecular reactive empirical bond order (AIREBO) or Tersoff potential is often employed. It is now known that AIREBO potential is capable to capture accurately the interaction between carbon

atoms, as well as bond breaking and making, while Tersoff potential is the most suitable to examine the mechanical properties. In addition, Tersoff potential was parametrized to describe the interaction between boron and nitrogen atoms in boronitrene (Sorkin et al. 2014). Stillinger-Weber and reactive empirical bond order (REBO) potentials are used to model interactions between molybdenum and sulfur atoms in MoS2 (Li et al. 2016a) and between phosphorus atoms in phosphorene (Sha et al. 2016).

Tight-binding (TB) technique is another method that has often been applied to study the structure, large deformation, and failure of 2D materials with defects. The TB technique has a very special place among the computational methods. On one hand, as discussed above, the DFT method is very precise but computationally demanding. As a result, the DFT calculations are not feasible for large or even moderate scale systems. On the other hand, applying molecular dynamics (MD) method for large systems is rather restricted as reliable and commonly accepted interatomic potentials for many 2D materials are not yet available. Consequently, empirical TB technique, placed in between DFT and MD in terms of computational cost and accuracy, is an appropriate method in dealing with the size problem. Thus, TB methods can be considered as a bridge between classical atomistic methods and quantum mechanical methods. TB simulations are usually performed by using DFTB+ package.

In the following, we briefly discuss computational models and methods that are commonly used to determine the mechanical properties of 2D materials, that is, defect mobility, elastic modulus, and failure behavior.

2.2   Determine Defect Mobility

One of the important characteristics of a defect is its mobility. In general, the mobility of a vacancy is determined by its diffusion (activation) barrier, which the defect must overcome while moving in the sample. Diffusion barriers and migration paths of vacancies are commonly explored by using the nudged elastic band (NEB) method in combination with DFT-based or TB-based techniques (Zobelli et al. 2007). The NEB method is constructed using the optimization of the total energy of a string of system images linking the initial and the final states, and no a priori assumption for the migration path is required. Thus, the NEB method is suitable for examining large regions of configuration space to obtain the set of minimum-energy transition paths. However, due to the slow convergence of the NEB method, the use of standard DFT techniques can be computationally expensive. A good substitute for DFT is density-functional tight-binding (DFTB) or MD method, which is able to reduce computational cost by orders of magnitude while still remaining a reasonable accuracy. It has already been demonstrated that combining DFTB and NEB gives results that are comparable to other higher-end techniques (Zobelli et al. 2007).

2.3   Determine Elastic Moduli

The presence of defects may significantly change the mechanical properties of 2D materials. A special interest is the modification of elastic moduli by defects. The most studied elastic quantity is the change of Young's modulus due to vacancies. By hosting a sample with vacancies (or other types of defects) at different concentrations, one is able to calculate the stress vs. strain curve by subjecting the constructed sample to small uniaxial tensile and compressive strain ($\varepsilon \leq 5\%$). The slope of the calculated stress-strain curve or the second derivative of the strain energy with respect to strain at $\varepsilon = 0$ can be used to find the Young's modulus (Jing et al. 2012). Other elastic moduli can be obtained in a similar way.

2.4   Determine Failure Behavior

In general, the failure behavior of 2D materials is dictated by defects. Due to the presence of defects, 2D materials can fail at much lower failure stress than their defect-free counterparts. In order to investigate the failure mechanism, 2D samples with defects are subjected to uniaxial tensile straining (Jing et al. 2012; Sorkin and Zhang 2016; Sorkin and Zhang 2015). Under uniaxial tensile straining, the strain is defined as the change in length, L, relative to the initial (reference) length, L_0: $\varepsilon = L/L_0 - 1$. In simulations, the sample can be stretched quasi-statically or at a specific strain rate. When a uniaxial tensile strain is applied quasi-statically, the 2D sample is strained along the specific direction by changing the length of the computational box and rescaling the atom positions in that direction. Subsequently, the sample geometry is optimized, and stress is relaxed in the direction perpendicular to the applied strain by minimizing the total energy of the sample. Conjugate gradient or quasi-Newton minimization techniques are commonly used for energy minimization. The defected sample is quasi-statically strained until it fails via bond fracture. Thus one can obtain the fracture strength and fracture strain and examine the fracture mechanism in details. Fracture strength and fracture strain are determined at the point on the stress-strain curve where the maximum stress is reached. The tensile stress, σ, is defined as $\sigma = F/A$, where F is the tensile force and $A = wt$ is the cross-sectional area (assuming thickness of $t = 0.335$ nm for graphene, $t = 0.61$ nm for MoS_2, $t = 0.34$ nm for boronitrene, and $t = 0.53$ nm for phosphorene).

The quasi-static calculations are performed at zero temperature using first-principle methods (Sorkin and Zhang 2015, 2016; Dettori et al. 2012; Rodrigo et al. 2016; Sensoy et al. 2017; Gan and Zhao 2014; Akinwande et al. 2016; Zhang et al. 2012). A 2D sample with defects can be stretched at finite temperature and specific strain rate using MD simulations (Akinwande et al. 2016; Zhang et al. 2012; Xiong et al. 2015; Sha et al. 2014, 2015; Guo et al. 2016). Initially, the sample is thermally equilibrated at the selected temperature using an NPT ensemble (constant number of atoms, pressure, and temperature). The internal residual stresses in the sample are relaxed to zero since during the thermal equilibration, the computational box is

allowed to change its shape and volume. Subsequently, a uniaxial tensile strain is applied by increasing the computational box length in a specific direction at a given strain rate ($\dot{\varepsilon} = 10^8$–10^{10} s^{-1}). At each time step of MD simulations, all the atomic positions are rescaled following the changes in the length of the computational box. Alternatively, a uniaxial tensile strain can be applied in a step-by-step manner. In each strain increment, the thermal equilibration of the strained sample is performed for a short period of time. During atomistic simulations, the 2D sample is kept at a specific temperature using a thermostat (Nose-Hoover thermostat or simple velocity rescaling method). The strain rate is chosen to be as slow as possible to reduce the potential strain rate-dependent effects. The typical engineering strain rate in atomistic simulations is within 10^8–10^{10} s^{-1}; this strain rate is by several orders of magnitude higher than those applied in experimental tensile tests. The prohibitive computational cost limits the ability to simulate the tensile deformation at a much lower strain rate. Like in the case of straining at zero temperature, the defected sample at finite temperatures is stretched until the ultimate failure strain is reached, from which the fracture strength and fracture mechanism can be determined.

In the following, we will discuss structures of four types of defects (vacancy, Stone-Wales defect, dislocation, and grain boundary) and their effects on the mechanical properties of four different 2D materials, that is, graphene, boronitrene, MoS$_2$, and phosphorene.

3 Vacancies

Vacancies are the most commonly studied point defects in 2D materials. Vacancies can be formed during the fabrication processes or can be intentionally introduced by ion (electron) irradiation. In simulations, vacancies are introduced by removing atoms from 2D materials. A single vacancy (monovacancy) requires a removal of a single atom, a double vacancy (divacancy) requires a removal of a couple adjacent atoms, and so on. Subsequently, geometry optimization is followed, through which the atomic structure around the vacancy is relaxed. The sample geometry is optimized by minimizing its total energy (usually at zero temperature using conjugate gradient method). Atomic structure around the vacancy can be obtained using a variety of modelling techniques, ranging from first-principles to atomistic simulations. The technique is selected as required by the preferred accuracy.

3.1 Structure and Formation Energy

The most important characteristic of a vacancy is its formation energy (E_f), which determines its concentration at thermodynamic equilibrium. The formation energy can be obtained by calculating the difference between the total energy of a sample with vacancy and the energy of the corresponding defect-free sample with the same number of atoms. The vacancy formation energy can be calculated by DFT with a high accuracy and precision.
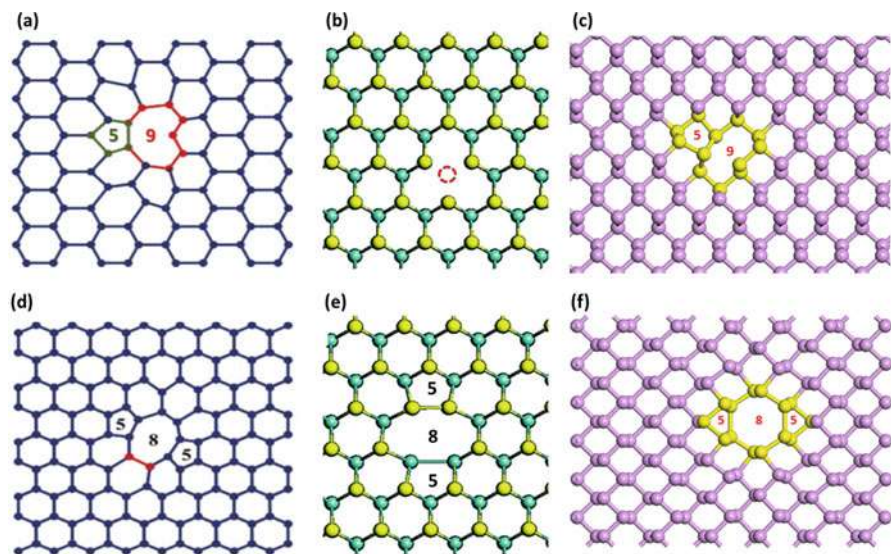
Fig. 1 Vacancies in 2D materials: (**a**) Monovacancy (5|9) in graphene (Adapted with permission from Rajasekaran et al. 2015: copyright 2015 Taylor & Francis). (**b**) Single Mo-vacancy in MoS_2 (Adapted from Yu et al. 2015a with permission of the PCCP Owner Societies). (**c**) Monovacancy (5|9) in phosphorene. (**d**) Double (5|8|5) vacancy in graphene (Reprinted with permission from Rajasekaran et al. 2015: Copyright 2015 Taylor & Francis). (**e**) Double (5|8|5) vacancy in MoS_2 (Adapted from Yu et al. 2015a with permission of the PCCP Owner Societies). (**f**) Double (5|8|5) vacancy in phosphorene

The atomic configurations around single and double vacancies in different 2D materials are illustrated in Fig. 1. The removal of a single atom in graphene produces a monovacancy with three under-coordinated neighboring atoms around it (each atom possesses a single dangling bond at the vacancy edge). The rearrangement of the under-coordinated atoms around the monovacancy leads to the formation of covalent bonds between them (El-Barbary et al. 2003). As a result, a single (5|9) configuration appears in graphene by forming five- and ninefold rings and leaving only one dangling bond (see Fig. 1a). In this process, small out-of-plane displacements of the under-coordinated atoms also occur. In contrast, rearrangement of under-coordinated atoms around a double vacancy in graphene eliminates all the dangling bonds and can lead to a pentagon-octagon-pentagon (5|8|5) structure (see Fig. 1d). A fully reconstructed divacancy actually has several structure variants, including (5|8|5), (555|777), and (5555|6|7777) "flower" defects (see Fig. 2d). Although these divacancies have a similar level of formation energy (of which the (555|777) divacancy has the lowest E_f), they are separated by high energy barriers ($\Delta E_b \approx 5$ eV), which effectively impede any transformation between them (via thermally activated bond rotation) at room temperature.

A double vacancy is more stable than two single vacancies (see Table 1). The formation energy of a double vacancy is lower since it does not contain

Table 1 Vacancy formation energy of single and double vacancies in 2D materials: graphene (Banhart et al. 2011; Faccio and Mombrú 2012), phosphorene (Liu et al. 2014), MoS_2 (Ataca and Ciraci 2011; Yu et al. 2015b; Hong et al. 2015), and boronitrene (Yu et al. 2015a; Zobelli et al. 2006; Okada 2009; Alem et al. 2009)

2D material	Vacancy formation energy	
	Single vacancy (eV)	Double vacancy (eV)
Graphene	7.57	6.37
Phosphorene	1.56	1.36
MoS_2 (Mo-vacancy)	5.59–6.2	
MoS_2 (S-vacancy)	1.22–3.03	1.51–4.14
h-BN (B-vacancy)	11.22	
h-BN (N-vacancy)	8.91	7.7

under-coordinated carbon atoms with dangling bonds. Thus, single vacancies have a tendency to migrate via thermally activated diffusion and coalesce to form divacancies by releasing energy (Do et al. 2005). In the same way, triple vacancies have a lower energy than separated monovacancy and divacancy and so forth. It is expected that the coalescence of large numbers of vacancies at elevated temperatures may lead to the formation of amorphous graphene structure (Skowron et al. 2015).

Typical configurations of atoms around monovacancies and divacancies in MoS_2 are shown in Fig. 1b, e, respectively. Since MoS_2 is composed of two atomic species, there are two types of monovacancies: S-vacancy and Mo-vacancy (see Fig. 1b). In contrast to graphene, covalent bond formation is not observed between the under-coordinated atoms around a single Mo-vacancy. The S-atoms surrounding Mo-vacancy show an outward relaxation, while the Mo- and S-atoms surrounding S-vacancy exhibit slightly inward relaxation (Feng et al. 2014a). First-principles calculations show that the formation energy for Mo-vacancy is significantly larger than that of S-vacancy (Ataca and Ciraci 2011) (see Table 1). Experimentally, S-vacancies (with a concentration of $c_v \approx 1\%$) are predominantly observed during CVD growth and mechanical exfoliation (Yu et al. 2015b). The type of observed S-vacancies depends on the nature of MoS_2 fabrication: single S-vacancies are found in exfoliated samples while double S-vacancies in CVD samples (Yu et al. 2015b). DFT calculations and experimental observations (Zhou et al. 2013) showed that the formation energy of double S-vacancies is roughly twice of that single vacancy (see Table 1). In sharp contrast to graphene (where the formation of double vacancies by combination of monovacancies is energetically favorable), in MoS_2, single S-vacancies do not have a strong propensity to combine to form double vacancies. The single S-vacancy is the most common point defect repeatedly observed in experiments because it has the lowest formation energy. This explains why S-vacancies are mainly found in exfoliated samples but fails to explain why double S-vacancies are mostly observed in CVD samples. This peculiarity of the double S-vacancy formation was explained by Yu et al. (2015b). Using DFT calculations, they revealed an intricate pathway for preferable formation of S-divacancies in MoS_2, which involves dislocations and grain boundaries abundant

in CVD samples. Formation energy for vacancies in MoS_2 (see Table 1) also strongly depends on the ambient conditions, i.e., S-rich or S-poor environment during CVD growth. The formation energies of Mo-monovacancies can be lower than those of S-monovacancies under S-rich condition. However, single Mo-vacancies are never observed experimentally. Instead Mo-vacancy complexes, which include a group of missing atoms (a Mo-atom and its three S-neighbors), are mainly found since once the Mo-vacancy is generated, the neighboring S-atoms become unstable and predisposed to be lost even under S-rich condition (Zhou et al. 2013).

Vacancies in boronitrene can be formed under electron beam exposure, with B-atoms being much more likely to be ejected than N-atoms. In boronitrene, like in MoS_2, formation of covalent homoelemental B-B (N-N) bonds between the under-coordinated neighboring B-atoms of N-monovacancy (N-atoms of B-monovacancy) is rarely observed (Zobelli et al. 2006) since it is energetically unfavorable. Both single B- and N-vacancies conserve the threefold symmetry around the missing atom without local reconstruction of dangling bonds. Only small local in- and out-of-plane atom relaxations around single vacancies are found (Zobelli et al. 2006). However, deletion of a couple of the adjacent B-N atoms leads to the formation of a double (5|8|5) B-N vacancy and results in the creation of homoelemental bonds (Okada 2009). Akin to MoS_2, the formation energy of boronitrene also depends on the ambient condition, i.e., chemical potentials of B- and N-atoms.

Atomic configurations around single (5|9) and double (5|8|5) vacancies in phosphorene are shown in Fig. 1c, f, correspondingly. Atoms around the single vacancy undergo reconstruction by forming a stable (5|9) pentagon-nonagon defect (see Fig. 1c), similar to that of graphene. Since the formation energy of mono-vacancy ($E_f = 1.56$ eV) is larger than that of divacancy ($E_f = 1.36$ eV), it is energetically favorable for two monovacancies to combine into a divacancy. First-principles calculations predict at least 14 variants of divacancies in phosphorene (Liu et al. 2014) (see Fig. 2b, c, e, f), while in graphene, there are only 3 types of fully reconstructed divacancies (see Fig. 2a, d). This staggering variety of divacancies is due to the anisotropic puckered lattice structure of phosphorene. Among these different variants, the (5|8|5) divacancy (see Fig. 2c) has the lowest formation energy. The vacancy formation energy in phosphorene is the smallest one (see Table 1), indicating that equilibrium vacancy concentration in phosphorene should be abnormally large in comparison with other 2D materials at the same conditions (Liu et al. 2014). The low value of E_f in phosphorene is related to its intrinsically softer P-P bonds (compared to the much stronger C-C bonds in graphene and other 2D materials) and also to the curvature effects of its puckered structure.

3.2 Vacancy Diffusion

Defects in 2D materials are not necessarily stationary, and their migration is able to affect the properties of the hosting materials. For example, vacancies can be
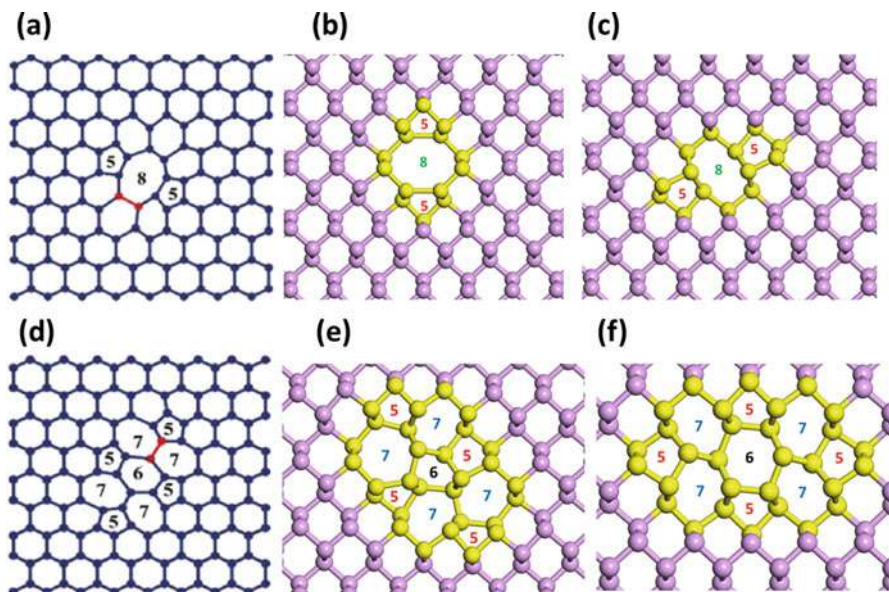
Fig. 2 Atomistic structures of reconstructed divacancies in graphene: (**a**) (5|8|5) divacancy, (**d**) (5555|6|7777) "flower" divacancy. The transformation between the neighboring divacancies can be realized through 90° rotation of the highlighted bond (Adapted from Rajasekaran et al. 2015: copyright 2015 Taylor & Francis). Atomistic structures of various reconstructed divacancies in phosphorene: (**b**, **c**) two variants of (5|8|5) divacancy and (**e**, **f**) two variants of (5555|6|7777) "flower" divacancy

Table 2 Diffusion energy barrier of single and double vacancies in 2D materials: graphene (El-Barbary et al. 2003; Rajasekaran et al. 2015), phosphorene (Cai et al. 2016), MoS_2 (Komsa et al. 2013), and h-BN (Zobelli et al. 2007)

2D material	Diffusion energy barrier	
	Single vacancy (eV)	Double vacancy (eV)
Graphene	1.39	7.1
Phosphorene	0.3	0.44–1.83
MoS_2 (S-vacancy)	2.27	8.5
h-BN (B-vacancy)	2.6	
h-BN (N-vacancy)	5.8	4.5–6.0

fairly mobile at high temperatures. The mobility of vacancies is determined by activation (diffusion, migration) energy barriers (see Table 2), which they have to overcome in the process of diffusion. In graphene, migration of a single vacancy occurs through the displacement of one of the carbon atoms located near the vacancy into a vacant site. This diffusion can be directly observed even at room temperature (El-Barbary et al. 2003). In contrast to single vacancy, the activation barrier for divacancy is almost five times larger (Banhart et al. 2011; Faccio and Mombrú 2012) (see Table 2). This high activation barrier indicates that divacancies in graphene are

immobile at room temperature, and their thermally activated migration takes place only at extremely elevated temperatures of T $\geq$ 3000 K (El-Barbary et al. 2003).

In boronitrene, only thermally activated B-monovacancies can diffuse at high temperatures (T $\geq$ 840 K) according to DFT calculations (Zobelli et al. 2007). Thermal migration of a single B-vacancy is a complex process: it occurs as one of its *second* neighbor atoms jumps into a vacant site (a boron atom reaches the vacant site through a few intermediate configurations). This B-vacancy migration preserves the heteroelemental B-N bonding between atoms. The activation energy barrier for N-monovacancies is so high (see Table 2) that their migration is energetically unfavorable at any temperature below the melting point (Zobelli et al. 2007). Migration of single B-vacancies often leads to the coalescence with immobile N-vacancies by forming a more stable B-N divacancies; their diffusion can be thermally activated at even higher temperatures (T $\geq$ 1400 K). The single and double vacancies agglomerate further by forming extended rotational and line defects (Zobelli et al. 2007).

Single S-vacancies in MoS_2 are found to be only mobile under electron beam radiation since the diffusion energy barrier (see Table 2) is prohibitively high for thermally activated diffusion at room temperature. The diffusion energy barrier for Mo-vacancies is even higher (Komsa et al. 2013). The rate of single S-vacancy diffusion caused by the impacts of energetic electrons in irradiation experiments is fairly small (Komsa et al. 2013; Feng et al. 2014b). However, DFT calculations suggest that the diffusion barrier can be markedly lowered (from $E_b = 2.3$ eV down to $E_b = 0.8$ eV) if the vacancy concentration is increased (Komsa et al. 2013). Using DFT calculations, Sensoy et al. (2017) found that the lattice reorganization around single S-vacancies is extremely localized, and there is no elastic interaction between two neighboring defects. In view of this, there is no thermodynamic driving force for the formation of vacancy clusters in MoS_2 (Sensoy et al. 2017).

Vacancy diffusion in phosphorene was studied by Cai et al. (2016) using first-principles calculations. They found that the single vacancy in phosphorene has the lowest activation energy barrier among all discovered 2D materials ($E_b \approx 0.3$ eV along the ZZ direction). The monovacancy diffusion in phosphorene is very anisotropic, with predominant migration in the ZZ direction. At room temperature, the motion of single vacancy in phosphorene is by $\sim$16 orders of magnitude faster than in graphene. The itinerant features of phosphorene vacancies are rooted in the unique puckering structure of phosphorene (Cai et al. 2016). It was predicted that experimental observation of vacancies in phosphorene would only be possible at low temperatures (T $\leq$ 70 K), where their thermally activated diffusion is amply suppressed. Furthermore, in contrast to other 2D materials exhibiting an extremely low mobility of double vacancies even at high temperatures, the migration divacancies in phosphorene can be detected around T=600 K. This is because the migration of double vacancies in phosphorene can occur in a completely novel way – through repeated association and dissociation of the fast-moving single vacancies (Cai et al. 2016). The presence of the itinerant vacancies in phosphorene can explain its fast structure evolution in vacuum and its rapid degradation in air.

3.3   Elastic Modulus

The effect of vacancies on the mechanical properties of 2D materials depends on the type, concentration, distribution of vacancies, and the loading direction (for anisotropic 2D materials). First we discuss the effect of vacancies on the elastic moduli of 2D materials, in particular, Young's modulus, which is a measure of stiffness of a material, one of the most important properties in engineering design.

It is expected that the presence of vacancies in graphene should lead to the deterioration of its mechanical properties. Both first-principles (Rodrigo et al. 2016) and atomistic simulations (Mortazavi and Ahzi 2013) indicate significant softening of graphene containing vacancies. However, one of the most surprising experimental discoveries was the observation of an increase in the Young's modulus at sufficiently low vacancy concentration (López-Polín et al. 2014). The Young's modulus increases up to almost twice of its original defect-free value at $c_v \approx 0.2\%$. DFT calculations revealed that at very low concentration, monovacancies stretch graphene and induce the strain fields that quench the out-of-plane vibrations, making the defective graphene effectively stiffer than pristine one (Rodrigo et al. 2016). At a higher concentration of vacancies, the Young's modulus decreases almost linearly with increasing vacancy concentration (Mortazavi and Ahzi 2013): at vacancy concentration $c_v = 2\%$, it decreases by about 20% (see Fig. 3a).

In contrast to graphene, introduction of sulfur monovacancies and divacancies in MoS_2 has only a relatively minor effect on its Young's modulus. DFT calculations indicate a relatively minor reduction of the elastic modulus about 4% for both types of vacancies at $c_v \approx 5\%$ (Gan and Zhao 2014). However, removal of a pair of neighboring Mo- and S-atoms (Mo-S divacancy) has a more dramatic effect – the Young's modulus decreases by about 36% along the AC direction and about 63% along the ZZ direction (Gan and Zhao 2014). The effect of V_{MoS3} point defects (a vacancy complex created by removal of a Mo-atom and its three nearby S-atoms) on the elastic properties of MoS_2 was examined by Li et al. (2016a) using MD simulations. They found that an increase of V_{MoS3} vacancy concentration (from $c_{MoS3} = 1\%$ to $c_{MoS3} = 10\%$) leads to a significant decrease in the Young's modulus ($\Delta Y/Y_0 \approx -25\%$). In addition, they identified an anomalous stiffening effect (the Young's modulus increases at elevated temperature) in defected MoS_2 containing V_{MoS3} vacancies at $c_{MoS3} = 0.1\%$ (Li et al. 2016a).

The impact of vacancies on the elastic moduli of boronitrene is rather insignificant. MD simulations showed that the Young's modulus is nearly unchanged ($\Delta Y/Y_0 \approx -2\%$) at relatively low concentration of double B-N vacancies ($c_v \approx 1\%$) (Le and Nguyen 2014).

Phosphorene is known for its distinctly anisotropic mechanical properties. In particular, its Young's modulus along the ZZ direction is considerably larger than that along the AC direction. Correspondingly, the effect of vacancies on the Young's modulus is also remarkably anisotropic. The magnitude of this effect is still the subject of considerable discussion. Using MD simulations, Sha et al. (Sha et al.
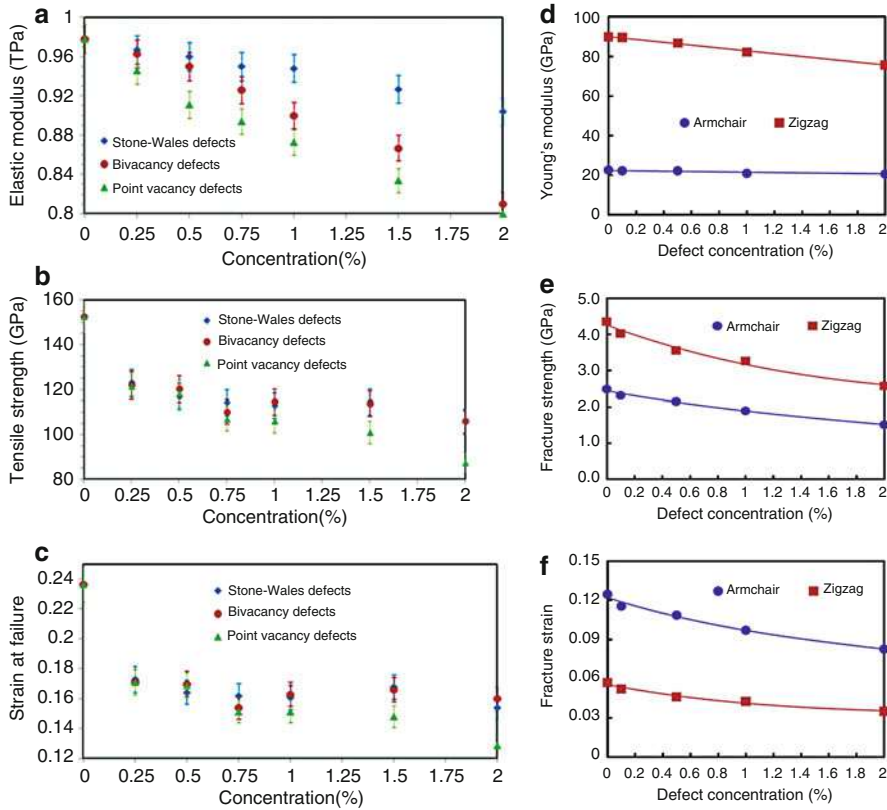
Fig. 3 Effect of single (green triangles) and double (red circle) vacancies on the Young's modulus (**a**), fracture strength (**b**), and fracture strain (**c**) of graphene (Adapted from Mortazavi and Ahzi 2013 with permission of Elsevier Ltd). Effect of single vacancies on the Young's modulus (**d**), fracture strength (**e**), and fracture strain (**f**) of phosphorene calculated along the AC and the ZZ directions (Adapted from Sha et al. 2016 with permission of the IOP publishing)

2016) found that vacancies significantly reduce the Young's modulus along the AC direction ($\Delta Y/Y_0 \approx -10\%$ at $c_v = 2\%$) but have negligible effect along the ZZ direction ($\Delta Y/Y_0 \leq -1\%$) as shown in Fig. 3d. On the other hand, first-principles calculations (Hao and Chen 2016) demonstrate that the Young's modulus of phosphorene is affected by vacancies along the AC direction, but along the ZZ direction, it can be reduced up to $\Delta Y/Y_0 \approx -12\%$ by monovacancies and up to $\Delta Y/Y_0 \approx -7\%$ by divacancies at $c_v = 2\%$. While by using atomistic simulations, Chen et al. (2017) indicated that vacancies can reduce the Young's modulus almost to the same extent in both directions: $\Delta Y/Y_0 \approx -9\%$ in the AC direction and $\Delta Y/Y_0 \approx -7\%$ in the ZZ (at $c_v \approx 1.2\%$). They also found that the softening effect of double vacancies is markedly larger than that of single vacancies.

3.4   Fracture Strength, Strain, and Failure Mechanism

Understanding of the effect of vacancies on the fracture strain and strength and failure behavior of 2D materials is crucial for their applications. The fracture strength is much more sensitive to defects than the Young's modulus. This can be understood in terms of the "weakest link" model, which conjectures that one weak link may be sufficient to cause failure.

Fracture strength of graphene decreases significantly with increasing vacancy concentration. Atomistic simulations demonstrate a sharp decline ($\Delta\sigma/\sigma_0 \approx -20\%$ -40%) in the tensile fracture strength of graphene even at sufficiently low vacancy concentration $c_v \approx 0.3\%$ at room temperature (Mortazavi and Ahzi 2013; Wang et al. 2012) (see Fig. 3b, c). The hosted vacancies act as stress concentration centers, causing a significant out-of-plane wrapping. Moreover, they coalesce into sizeable holes in graphene; hence fracture strength is greatly reduced.

The dependence of the fracture strain on the vacancy concentration in graphene is more complicated: it decreases at low concentrations but increases again at higher ones. At very high vacancy concentration, the graphene deformation becomes plastic-like and exhibits superductility (Xu et al. 2013). Near the failure strain, multiple monoatomic carbon chains are formed along the AC direction. As the tensile strain increases, the elongation of these monoatomic carbon chains holds the graphene connected until the last connecting carbon chain ruptures (Mortazavi and Ahzi 2013).

MD simulations indicate that monovacancies reduce both the fracture strength and failure strain of graphene to a larger extent than divacancies (Mortazavi and Ahzi 2013). Mortazavi et al. (Mortazavi and Ahzi 2013) found that the failure strain decreases by $\Delta\varepsilon/\varepsilon_0 \approx -47\%$ due to monovacancies and only by $\Delta\varepsilon/\varepsilon_0 \approx -31\%$ due to divacancies at $c_v = 2\%$ (see Fig. 3b). The fracture strength is reduced by $\Delta\sigma/\sigma_0 \approx -43\%$ due to single vacancies and only by $\Delta\sigma/\sigma_0 \approx -30\%$ due to double vacancies as shown in Fig. 3c. This indicates that double vacancies are more stable since they undergo a series of configuration changes, allowing them to sustain a larger tensile strain (Mortazavi and Ahzi 2013).

The failure of defected graphene initiates when a few bonds in the vicinity of a vacancy rupture (see Fig. 4a). Consequently, a nano-size void is formed around the ruptured bonds. The void grows under applied tensile strain, while the C-C bonds along its edge rupture. This process ultimately leads to the failure of graphene (Degefe and Parashar 2016).

Likewise, the fracture strength and the ultimate strain of boronitrene are abruptly reduced by vacancies. N-monovacancies lessen the fracture strain along the ZZ direction by $\Delta\varepsilon/\varepsilon_0 \approx -36\%$ (at $c_v = 6\%$) while B-monovacancies only by $\Delta\varepsilon/\varepsilon_0 \approx -20\%$. Equally, the fracture strength of boronitrene is significantly reduced by N-monovacancies ($\Delta\sigma/\sigma_0 \approx -16\%$) in comparison with B-monovacancies ($\Delta\sigma/\sigma_0 \approx -8\%$) (Xiong et al. 2015). For boronitrene, the detrimental effect of N-monovacancies and BN divacancies on the fracture strain and the ultimate strain is practically the same (Le and Nguyen 2014).

The effect of single and double S-vacancies on the fracture strength of MoS_2 depends on the direction of applied strain. It is relatively weak along the ZZ
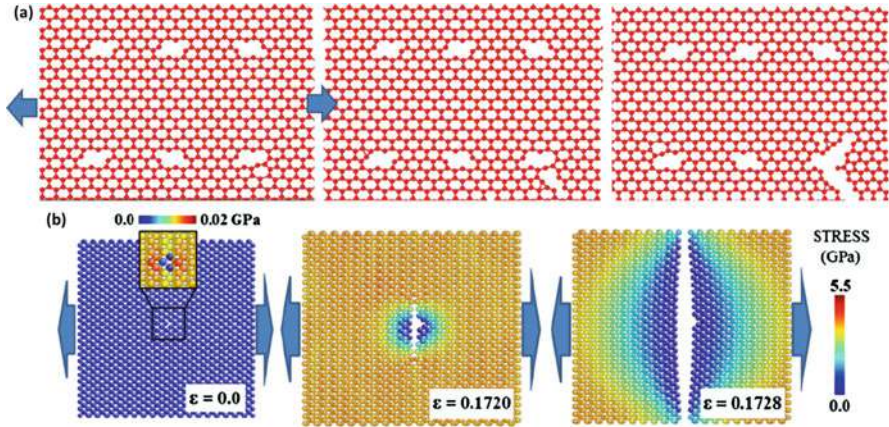
Fig. 4 (**a**) Tensile fracture of graphene containing double vacancies under applied strain along the AC direction (Adapted from Degefe and Parashar 2016 with permission of the IOP publishing). (**b**) Crack nucleation and fracture of phosphorene containing a single vacancy under applied tensile strain along the AC direction (Adapted from Sha et al. 2016 with permission of the IOP publishing)

direction: the fracture strength is reduced by less than $\Delta\sigma/\sigma_0 \approx -4\%$ by mono- and divacancies at vacancy concentration $c_v = 5\%$. However, along the AC direction, it is reduced by $\Delta\sigma/\sigma_0 \approx -48\%$ by monovacancies and $\Delta\sigma/\sigma_0 \approx -21\%$ by divacancies (Gan and Zhao 2014). Besides, the fracture strain of defected MoS_2 decreases by $\Delta\varepsilon/\varepsilon_0 \approx -42\%$ due to introduction of S-monovacancies along the ZZ direction and by $\Delta\varepsilon/\varepsilon_0 \approx -32\%$ due to S-divacancies. Along the AC direction, the fracture strain is reduced by $\Delta\varepsilon/\varepsilon_0 \approx -48\%$ due to the monovacancies and by $\Delta\varepsilon/\varepsilon_0 \approx -32\%$ due to the divacancies (Gao and Zhao 2016). However, the hosting of Mo-S divacancies has the most dramatic effect. The fracture strength is reduced by $\Delta\sigma/\sigma_0 \approx -80\%$ along the ZZ direction and by $\Delta\sigma/\sigma_0 \approx -17\%$ along the AC direction. The corresponding reduction in the fracture strain is $\Delta\varepsilon/\varepsilon_0 \approx -56\%$ along the ZZ direction and $\Delta\varepsilon/\varepsilon_0 \approx -34\%$ along the AC (Gan and Zhao 2014). The effect of the Mo-S divacancies is so strikingly detrimental that the stability of defected MoS_2 for fracture along the ZZ direction was questioned (Gan and Zhao 2014).

In phosphorene, the fracture strength and fracture strain are greatly reduced by vacancies (see Fig. 4). MD simulations indicate that due to its flexible puckered structure, the atomic stress distribution is nearly uniform in defected phosphorene, except for the region around the vacancy (the stress concentration is shown in Fig. 4b). The stress concentration causes the bond failure in this region, resulting in the crack nucleation. The emerged crack propagates along a groove of the puckered structure, forming atomically flat fracture surface (Sha et al. 2016).

Both the fracture strength and ultimate strain of phosphorene along the AC and the ZZ directions decrease with an increase in the vacancy concentration (Sha et al. 2016) (see Fig. 3e, f, respectively). At a vacancy concentration of $c_v = 2\%$, the fracture strength decreases by $\Delta\sigma/\sigma_0 \approx -39\%$ along the AC direction

Table 3 Effect of monovacancies on the mechanical properties of 2D materials. Reduction of the Young's modulus, fracture strength, and fracture strain of defected 2D materials: graphene (Mortazavi and Ahzi 2013), phosphorene (Sha et al. 2016; Hao and Chen 2016), boronitrene (Le and Nguyen 2014), and MoS_2 (Gao and Zhao 2016) relative to the defect-free ones. Only effects of N-vacancies in boronitrene and S-vacancies in MoS_2 are indicated

2D material	Chirality	Vacancy concentration	$\Delta Y/Y_0$	$\Delta\sigma/\sigma_0$	$\Delta\varepsilon/\varepsilon_0$
Graphene	AC/ZZ	2%	−20%	−43%	−36%
Phosphorene	AC	2%	−10%	−39%	−34%
	ZZ	2%	−6%	−43%	−47%
Boronitrene	AC	6%	−3%	−12%	−23%
	ZZ	6%		−15%	−35%
MoS_2	AC	5%	−4%	−25%	−48%
	ZZ	5%		−4%	−42%

and by $\Delta\sigma/\sigma_0 \approx -41\%$ along the ZZ direction. The ultimate strain reduces by $\Delta\varepsilon/\varepsilon_0 \approx -34\%$ and by $\Delta\varepsilon/\varepsilon_0 \approx -39\%$ along the AC and ZZ directions, correspondingly. First-principles calculations (Hao and Chen 2016) also indicate that the fracture strength ($\Delta\sigma/\sigma_0 \approx -45\%$) and fracture strain ($\Delta\varepsilon/\varepsilon_0 \approx -56\%$) of defected phosphorene ($c_v = 2\%$) are particularly reduced along the ZZ direction. Therefore, it is much easier to fracture the defected phosphorene along the ZZ direction than the defect-free phosphorene.

The type of vacancy also has influence on the failure mechanism of defected phosphorene: introduction of single vacancy has stronger effect than that of double vacancy. Due to the anisotropy of its mechanical properties, the effect of single and double vacancies differs along the AC and the ZZ directions. Effective reduction in the fracture strength due to double vacancies is much larger than that due to the single ones along the AC direction, while the reduction is similar along the ZZ direction (Sha et al. 2016).

The effects of single vacancies (at a similar vacancy concentration) on the mechanical properties of 2D materials are summarized in Table 3. It can be seen that the mechanical properties of graphene and phosphorene are most sensitive to vacancy defects among the listed 2D materials. In particular, the mechanical properties of graphene are modified to the largest extent by vacancies.

4 Stone-Wales Defects

One of the exceptional properties of 2D materials with honeycomb lattice structure is their capacity to undergo local restructuring via bond rotations and form non-hexagonal rings. Hence a novel class of topological defects can be introduced in 2D materials without adding or removing atoms. The most extensively studied topological defects are the Stone-Wales defects in graphene (Ma et al. 2009; Stone and Wales 1986) (see Fig. 5a). A Stone-Wales (SW) defect in graphene can be
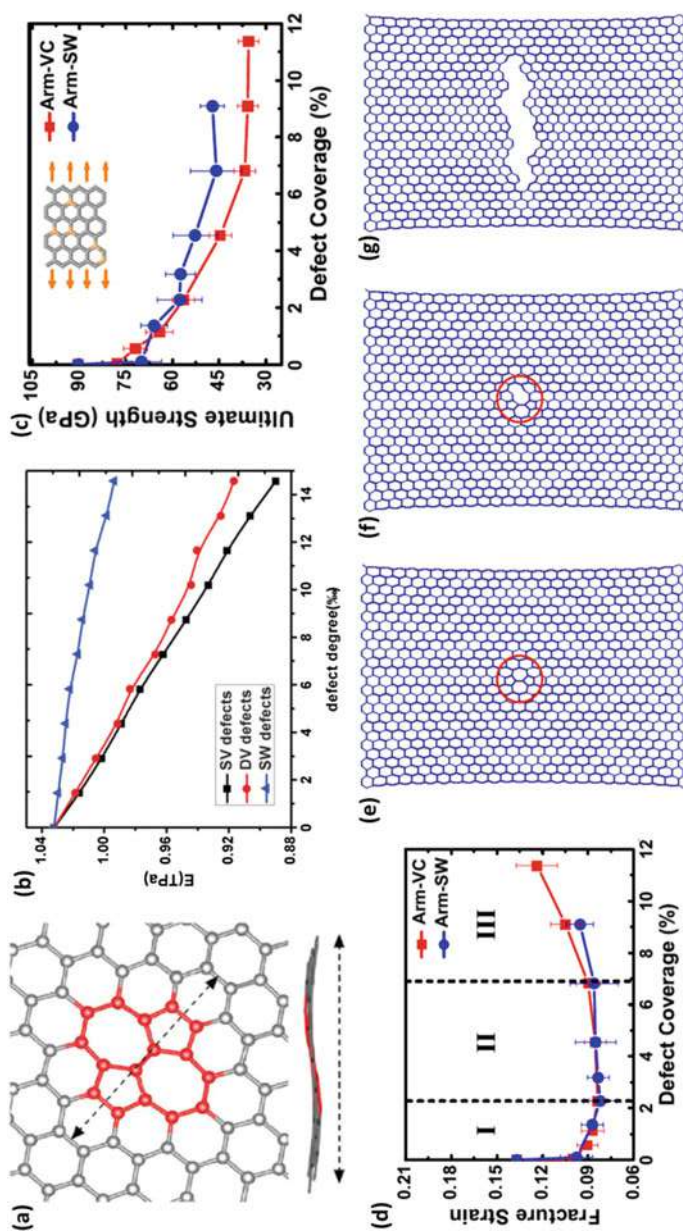
Fig. 5 (**a**) Structure of Stone-Wales (SW) defect in graphene: top (upper panel) and side (lower panel) views. The red atoms represent the SW defect. Arrows indicate the size of the buckled region (Adapted with permission from Ulman and Narasimhan 2014. Copyright 2014 American Physical Society). (**b**) The Young's modulus of graphene vs. the concentration of SW defect (triangles) and vacancies (squares, circles) (Adapted from Jing et al. 2012 with permission of The Royal Society of Chemistry). (**c, d**) The fracture strength and the ultimate strain as a function of defect concentration: SW defects (circles) and monovacancies (squares) (Adapted from Xu et al. 2013 with permission of IOP Publishing) (**e–g**). Failure mechanism of defected graphene containing a Stone-Wales defect under the strain applied along the AC direction: (**e**) initial structure, (**f**) bond rupture, and (**g**) crack propagation during the fracture (Adapted with permission from Ansari et al. 2012: Copyright 2012 Elsevier Ltd)

formed by the 90° rotation of two adjacent carbon atoms around the midpoint of the connecting bond (see Fig. 5a). This rotation transforms four neighboring hexagons into two pentagons and two heptagons, forming a (5|7|7|5) defect. The bonds between pentagon and nearby hexagons, as well as the central rotated bond (separating two adjacent heptagons), are compressed, while the bonds between heptagons and pentagons (and the nearby hexagons) are stretched. Besides, the out-of-plane buckling area ($\Delta z \approx 2$ Å, see Fig. 5a) containing the SW defect (the buckled region extends over several nanometers) lowers further the defect energy (Skowron et al. 2015; Ertekin et al. 2009).

According to the first-principles calculations, the formation energy of the SW defect in graphene is $E_f = 4.6$–5.8 eV (Skowron et al. 2015; Ertekin et al. 2009). The DFT value for the activation energy of SW transformation (90° bond rotation) is $\Delta E_b = 9.2$ eV (Ertekin et al. 2009; Li et al. 2005a), reflecting the fact that two carbon bonds must be simultaneously broken and two C-atoms must be concurrently rotated during the transformation, which is an energetically expensive process. The energy barrier for the reverse transformation is half of the direct SW transformation (~ 5 eV) (Banhart et al. 2011). The energy cost of SW transformation is prohibitively high; hence SW defects cannot form spontaneously in equilibrium conditions at room temperature. However, this high energy barrier can be significantly reduced (2–7 eV) by chemical functionalization (with transition metals (Meng et al. 2014), carbon adatoms (Wang and Ding 2013), hydrogen and hydroxyl groups (Nascimento and Nunes 2013)), applied tensile strain (Samsonidze et al. 2002), or electron irradiation (Kotakoski et al. 2011). For example, DFT calculations predict that formation of SW defects can be thermodynamically favorable at room temperature when the applied strain exceeds $\varepsilon \approx 0.2$ along the AC direction and $\varepsilon \approx 0.17$ along the ZZ direction (Skowron et al. 2015; Li et al. 2005b). In most cases, SW defects are introduced in graphene during the synthesis (rapid quench from high temperature) or under electronic irradiation (Stone and Wales 1986; Samsonidze et al. 2002). As soon as SW defects are formed under the non-equilibrium conditions, the high energy barrier for the reverse transformation ensures their presence at room temperature.

The effect of SW defects on the mechanical properties of graphene was studied by both first-principles and atomistic simulations. According to atomistic simulations, the effect of SW defects on the Young's modulus ranges from negligible $\Delta Y/Y_0 \approx -1\%$ (Jing et al. 2012; Ansari et al. 2012) to moderate $\Delta Y/Y_0 \approx -6\%$–13% (Mortazavi and Ahzi 2013; Hao et al. 2011), although the detrimental effect of SW defects becomes more prominent with increasing defect concentration (see Figs. 3a–c and 5b). Using DFT calculations, Ren et al. (Ren and Cao 2016) also found that the Young's modulus is only slightly affected by SW defects. As the SW defects prestress graphene by making it slightly anisotropic, the deformation along the AC direction is easier than along the ZZ (Ren and Cao 2016). In sharp contrast, Yanovsky et al. (2010) found a considerable reduction ($\Delta Y/Y_0 \approx -60\%$) in the Young's modulus (and failure strain) of graphene by using DFT method. This discrepancy may be partially explained by the significantly smaller size of graphene samples used in DFT calculations (Yanovsky et al. 2010).

The reduction of Young's modulus due to SW defects is noticeably smaller than that due to vacancies (see Figs. 3a and 5b). The SW defects preserve interatomic sp^2 bonding and do not contain dangling bonds; hence they weakly reduce the Young' modulus (Jing et al. 2012; Ansari et al. 2012). The Young's modulus decreases almost linearly with increasing vacancy concentration (Hao et al. 2011), while it deteriorates nonlinearly (and more gradually) with increasing SW concentration (see Figs. 3a and 5b). The hydrogenation of SW defects reduces the Young's modulus furthermore. Hydrogen atoms attached to the SW defects change their hybridization from sp^2 to sp^3, resulting in an additional reduction of Young's modulus. In contrast, hydrogenation of monovacancies eliminates dangling bonds and thus enhances the Young's modulus of defective graphene (Jing et al. 2012).

The effect of SW defects on the ultimate strain and intrinsic strength of graphene was studied by using atomistic simulations (Wang et al. 2012; Ansari et al. 2012). The SW defects can significantly reduce the failure strain and strength of graphene. Their effect on the mechanical properties is strongly anisotropic (even at very low defect concentration $c_{sw} \approx 1\%$). The failure strain reduction along the AC direction ($\Delta\varepsilon/\varepsilon_0 \approx -22\%$) is considerably larger than that along the ZZ direction ($\Delta\varepsilon/\varepsilon_0 \approx -3\%$) (Wang et al. 2012; Ansari et al. 2012). Similarly, the reduction in the intrinsic strength along the AC direction ($\Delta\sigma/\sigma_0 \approx -11\%$) is considerably larger than along the ZZ direction ($\Delta\sigma/\sigma_0 \approx -2\%$) (Ansari et al. 2012). Compared to single vacancies, SW defects lead to a smaller reduction in the failure strength and strain of graphene. Introduction of single vacancies ($c_v \approx 1\%$) decreases the intrinsic strength by $\Delta\sigma/\sigma_0 \approx -12.2\%$ along the AC direction and by $\Delta\sigma/\sigma_0 \approx -5\%$ along the ZZ direction (Ansari et al. 2012). DFT calculations predict the similar reduction in the intrinsic strength of defected graphene ($\Delta\sigma/\sigma_0 \approx -10\%$). But in contrast to MD simulations, they predict an extra sharp reduction in the failure strain ($\Delta\varepsilon/\varepsilon_0 \approx -60\%$) (Ren and Cao 2016; Yanovsky et al. 2010). This is because the stress field around the SW defects obtained by DFT calculations differs significantly from that calculated by MD simulations. The stress field causes a noticeable pre-straining of defected graphene. Due to the intricate coupling effects among the pre-strained graphene, tensile strain response, and highly nonlinear bond stretching behavior, DFT calculations predict a rather different failure strain (Ren and Cao 2016; Yanovsky et al. 2010).

The impact of SW defects on the failure behavior of graphene greatly amplifies with their accumulation. Using MD simulations, Xu et al. (2013) varied the concentration of SW defects and studied their effect on the failure characteristics of defected graphene. They found that at low concentration ($c_{sw} \approx 1\%$), the ultimate strength of graphene gradually decreases with increasing defect concentration, and it remains constant at the high concentration ($c_{sw} \approx 10\%$, see Figs. 3b and 5c). In contrast, the dependence of fracture strain on defect concentration is very unusual (see Figs. 3c and 5d): the fracture strain initially decreases at low defect concentration, reaches a plateau, and then increases at higher concentration. At higher concentration of SW defects, graphene becomes weaker but more ductile. This transition from brittle rupture to ductile fracture is caused by the specific

rearrangement of atoms around the SW defects. The rearranged atoms trap and blunt the propagating crack tips; hence the ductile fracture is observed. At high strain and elevated temperatures, SW defects participate in plastic deformation of graphene (Xu et al. 2013).

Using atomistic simulations, Ansari et al. (2012) examined the fracture of graphene containing randomly distributed SW defects (Fig. 5e). They found that SW defects are nucleation centers for fracture. Fracture initiates as a nano-crack appearing at a SW defect, where the C-C bond shared by the heptagon and nearby hexagon ruptures at the critical strain (see Fig. 5f). As the neighboring bonds break, the emerging crack propagates perpendicularly to the loading direction, as shown in Fig. 5g. The final rupture occurs when the last connecting carbon chain fails (Mortazavi and Ahzi 2013; Xu et al. 2013; Ansari et al. 2012).

Using first-principles calculations, Wang et al. (2016) investigated the energetics and the structural properties of SW defects in boronitrene. In boronitrene, a SW defect can be formed by the 90° rotation of two adjacent B- and N-atoms around the midpoint of the bond connecting them (see Fig. 6a). However, this rotation results in the formation of two homoelemental bonds (B-B and N-N), which are energetically unfavorable. The formation energy of SW defects in boronitrene $E_f = 6.5$ eV (Wang et al. 2016) is higher than that in graphene $E_f = 4.6$–5.8 eV (Skowron et al. 2015; Ertekin et al. 2009). In order to create a SW defect in boronitrene, one needs to overcome the energy barrier of $\Delta E_b = 10.4$ eV (Wang et al. 2016), which is greater than that of graphene $\Delta E_b = 9.2$ eV (Ertekin et al. 2009; Li et al. 2005a). This high energy barrier indicates that boronitrene with SW defects is stable. Though the formation energy of SW defects is rather high, they can be formed by rapid quenching from high temperature during the synthesis or under electronic irradiation (Wang et al. 2016).

The effect of SW defects on the mechanical properties and failure behavior of boronitrene was studied by Le et al. (Le and Nguyen 2014) using MD simulations coupled with finite element method (FEM). They found that uniaxial tensile stress-strain curves of pristine and defective boronitrene are nearly identical up to the fracture points (see Fig. 6b, c).

Similar to graphene, fracture of defected boronitrene is brittle, as shown in Fig. 6b, c. It indicates that the effects of SW defects on the mechanical properties of 2D materials exhibit certain universal features. Brittle fracture starts near the SW defects when boronitrene is stretched along the AC direction. But if it is strained along the ZZ direction, the fracture begins far away from the SW defects, because when the tensile strain is applied in parallel to the orientation of the most compressed central bond of the SW defect, the possibility of fracture onset at the defect is significantly reduced (Le and Nguyen 2014). The SW defects ($c_{sw} \sim 5\%$) decrease the fracture stress by $\Delta\sigma/\sigma_0 \approx -10\%$ along the AC direction (similar to graphene) and by $\Delta\sigma/\sigma_0 \approx -7\%$ along the ZZ direction (which is notably larger than that for graphene). The fracture strain is reduced by $\Delta\varepsilon/\varepsilon_0 \approx -25\%$ along the AC direction (similar to graphene) and by $\Delta\varepsilon/\varepsilon_0 \approx -18\%$ along the ZZ direction (which is considerably larger than that for graphene). Overall the effects of SW defects on the mechanical properties of boronitrene are the same as on those of
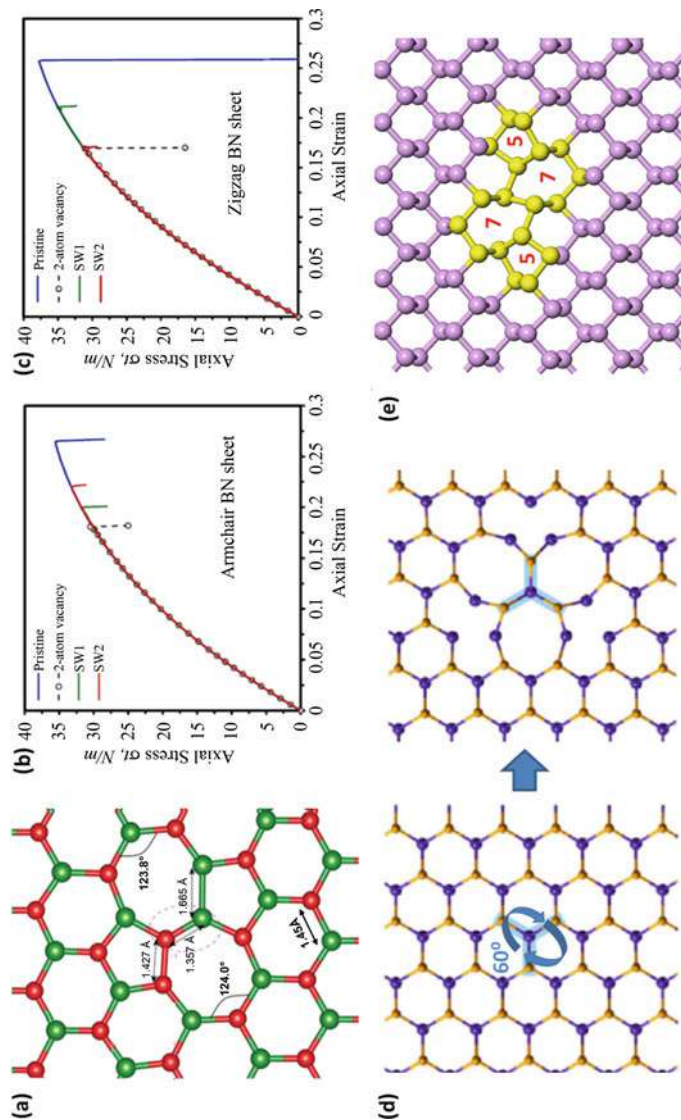
Fig. 6 (**a**) Atomistic structure of SW defects in boronitrene (N-atoms marked by red and B-atoms by green). Note the formation of two homoelemental bonds B–B and N–N bonds and the corresponding angle and bond adjustments of the pentagons and heptagons (Adapted from Wang et al. 2016 with permission of The Royal Society of Chemistry). (**b**, **c**) Stress as a function of tensile strain applied along the AC (**b**) and ZZ (**c**) directions to boronitrene sample (Adapted from Le and Nguyen 2014 with permission of AIP Publishing). (**d**) Atomistic structure of threefold rotational defects in transition metal dichalcogenides: (left panel) a triplet of M-X bonds are rotated by 60° around an M-atom and (right panel) the atomic model of a trefoil defect. Three rotated M-X bonds are marked by blue, chalcogenide X-atoms by yellow, and metallic M-atoms by magenta (Adapted from Lin et al. 2015): Open Access, Nature Communication. (**e**) The atomic model of Stone–Wales defects in phosphorene

graphene: the SW defects significantly reduce the ultimate strain and fracture stress of boronitrene but do not affect the Young's modulus to the same extent (Le and Nguyen 2014).

The formation of SW defects through 90° rotation in transition metal dichalcogenides (like in boronitrene) results in the formation of two energetically unfavorable homoelemental bonds. To avoid the formation of homoelemental bonds, Lin et al. (2015) suggested that the simultaneous 60° rotation of three bonds centered on a metal atom can sustain both the heteroelemental nature of bonding and the trigonal lattice symmetry, leading to the formation of a "trefoil"-shaped SW-like defect (see Fig. 6d). Using transmission electron microscopy (TEM), they observed these rotational defects in WS_2, WSe_2, and $MoSe_2$ at elevated temperatures. According to the first-principles calculations, the formation energy of these defects is within $E_f \approx 3$–5 eV range (Lin et al. 2015). The trefoil defects are mobile at elevated temperatures, although their mobility is much lower than that of vacancies, apparently due to the cooperative readjustment of a larger number of bonds. The diffusion barrier for migration of these defects is $\Delta E_b \approx 2.3$ eV (Lin et al. 2015), which can be overcome under electron irradiation. In addition, Lin et al. (2015) found that these SW-like defects only weakly depreciate the mechanical properties of transition metal dichalcogenides: at moderate concentration of defects ($c_{sw} = 3\%$), the Young's modulus of WSe_2 is reduced only by $\Delta Y/Y_0 \approx -4\%$ (Lin et al. 2015).

SW defects in phosphorene (see Fig. 6e) were investigated by Hu et al. (Hu and Yang 2015) using first-principles calculations. The formation energy of SW defect in phosphorene is $E_f = 1.01$ eV (Hu and Yang 2015), which is the lowest formation energy among the discussed 2D materials. By using DFT calculations, Li et al. (2016b) found that SW defects can migrate along the ZZ direction and aggregate to form line defects near grain boundaries. This migration is markedly anisotropic, since the lowest migration energy barrier along the ZZ direction ($E_m = 1.39$ eV) is half of that along the AC direction ($E_m = 2.58$ eV).

5   Dislocations

Dislocations are another type of point defects in 2D materials. There are several routes leading to the formation of dislocations. For example, in graphene, they can be formed during the chemical vapor deposition (CVD) growth, electron beam sputtering of carbon dimers, or as a combination of other defects, i.e., monovacancies and Stone-Wales defects. The effect of dislocations on the mechanical properties can be substantial. In 2D materials, dislocations form grain boundaries and lead to the out-of-plane wrapping of 2D materials, which can significantly change the fracture strength (Grantab et al. 2010). They also can serve as carriers of plastic flow under electron beam irradiation.
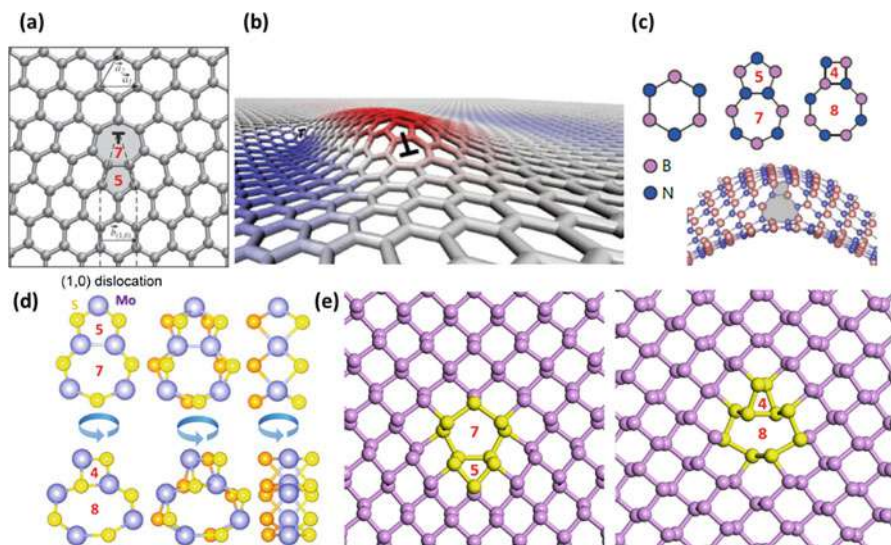
Fig. 7 (**a**) Atomic model of (1, 0) edge dislocations in graphene. Pentagons and heptagons are shaded. (Adapted with permission from Yazyev and Louie 2010. Copyright 2010 American Physical Society). (**b**) Out-of-plane buckling of graphene due to the presence of a (1, 0) edge dislocation (Adapted by permission from Macmillan Publishers Ltd.: Nature Communications Lehtinen et al. 2013, copyright 2013). (**c**) Atomic structures of edge dislocations in boronitrene (top panel) and buckling of boronitrene around (4|8) dislocation (bottom panel). (Adapted with permission from (Liu et al. 2012) (Copyright 2012 American Chemical Society). (**d**) Atomic structures of dislocation cores in monolayer MoS_2: (5|7) and (4|8) dislocations are shown in front, isometric, and side views (arrows indicate axial rotation). (Adapted with permission from Zou et al. 2013. Copyright 2013 American Chemical Society). (**e**) Edge dislocations in phosphorene: (5|7) (left panel) and (4|8) (right panel)

5.1   Dislocation Structure and Energies

Dislocations in graphene (see Fig. 7a, b) can be represented by pentagon-heptagon (5|7) pairs. Since the strain energy of a dislocation is proportional to the square of its Burgers vector, a (0,1) edge dislocation with the smallest Burgers vector (Zou et al. 2013) is the most energetically favorable one (see Fig. 7a). In this context, a Stone-Wales defect can be considered as a dislocation dipole in which two (5|7) dislocations with opposite Burgers vectors are moved by one lattice unit (Zou et al. 2013). Dislocations in graphene create a long-ranged stress field with high strain energy. Since neighboring atomic planes inhibiting out-of-plane deformations are absent in graphene, this strain energy can be partially released by its buckling around the dislocations (Alem et al. 2011; Yazyev and Louie 2010) (see Fig. 7b). According to DFT calculations, the formation energy of an isolated (5|7) edge dislocation in graphene is $E_d \approx$ 5–7.5 eV (Alem et al. 2011; Yazyev and Louie 2010).

The presence of two atomic species in boronitrene leads to a new (4|8) type of dislocations (see Fig. 7c). In odd-membered (5|7) dislocations, the alternating order of the two atomic species is interrupted when the bond is formed between atoms of the same type and charge (see N-N bond of a (5|7) dislocation in Fig. 7c). The homoelemental bonds significantly increase the formation energies of (5|7) dislocations and lead to positively (or negatively) charged dislocations. In contrast, the (4|8) dislocations, containing only even-membered rings, preserve the alternating order of B- and N-atoms (see a (4|8) dislocation in Fig. 7c) and therefore have a lower electrostatic energy. But their strain energy is larger than that of (5|7) dislocations. Similar to graphene, this strain energy is lowered via buckling of free-standing boronitrene in the out-of-plane direction around the defects (see Fig. 7c, bottom panel). The available experimental data indicate the presence of both types of dislocations but with noticeable abundance of the (5|7) dislocations along grain boundaries in boronitrene (Gibb et al. 2013).

The triple-layered structure of MoS_2 (a triangular lattice of Mo-atoms located in between two triangularly packed planes of S-atoms) brings an extra complexity to the atomic structure of its dislocations. The two types of dislocations in MoS_2 are shown in Fig. 7d: DFT calculations revealed complex concave polyhedrons extending in the third dimension (Zhou et al. 2013). As in the case of boronitrene, the homoelemental bonds in the (5|7) dislocations connect a couple of charged atoms of the same type (Mo-Mo and S-S), thus forming a high energy charged defect, while the strain energy of the (4|8) dislocation is higher than that of the (5|7) one. The high in-plane stiffness of triple-layered lattice prevents the buckling around the dislocations in MoS_2 (Zhou et al. 2013). Experimentally, both the (5|7) and the (4|8) dislocations were observed at grain boundaries of MoS_2 (Azizi et al. 2014).

In phosphorene, due to its distinctive anisotropic puckered structure, both slightly distorted (5|7) and (4|8) dislocations can be introduced (see Fig. 7e). In contrast to the strong sp^2-hybridized C-C bonds of graphene, much softer sp^3-hybridized P-P bonds of phosphorene constitute a malleable lattice structure, which allows to host dislocations without the detectable out-of-plane buckling around them (Sorkin and Zhang 2017).

5.2   Dislocation Mobility

The edge dislocations in 2D materials can migrate either via bond rotations or sputtering of lattice atoms: gliding and climbing, respectively. The glide of edge (5|7) dislocations is possible through the 90° rotation of the bond between the heptagon and its neighboring hexagon. Besides gliding, dislocation climbing is also feasible when atoms are either added or removed near the dislocation core (Lehtinen et al. 2013; Ding et al. 2007). In graphene under electronic irradiation, removal of lattice atoms happens frequently, while addition of new carbon atoms into the lattice is a rarer event (Lehtinen et al. 2013; Ding et al. 2007).

Energy cost for dissociation of SW defects into (5|7) edge dislocation pairs and their consecutive migration is prohibitively high at room temperature. Energy cost

for separation of two (5|7) dislocation pairs with an intervening hexagon is $E_d \approx$ 4–6 eV (Warner et al. 2012), although this energy decreases with increasing distance between them. Hence, plastic deformation in graphene starts through the formation of additional defects. This plastic deformation leads in the end to the formation of isolated (5|7) pairs (Skowron et al. 2015). Dislocations are also created via reconstruction of vacancies and adatoms introduced by electron irradiation.

6 Grain Boundaries

Grain boundary (GB) is a line defect in 2D materials. Typically, large-scale films of graphene produced by CVD are polycrystalline – they are composed of single-crystalline domains (grains) with different lattice orientations. This variety in grain orientations unavoidably leads to the presence of grain boundaries – the interfaces between the grains.

GB can be regarded as an array dislocations arranged in a linear way. In general, GBs can be classified by the value of the tilt angle separating the adjacent grains: low-angle and high-angle grain boundaries. In low-angle GBs, dislocation cores are separated by sufficiently large distance, while in high-angle GBs, dislocation cores are very close, and they can even overlap. The tilt angle is set in the course of non-equilibrium growth processes, rather than by equilibrium energetics. GBs in 2D materials are non-thermodynamic objects since they do not emerge as a result of thermal fluctuations. They are considered as kinetically "frozen" during growth process.

6.1 Construction of GB

A few 2D samples with GB composed of linear arrays of dislocations are shown in Fig. 8. In order to construct a sample with GB, a group of atoms, forming "wedge"-like region, is removed. Subsequently, the two adjacent crystalline domains are pivoted around a selected origin in the clockwise and anti-clockwise directions, respectively, and welded seamlessly together (with a minor bond length adjustment). The created dislocation cores (see Fig. 8) form a vertically oriented linear array of dislocations. This is the preferred alignment of identical edge dislocations (Liu and Yakobson 2010). In order to implement periodic boundary conditions in the direction perpendicular to the created grain boundary, the constructed sample is duplicated. The duplicate is rotated by 180° and joined to the original one. Therefore, the periodic boundary conditions are implemented by constructing a sample with two grain boundaries formed by linear arrays of dislocations oriented in the opposite directions (Liu and Yakobson 2010).

The distance between the nearest dislocations along GBs is associated with the angular rotation of the adjacent grains, which is described by tilt (misalignment or misorientation) angle, α. Low-angle grain boundaries are those with the misorientation angle less than about $\alpha \leq 15°$, while high-angle GBs have the larger tilt angle
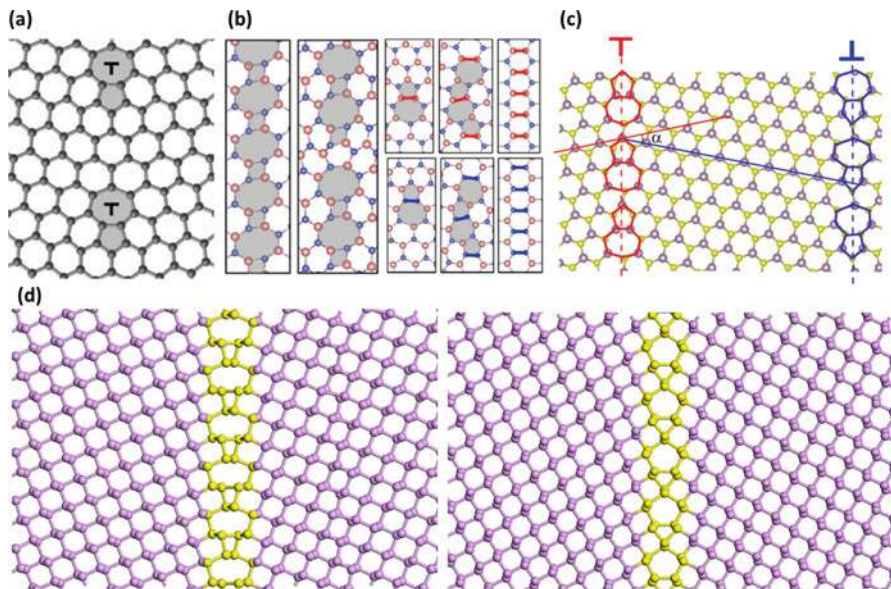
Fig. 8 Grain boundaries: (**a**) graphene (Taken from Grantab et al. 2010), (**b**) boronitrene (Adapted with permission from Liu et al. 2012. Copyright 2015 American Chemical Society), (**c**) MoS_2 (Adapted with permission from Yu et al. 2015b. Copyright 2012 American Chemical Society), (**d**) phosphorene. Arrays of (4|8) dislocations (left panel) and (5|7) dislocations (right panel) form GBs between AC-oriented and ZZ-oriented crystalline domains

$\alpha > 15°$ (Liu and Yakobson 2010). The constructed GB is often oriented either along the AC or the ZZ direction, and hence two primary types of GBs: AC-GB and ZZ-GB, are usually studied. In addition, introduction of GBs in graphene and boronitrene leads to out-of-plane warping and inflection near the GBs (Liu and Yakobson 2010), which effectively relieve the in-plane stress by "screening" the in-plane elastic fields of dislocations (Liu and Yakobson 2010). The degree of warping is characterized by inflection angle, which in some cases can reach up to $\sim 72°$ in graphene and $\sim 77°$ in boronitrene, thus significantly reducing their mechanical strength (Zhang et al. 2012).

In order to construct a polycrystalline sample with a large number of grains of various sizes and random crystal orientations, the Voronoi cell method (Voronoi tessellation) is often applied (Yazyev and Louie 2010). The Voronoi tessellation is an assembly of convex polygons separated by line segments (walls perpendicular to the lines connecting the grain centers). The grain centers are seeded randomly or uniformly inside the simulation box. Each convex polygon (Voronoi cell) represents a grain domain filled with atoms. Usually, the grain sizes are normally distributed around the specified average grain size. Each grain is given an arbitrary in-plane orientation in order to mimic the grain orientations observed in experiments. The constructed polycrystalline sample is usually annealed at a high temperature and

then rapidly quenched to room temperature to remove any abnormally low or high density regions at grain boundaries (defined as the atomically thick layers adjacent to the Voronoi cell walls).

6.2   Grain Boundaries: Structures and Energies

Grain boundaries in graphene are generally composed of linear arrays of aligned (5|7) edge dislocations (see Fig. 8a). The array of (5|7) dislocations bisects the tilt angle between the two adjacent graphene domains. This configuration is optimal and energetically favorable: departure from the bisect increases the interface energy (Liu and Yakobson 2010). Even so, experimentally observed GBs in graphene are not strictly symmetrical and occasionally have serpentine-like geometry (Huang et al. 2011). Note that graphene GBs are composed only of (5|7) dislocations since their formation energy is considerably lower than that of (4|8) dislocations. Introduction of GBs in free-standing graphene leads to out-of-plane warping near the low-angle GBs, which relieves the in-plane strain at GBs (Yazyev and Louie 2010; Liu and Yakobson 2010; Kurasch et al. 2012). The high-angle grain boundaries are practically flat, as confirmed by the experimental data (Yazyev and Louie 2010).

In polycrystalline graphene, adjacent grain domains are competing to add carbon atoms to their outer boundaries at each other's expense. This process results in GB migration (via gliding and climbing of (5|7) dislocations), which reduces the fraction occupied by GBs in the sample and decreases the free energy contributed by them. The GB migration is observed in experiments and simulations only in the presence of significant GB curvature (Kurasch et al. 2012).

GBs in boronitrene formed by linear arrays of (5|7) and (4|8) dislocations are shown in Fig. 8b. In boronitrene, both the strain energy (associated with defect structure) and the type of bonding (homo- or heteroelemental bonding between B- and N-atoms) define the formation energy of dislocations. Energetically unfavorable homoelemental B–B and N–N bonding in the (5|7) dislocations makes the (4|8) dislocations, with heteroelemental B-N bonding, equally preferable (Liu et al. 2012). The higher strain energy of the (4|8) dislocations is compensated by the out-of-plane buckling along GBs (Ding et al. 2014) observed experimentally (Gibb et al. 2013). Hence, boronitrene is capable to avoid the unfavorable homoelemental bonding at the cost of increasing lattice strain. First-principles calculations indicate that ZZ-GBs are composed of (4|8) dislocations, while AC-GBs consist of (5|7) dislocations (Liu et al. 2012).

An array of dislocations forming GBs between the tilted adjacent grains in MoS_2 is shown in Fig. 8c. In contrast to boronitrene, where the (4|8) dislocations with large Burgers vectors are feasible due to its high flexibility, GBs in MoS_2 are formed only by (5|7) dislocations with small Burgers vectors. The reason is that the high flexural rigidity of triple-layered MoS_2 excludes the possibility of the out-of-plane buckling along its GBs (Zou et al. 2013).

Similar to boronitrene, GBs in phosphorene are formed by (4|8) dislocations along the ZZ direction (see Fig. 8d, left panel) and (5|7) dislocations along the

AC direction (see Fig. 8d, right panel), where the dislocation rings are highlighted. Phosphorene with GBs maintains its original planar shape, without any kind of out-of-plane wrapping, wrinkles, or inflections due to its unique puckered structure (see Fig. 8d). Its lattice structure, defined by tetrahedral geometry of sp^3-hybridized P-P bonds, is very flexible and capable to accommodate arrays of (4|8) and (5|7) dislocations forming GBs between neighboring crystalline domains. The energy per unit length of grain boundary E_{gb} is calculated according to:

$$E_{gb} = \frac{E_{tot} - NE_b}{2L}$$

where E_{tot} is the total energy of the simulation cell containing two grain boundaries, N is the total number of atoms, $2L$ is the total length of the two grain boundaries, and E_b is the energy per atom in defect-free bulk 2D material. The general observation is that the GB energy increases in direct proportion to the tilt angle, i.e., it increases in proportion to the dislocation density at GBs.

Graphene GB energies were studied by using both first-principles and atomistic simulations. The energy as a function of tilt angle, α, for symmetric GBs formed by an array of evenly spaced dislocations (Liu and Yakobson 2010) is shown in Fig. 9a. The GB energy can be approximately fitted to sine function. At low-angle GBs, the energy increases proportionally with the tilt angle. The GB energy is proportional to the dislocation density, which increases with the tilt angle (the larger the tilt angle, the smaller the distance between the nearby (5|7) dislocations, hence the larger the dislocation density). When the distance between adjacent dislocations reaches the minimum (i.e., they separated only by a single hexagon forming an array of – (5|7)6(5|7)6(5|7)6(5|7)6 – see Fig. 9a), the GB energy reaches its maximum. Subsequently, the energy decreases since the spacing between the neighboring dislocations increases again as shown in Fig. 9a, which eventually leads to the wavy interface between the adjacent grains.

The presence of two types of atoms in boronitrene brings more complexity to the energetics of its GBs in comparison to graphene. Three classes of boronitrene GBs can be constructed: symmetric GBs, where the grains have mirror symmetry about the bisecting GB; asymmetric GBs, where a grain on one side is an inverse reflection of the grain on the other side (with every B-atom replaced by N-atom and vice versa); and nonsymmetric GBs. Moreover, the homoelemental B-B and/or N-N bonds may appear along the GBs, and the unequal numbers of B-B and N-N bonds along a GB line cause elemental polarity (B-rich or N-rich GBs). Even so, the GB energy of boronitrene with symmetric grain boundaries is similar to that of graphene (Ding et al. 2014) (see Fig. 9b): It increases with increasing tilt angle, reaches the maximum, and then decreases once the tilt angle exceeds $\alpha \approx 60°$. This change of the GB energy reflects the hexagonal periodicity of boronitrene lattice: when the tilt angle is larger than $60°$, the structure would come closer to its original configuration at $\alpha \approx 0°$ (or $\alpha \approx 120°$) due to the angular periodicity of a hexagonal ring.

According to the first-principles calculations, the GB energy in MoS_2 is relatively low (see Fig. 9c), which signifies the easy formation of GBs and explains the relative
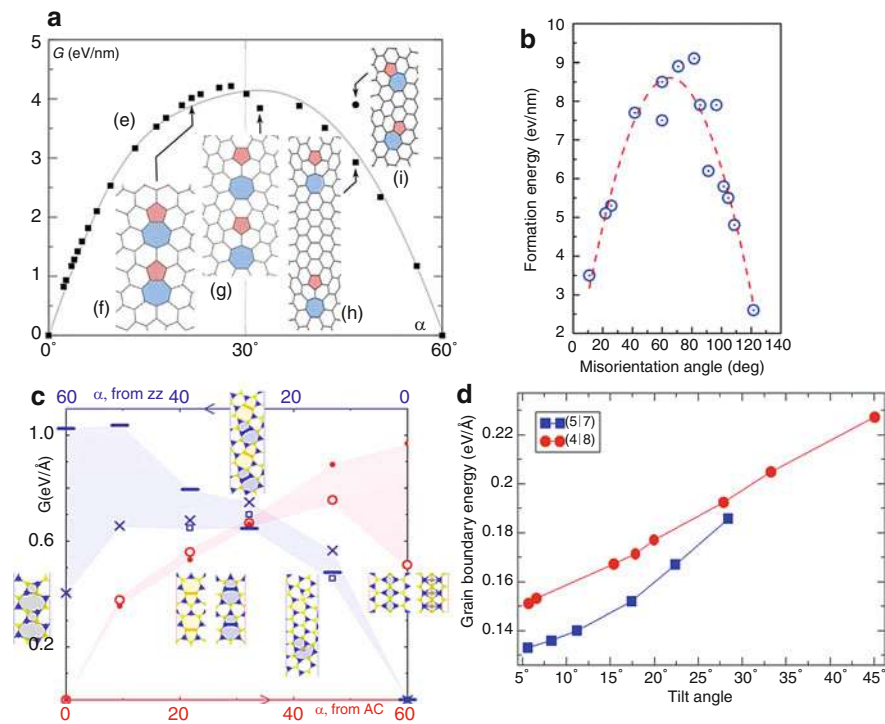
Fig. 9 Energy of grain boundary per unit length as a function of tilt angle: (**a**) graphene (Adapted with permission from Liu and Yakobson 2010. Copyright 2010 American Chemical Society), (**b**) boronitrene (Adapted from Ding et al. 2014 with permission of The Royal Society of Chemistry), (**c**) MoS$_2$ (Adapted with permission from Yu et al. 2015b. Copyright 2012 American Chemical Society), and (**d**) phosphorene formed by (4|8) defect pairs (red circles) and (5|7) defect pairs (blue squares) as a function of tilt angle (Reprinted from Sorkin and Zhang 2017 with permission of IOP Publishing)

abundance of GBs in MoS$_2$ observed experimentally (Yu et al. 2015b). The GB energies for both AC-GB and ZZ-GB are shown in Fig. 9c.

Grain boundary energy in phosphorene as a function of tilt angle is shown in Fig. 9d. The energy per unit length as a function of tilt angle is shown in Fig. 9c for the phosphorene samples containing grain boundaries composed of (4|8) dislocations (red circles) and (5|7) defect pairs (blue squares). The GB energies for the two types of grain boundaries are quite similar (Sorkin and Zhang 2017). Nonetheless, the grain boundaries formed by (5|7) defects have slightly lower energy, and therefore they are slightly more stable. Similar to graphene (Liu and Yakobson 2010) and boronitrene (Ding et al. 2014), GB energy in phosphorene increases with the increase in the tilt angle. The GB energies in phosphorene are lower than those in graphene (Yazyev and Louie 2010; Liu and Yakobson 2010), boronitrene (Hao and Chen 2016), and MoS$_2$ (Zou et al. 2013), implying a higher stability of GBs in phosphorene than in other 2D materials.

6.3   Failure Mechanism of Grain Boundary

Investigation on the effects of GBs on the mechanical properties of graphene has ended up with a few unexpected discoveries. It was at first predicted and then experimentally verified that contrary to standard reasoning, graphene with large-angle tilt boundaries (high GB defect density) is almost as strong as the pristine one and is much stronger than that with small-angle boundaries (see Fig. 10) (Grantab et al. 2010). These surprising results cannot be explained by Griffith-type fracture mechanics criterion since it does not take the details of bond stretching and breaking into account. It was observed that the failure always starts at GBs of graphene where the bonds shared by hexagon-heptagon rings of the (5|7) dislocations rupture (see Fig. 10a). Subsequently, nanovoids are generated around these broken bonds. The nanovoids gradually grow, leading to the fracture of graphene (Jhon et al. 2012). These broken bonds are more pre-strained at small-angle GBs than those at large-angle GBs, which can better accommodate the strained bonds (Grantab et al. 2010). The closer the (5|7) dislocations are to each other at large-angle GBs, the more significant the mutual cancellation of their overlapping compressive and tensile stress fields. Therefore, the tensile pre-strain of the critical bonds is reduced in large-angle GBs. The initial pre-strain of these critical bonds defines the intrinsic strength, which increases with the tilt angle (see Fig. 10b). But, this GB strengthening with tilt angle is valid only for symmetric GBs formed by linear array of evenly spaced (5|7) dislocations. It was later found that GB strength can decrease with the tilt angle for asymmetrical GB formed by irregularly spaced dislocations. This indicates that it is not just the density of dislocation defects that affects the mechanical properties but the detailed arrangements of defects are also important (Wei et al. 2012).

Similar to graphene, the high-angle symmetrical GBs of boronitrene with a larger dislocation density are stronger than the low-angle GBs (Ding et al. 2014). The atomistic simulations also demonstrated that failure of boronitrene is similar to that of graphene. At the failure strain, boronitrene with GBs undergoes a brittle fracture: the rupture occurs sharply without any noticeable preceding changes in the rate of tensile stretching (Ding et al. 2014). This sudden failure mainly stems from the rupture of overstretched homoelemental GB bonds. At grain boundaries formed by array of edge dislocations, homoelemental B-B and N-N bonds occasionally appear, which can be broken much easier than heteroelemental ones. Under applied tensile strain, they rupture first (Becton and Wang 2015). This fracture of homoelemental bonds leads to cumulative stress concentration in the remaining heteroelemental bonds. Since there are only few relatively highly stress-concentrated bond connections between the grains, if one part fails, the rest fails in a rapid fashion. The cracks in boronitrene propagate typically along the ZZ direction (Xiong et al. 2015; Ding et al. 2014). The fracture strength of boronitrene also drops rapidly with the increase in the inflection angle at its GBs. This is because the local sp^3 hybridization at the inflected GBs reduces the fracture strength (Ding et al. 2014).
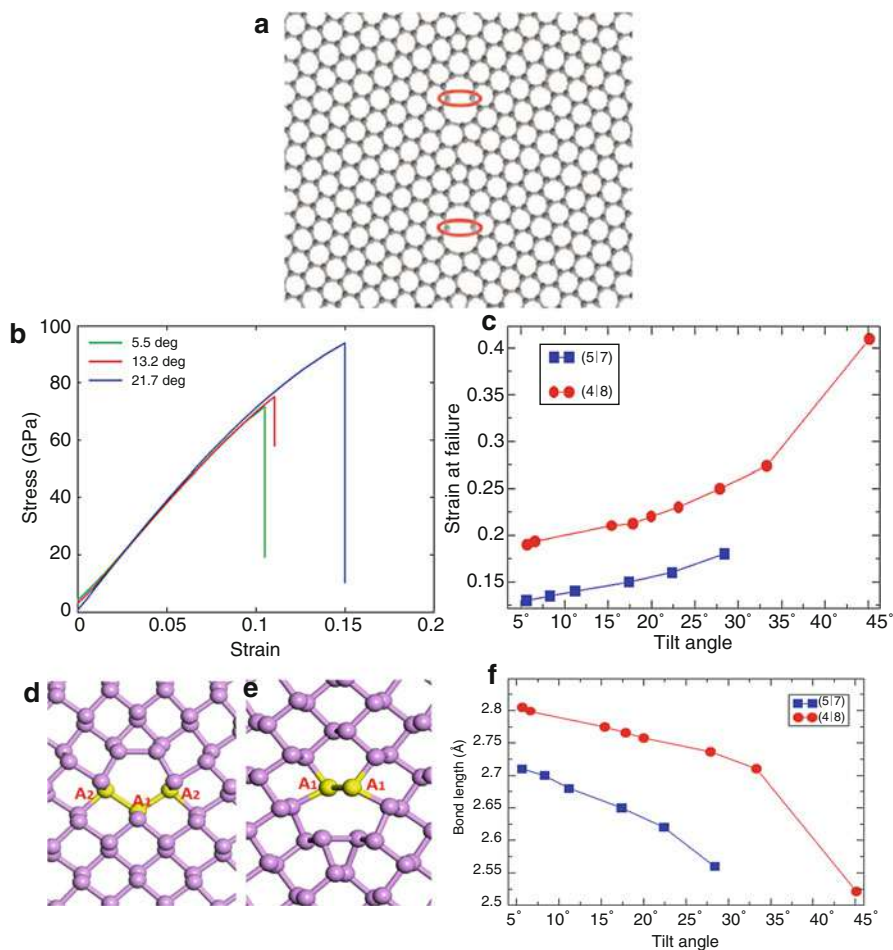
Fig. 10 (**a**) Initial stage of failure in AC-oriented graphene with GBs. The broken critical hexagon-heptagon bonds of (5|7) defects are outlined (Grantab et al. 2010). (**b**) The stress-strain curves of ZZ-oriented graphene sheets strained perpendicular to the GBs (Grantab et al. 2010). (**c**) Failure strain as a function of tilt angle for phosphorene with linear grain boundaries formed by evenly spaced (4|8, red circles) and (5|7, blue squares) edge dislocations. (**d**) Critical bonds $|A_1A_2|$ in heptagon of a (5|7) defect pair. (**e**) Critical bonds $|A_1A_1|$ in octagon of a (4|8) defect pair rupture first at the onset of tensile failure. (**f**) Initial bond length of the critical $|A_1A_1|$ bonds in (4|8) defect pairs (red circles) and of the $|A_1A_2|$ bonds in (5|7) defect pairs (blue squares) as a function of tilt angle at zero strain. (**d–f**) (Reprinted from Sorkin and Zhang 2017 with permission of IOP Publishing)

Effects of grain boundaries on the mechanical behavior of MoS_2 subjected to uniaxial tensile were studied by Dang et al. (Dang and Spearot 2014) using MD simulations. They found that similar to other 2D materials, the failure initiates at GBs (Dang and Spearot 2014). Similar to graphene, the nucleated crack mainly propagates along the ZZ direction.

The unusual failure characteristics of tilt GBs in graphene and boronitrene are also shared by phosphorene (Sorkin and Zhang 2017) (see Fig. 10c), which signifies that it may be universal for all 2D materials. In phosphorene, large-angle GBs are stronger than low-angle ones. In analogy with graphene, the failure is due to the rupture of the most pre-strained bonds of the (5|7) and (4|8) dislocations forming GBs. The first bonds rupturing at the onset of brittle failure in phosphorene are the most stretched ones in (5|7) and (4|8) dislocations (see the highlighted bonds in Fig. 10d, e). These bonds connect the heptagons and the nearby hexagons in (5|7) defects (as well as the octagons and the adjacent hexagons in (4|8) defects) just as in graphene and boronitrene. Phosphorene with large-angle GBs formed by densely packed dislocations is much stronger than that with low-angle grain boundaries. Similar to graphene and boronitrene, the higher defect concentration at the grain boundaries does not inevitably lead to greater deterioration of its mechanical properties. The large-angle GBs are better able to accommodate the pre-strain and prevent failure that originates from the rupture of the critical bonds of the (4|8) and (5|7) edge dislocations (see Fig. 10f). The closer the (4|8) or (5|7) dislocations are to each other, the more significant the mutual cancellation of their overlapping compressive and tensile stress fields is. Therefore, the tensile pre-strain of the critical bonds is reduced in larger-angle GBs. The initial pre-strain of the critical bonds defines the intrinsic strength of defected phosphorene (Sorkin and Zhang 2017).

6.4   Failure Mechanism of Polycrystalline Materials

After examination of the effects of GB on the mechanical properties of 2D materials, the next logical step is to examine the effects of GBs in polycrystalline samples. The terminations of GBs or GB junctions (one-dimensional regions where three or more adjacent grain boundaries meet) in polycrystalline materials are crucial for failure: they generally serve as accumulators of pre-strain (stress concentrators), sites for crack nucleation, and therefore lower the fracture strength of 2D materials (see Fig. 11).

It was found that in polycrystalline graphene, the GB junctions are the "weakest links." In particular, triple junctions (TJs) of graphene or vacancy clusters at GBs are the nucleation centers for cracks (Li et al. 2015). Initial bond fracture at TJs or GBs leads to nano-size void formation, which in turn results in crack nucleation. Occasionally, a row of nano-size voids appear along GBs (see Fig. 11a), and when applied strain increases, these voids coalesce into cracks. The cracks in graphene propagate along GBs and traverse grain domains (see Fig. 11a), indicating that fracture strength of GBs is comparable to that of defect-free bulk (Li et al.
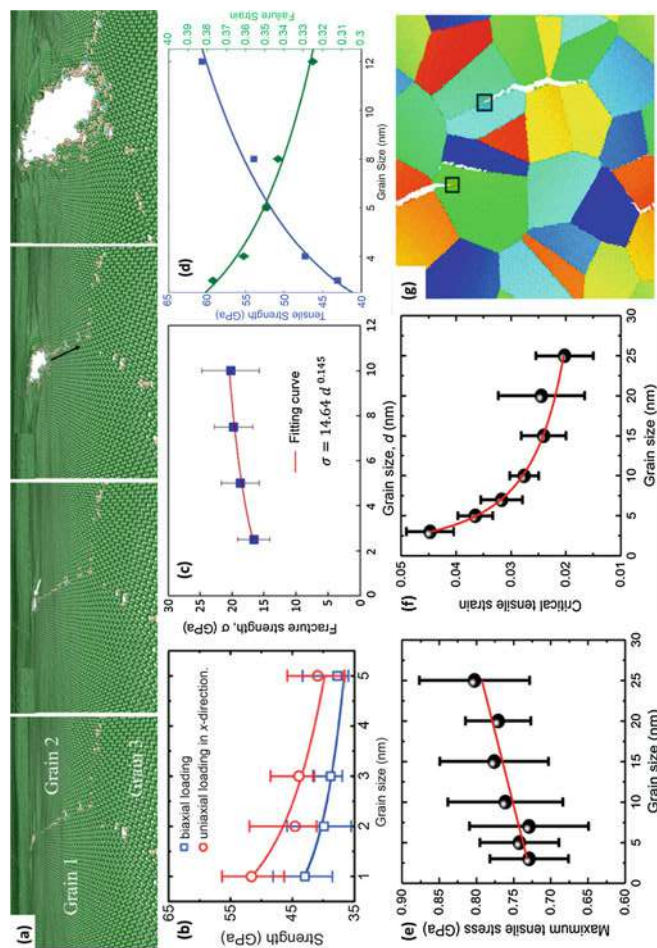
Fig. 11 (**a**) Snapshots of crack nucleation and unzipping mechanism leading to brittle intergranular fracture in graphene. The arrow indicates the unzipping direction (Adapted from Cao and Qu 2013 with permission of AIP Publishing). (**b**) Pseudo Hall-Petch: size-dependence of fracture strength for polycrystalline graphene monolayer with different grain sizes (Reprinted with permission from Song et al. 2013. Copyright 2013 American Chemical Society). (**c**) Inverse Hall-Petch: size-dependence of fracture strength for polycrystalline graphene monolayer with different grain sizes (Reprinted with permission from Li et al. 2014. Copyright 2014 American Chemical Society). (**d**) Size-dependence of tensile fracture strength (blue line) and failure strain (green) for polycrystalline boronitrene with different grain sizes (Adapted from Becton and Wang 2015 with permission of the PCCP Owner Societies). (**g**, **e**) Polycrystalline phosphorene: fracture strength (**g**) and tensile failure strain (**e**) as a function of grain size (Adapted from Cao et al. 2017). (**f**) Crack nucleation along grain boundaries in polycrystalline phosphorene (Adapted from Cao et al. 2017 with permission of IOP Publishing)

2015). The vacancies and dislocations seriously complicate crack propagation in polycrystalline graphene since they not only induce stress field but also lead to wrinkling deformation affecting its fracture mechanism (Liu and Yakobson 2010). Under the applied tensile strain, the deformation in polycrystalline graphene is completely elastic up to the brittle fracture since the (5|7) edge dislocations are immobile, and therefore plastic deformation is inhibited (Cao and Qu 2013). The predicted fracture strain ($\varepsilon_{cr} \approx 0.16$) in polycrystalline graphene, although lower than in monocrystalline graphene ($\varepsilon_{cr} \approx 0.23$), is still the highest ultimate strain in polycrystalline materials (Cao and Qu 2013). The transition to plastic deformation occurs only at elevated temperatures (T $\geq$ 1000 K). MD simulations (Yang et al. 2015) predict that the ductility of polycrystalline graphene can deteriorate unexpectedly due to accelerated bond breaking at high temperatures. Furthermore, if the applied strain rate is sufficiently fast at elevated temperatures, then polycrystalline graphene undergoes brittle fracture again. The reason is that the short relaxation time prevents the bond rotations and vacancy coalescence, resulting in simultaneous collective bond breaking (Yang et al. 2015).

A number of studies focusing on the strength of polycrystalline graphene show that there are two opposing trends in the dependence of the tensile fracture strength on the grain size: pseudo Hall-Petch, where the smaller the grain size, the larger the fracture strength (Song et al. 2013) (see Fig. 11b), and inverse Hall-Petch, where the smaller the average grain size, the lower the tensile failure strength (Li et al. 2014; Mortazavi and Cuniberti 2014) (see Fig. 11c). At the moment, the consensus is that the fracture strength of polycrystalline graphene is sensitive to the details of atomic structure (number, type, and distribution of various defects), which are dependent on the sample generating methods. The reduction in failure strength in polycrystalline graphene is up to $\Delta\sigma/\sigma_0 \approx -40\%$ for the smallest grains in the case of the inverse Hall-Petch dependence (Mortazavi and Cuniberti 2014).

The Young's modulus diminishes with the decrease in the average grain size. The Young's modulus of graphene with ultrafine grains (average grain size $\Delta \approx 1$ nm) can be reduced by $\Delta Y/Y_0 \approx -50\%$ in comparison with pristine monocrystalline graphene. However, this effect is negligible ($\Delta Y/Y_0 \approx -3\%$) when the average grain size exceeds $\Delta \approx 10$ nm (Mortazavi and Cuniberti 2014). The failure strain increases gradually when the average grain size decreases. This is because the smaller the average grain size, the more uniformly distributed the defective areas (containing GBs, TJs, and nano-sized voids) over the sample, making it more homogeneous. The more homogeneous the sample, the more uniformly it deforms under loading conditions and therefore fails at a higher failure strain. Besides that, the formation of monoatomic carbon chains between the adjacent GBs, separated by crack nucleation, also allows for polycrystalline graphene to withstand a higher strain and leads to the formation of irregular crack edges (Mortazavi and Cuniberti 2014).

The fracture strength of polycrystalline boronitrene is less studied. Atomistic simulations indicate the inverse Hall-Petch dependence of its failure strength on the grain size. There is a noticeable reduction in the fracture strength (and Young's modulus) when the average grain size decreases (Becton and Wang 2015) (see

Fig. 11d). The smaller the average grain size, the larger the percentage area (fraction) of GBs. These GBs are not as stable as in the pristine crystalline lattice and may serve as sources for crack nucleation. Thus a larger number of GBs reduce the fracture strength of polycrystalline boronitrene (Becton and Wang 2015). Like in graphene, the brittle failure of boronitrene is due to abrupt rupture of the bonds at TJs and GBs (Becton and Wang 2015).

Similar to graphene and boronitrene, fracture in MoS_2 is brittle: as soon as one bond ruptures near TJs or GBs, the released stress from the broken bond raises the stress on the nearby bonds, which consequently rupture. This process proliferates, causing a cascade of bond breaking and resulting in the brittle fracture of polycrystalline MoS_2 (Wang et al. 2015). The underlying reason for the brittleness is that the bond breaking is a localized event since the stress concentration near the ruptured bonds cannot be redistributed due to immobile dislocations in MoS_2. Thus, the cascade of bond failure results in crack propagation, which preferably takes place along the ZZ direction of MoS_2 (Wang et al. 2015).

The mechanical properties and the failure mechanism of polycrystalline phosphorene were studied by Cao et al. (2017) using atomistic simulations. They found that the Young's modulus of polycrystalline phosphorene decreases when the average grain size is reduced. Polycrystalline phosphorene also exhibits an inverse Hall-Petch dependence: its tensile fracture strength decreases with decreasing grain size (see Fig. 11e). The fracture strength of polycrystalline phosphorene is reduced by $\Delta\sigma/\sigma_0 \approx -20\%$ relative to monocrystalline defect-free phosphorene (Cao et al. 2017). In contrast, the fracture strain increases with reduced grain size (see Fig. 11f). The failure mechanism of polycrystalline phosphorene is also similar to that of graphene and boronitrene: the initial bond breaking and subsequent crack nucleation occur mainly at triple junctions and grain boundaries (nearly perpendicular to the stretching direction). The initiated cracks rapidly propagate along the GBs and occasionally across the grain interior (see Fig. 11g). In the grain interior, the cracks in phosphorene typically move along the ZZ direction.

7 Conclusions

Defects in 2D materials have a profound impact on their mechanical properties and failure behavior. In this chapter, we discussed the effect of four fundamental defects: vacancies, Stone-Wales defects, dislocations, and grain boundaries. Four representatives of 2D materials: graphene, boronitrene, molybdenum disulfide, and phosphorene, were considered. The intrinsic defects in these 2D materials have diverse atomic structures and energetics, and their effects on the mechanical properties also vary. Nevertheless, many of the discussed effects of defects on the mechanical properties of 2D materials such as softening of elastic moduli, deterioration in fracture strength, and reduction of ultimate strain are manifestly universal. Besides these effects, there are a few novel effects shared by 2D materials like the unexpected strengthening of high-angle grain boundaries and inverse Hall-Petch dependence of fracture strength on grain size in polycrystalline samples.

Despite the evident progress in studying defected 2D materials, many challenging questions still remain open. Here we mention just a few of them: How can one deliberately engineer defects in a controllable way? Can one control the fracture behavior of 2D materials by manipulating defects, chemical functionalization, and strain engineering? How do the out-of-plane effects induced by defects influence their properties? Can we enhance the toughness of 2D materials by defect engineering?

The family of 2D materials is rapidly expanding. Here, we only discussed four representative 2D materials. Certainly, defects in other 2D materials, such as silicene, germanene, borophene, and metal-organic frameworks, which are not included in the present review, are equally interesting. In addition, defect structures, defect interactions, and their impact on the mechanical properties of newly emerging 2D materials need also great research attention.

Finally, the grand challenge is to develop new models and computational methods for multiscale simulations of defects in 2D materials. Understanding the interactions between a variety of defects across length and time scales will continue to motivate further research.

References

Akinwande D, Brennan CJ, Bunch JS, Egberts P, Felts JR, Gao H, Huang R, Kim J, Li T, Li Y, Liechti KM, Lu N, Park HS, Reed EJ, Wang P, Yakobson BI, Zhang T, Zhang Y-W, Zhou Y, Zhu Y (2016) A review on mechanics and mechanical properties of 2D materials – graphene and beyond. Extreme Mech Lett 13:1–86. https://doi.org/10.1016/j.eml.2017.01.008

Alem N, Erni R, Kisielowski C, Rossell MD, Gannett W, Zettl A (2009) Atomically thin hexagonal boron nitride probed by ultrahigh-resolution transmission electron microscopy. Phys Rev B 80:155425. https://doi.org/10.1103/PhysRevB.80.155425

Alem N, Yazyev OV, Kisielowski C, Denes P, Dahmen U, Hartel P, Haider M, Bischoff M, Jiang B, Louie SG, Zettl A (2011) Probing the out-of-plane distortion of single point defects in atomically thin hexagonal boron nitride at the picometer scale. Phys Rev Lett 106:126102. https://doi.org/10.1103/PhysRevLett.106.126102

Ansari R, Ajori S, Motevalli B (2012) Mechanical properties of defective single-layered graphene sheets via molecular dynamics simulation. Superlattice Microst 51:274–289. https://doi.org/10.1016/j.spmi.2011.11.019

Ataca C, Ciraci S (2011) Functionalization of single-layer MoS_2 honeycomb structures. J Phys Chem C 115:13303–13311. https://doi.org/10.1021/jp2000442

Azizi A, Zou X, Ercius P, Zhang Z, Elías AL, Perea-López N, Stone G, Terrones M, Yakobson BI, Alem N (2014) Dislocation motion and grain boundary migration in two-dimensional tungsten disulphide. Nat Commun 5:4867. https://doi.org/10.1038/ncomms5867

Banhart F, Kotakoski J, Krasheninnikov AV (2011) Structural defects in graphene. ACS Nano 5:26–41. https://doi.org/10.1021/nn102598m

Becton MD, Wang X (2015) Grain-size dependence of mechanical properties in polycrystalline boron-nitride: a computational study. Phys Chem Chem Phys 17:21894–21901. https://doi.org/10.1039/C5CP03460D

Cai Y, Ke Q, Zhang G, Yakobson BI, Zhang YW (2016) Highly itinerant atomic vacancies in phosphorene. J Am Chem Soc 138:10199–10206. https://doi.org/10.1021/jacs.6b04926

Cao A, Qu J (2013) Atomistic simulation study of brittle failure in nanocrystalline graphene under uniaxial tension. Appl Phys Lett 102:071902. https://doi.org/10.1063/1.4793088

Cao P, Wu J, Zhang Z, Ning F (2017) Mechanical properties of monocrystalline and polycrystalline monolayer black phosphorus. Nanotechnology 28:045702. https://doi.org/10.1088/1361-6528/28/4/045702

Chen WH, Chen IC, Cheng HC, Yu CF (2017) Influence of structural defect on thermal-mechanical properties of phosphorene sheets. J Mater Sci 52:3225–3232. https://doi.org/10.1007/s10853-016-0611-3

Dang KQ, Spearot DE (2014) Effect of point and grain boundary defects on the mechanical behavior of monolayer MoS_2 under tension via atomistic simulations. J Appl Phys 116:013508. https://doi.org/10.1063/1.4886183

Degefe M, Parashar A (2016) Effect of non-bonded interactions on failure morphology of a defective graphene sheet. Mater Res Express 3:045009. https://doi.org/10.1088/2053-1591/3/4/045009

Dettori R, Cadelano E, Colombo L (2012) Elastic fields and moduli in defected graphene. J Phys Condens Matter 24:104020. https://doi.org/10.1088/0953-8984/24/10/104020

Ding F, Jiao K, Wu M, Yakobson BI (2007) Pseudoclimb and dislocation dynamics in superplastic nanotubes. Phys Rev Lett 98:075503. https://doi.org/10.1103/PhysRevLett.98.075503

Ding N, Wu C-ML, Li H (2014) The effect of grain boundaries on the mechanical properties and failure behavior of hexagonal boron nitride sheets. Phys Chem Chem Phys 16:23716–23722. https://doi.org/10.1039/c4cp02521k

Do LG, Wang CZ, Yoon E, Hwang NM, Kim DY, Ho KM (2005) Diffusion, coalescence, and reconstruction of vacancy defects in graphene layers. Phys Rev Lett 95:205501. https://doi.org/10.1103/PhysRevLett.95.205501

El-Barbary AA, Telling RH, Ewels CP, Heggie MI, Briddon PR (2003) Structure and energetics of the vacancy in graphite. Phys Rev B 68:144107. https://doi.org/10.1103/Physrevb.68.144107

Ertekin E, Chrzan DC, Daw MS (2009) Topological description of the Stone-Wales defect formation energy in carbon nanotubes and graphene. Phys Rev B 79:155421. https://doi.org/10.1103/PhysRevB.79.155421

Faccio R, Mombrú AW (2012) Magnetism in multivacancy graphene systems. J Phys Condens Matter 24:375304. https://doi.org/10.1088/0953-8984/24/37/375304

Feng L-P, Su J, Liu Z (2014a) Effect of vacancies on structural, electronic and optical properties of monolayer MoS_2: a first-principles study. J Alloys Compd 613:122–127. https://doi.org/10.1016/j.jallcom.2014.06.018

Feng L-P, Su J, Chen S, Liu ZT (2014b) First-principles investigations on vacancy formation and electronic structures of monolayer MoS_2. Mater Chem Phys 148:5–9. https://doi.org/10.1016/j.matchemphys.2014.07.026

Gao Y, Zhao H (2016) Chirality and vacancy effect on phonon dispersion of MoS_2 with strain. Phys Lett A 380:745–752. https://doi.org/10.1016/j.physleta.2015.11.029

Gan Y, Zhao H (2014) Chirality effect of mechanical and electronic properties of monolayer MoS_2 with vacancies. Phys Lett A 378:2910–2914. https://doi.org/10.1016/j.physleta.2014.08.008

Gibb AL, Alem N, Chen JH, Erickson KJ, Ciston J, Gautam A, Linck M, Zettl A (2013) Atomic resolution imaging of grain boundary defects in monolayer chemical vapor deposition-grown hexagonal boron nitride. J Am Chem Soc 135:6758–6761. https://doi.org/10.1021/ja400637n

Grantab R, Shenoy VB, Ruoff RS (2010) Anomalous strength characteristics of tilt grain boundaries in graphene. Science 330:946–948. https://doi.org/10.1126/science.1196893

Guo Y, Qiao C, Wang A, Zhang J, Wang S, Su W-S, Jia Y (2016) The fracture behaviors of monolayer phosphorene with grain boundaries under tension: a molecular dynamics study. Phys Chem Chem Phys 18:20562–20570. https://doi.org/10.1039/C6CP03655D

Hao F, Chen X (2016) First-principles study of the defected phosphorene under tensile strain. J Appl Phys 120:165104. https://doi.org/10.1063/1.4966167

Hao F, Fang D, Xu Z (2011) Mechanical and thermal transport properties of graphene with defects. Appl Phys Lett 99:041901. https://doi.org/10.1063/1.3615290

Hong J, Hu Z, Probert M, Li K, Lv D, Yang X, Gu L, Mao N, Feng Q, Xie L, Zhang J, Wu D, Zhang Z, Jin C, Ji W, Zhang X, Yuan J, Zhang Z (2015) Exploring atomic defects in molybdenum disulphide monolayers. Nat Commun 6:6293. https://doi.org/10.1038/ncomms7293

Hu W, Yang J (2015) Defects in phosphorene. J Phys Chem C 119:20474–20480. https://doi.org/
    10.1021/acs.jpcc.5b06077

Huang PY, Ruiz-Vargas CS, van der Zande AM, Whitney WS, Levendorf MP, Kevek JW, Garg
    S, Alden JS, Hustedt CJ, Zhu Y, Park J, McEuen PL, Muller DA (2011) Grains and grain
    boundaries in single-layer graphene atomic patchwork quilts. Nature 469:389–392. https://doi.
    org/10.1038/nature09718

Jhon YI, Zhu SE, Ahn JH, Jhon MS (2012) The mechanical responses of tilted and non-tilted
    grain boundaries in graphene. Carbon 50:3708–3716. https://doi.org/10.1016/j.carbon.2012.03.
    044

Jing N, Xue Q, Ling C, Shan M, Zhang T, Zhou X, Jiao Z (2012) Effect of defects on Young's
    modulus of graphene sheets: a molecular dynamics simulation. RSC Adv 2:9124. https://doi.
    org/10.1039/c2ra21228e

Komsa H-P, Kurasch S, Lehtinen O, Kaiser U, Krasheninnikov AV (2013) From point to extended
    defects in two-dimensional MoS_2: evolution of atomic structure under electron irradiation. Phys
    Rev B 88:35301.https://doi.org/10.1103/PhysRevB.88.035301

Kotakoski J, Meyer JC, Kurasch S, Santos-Cottin D, Kaiser U, Krasheninnikov AV (2011) Stone-
    Wales-type transformations in carbon nanostructures driven by electron irradiation. Phys Rev B
    83:245420. https://doi.org/10.1103/PhysRevB.83.245420

Kurasch S, Kotakoski J, Lehtinen O, Skakalova V, Smet J, Krill CE, Krasheninnikov AV, Kaiser
    U (2012) Atom-by-atom observation of grain boundary migration in graphene. Nano Lett
    12:3168–3173. https://doi.org/10.1021/nl301141g

Le M-Q, Nguyen D-T (2014) Atomistic simulations of pristine and defective hexagonal BN and
    SiC sheets under uniaxial tension. Mater Sci Eng A 615:481–488. https://doi.org/10.1016/j.
    msea.2014.07.109

Lehtinen O, Kurasch S, Krasheninnikov AV, Kaiser U (2013) Atomic scale study of the life cycle
    of a dislocation in graphene from birth to annihilation. Nat Commun 4:2098. https://doi.org/10.
    1038/ncomms3098

Li L, Reich S, Robertson J (2005a) Defect energies of graphite: density-functional calculations.
    Phys Rev B 72:184109. https://doi.org/10.1103/PhysRevB.72.184109

Li ML, Wan YL, Tu LP, Yang YC, Lou J (2016a) The effect of V_{MoS3} point defect on the elastic
    properties of monolayer MoS_2 with REBO potentials. Nanoscale Res Lett 11:1–7. https://doi.
    org/10.1186/s11671-016-1377-x

Li NN, Sha ZD, Pei QX, Zhang YW (2014) Hydrogenated grain boundaries control the strength
    and ductility of polycrystalline graphene. J Phys Chem C 118:13769–13774. https://doi.org/10.
    1021/jp502084f

Li X, Guo P, Cao T-F, Liu H, Lau W, Liu L (2015) Structures, stabilities, and electronic
    properties of defects in monolayer black phosphorus. Sci Rep 5:10848. https://doi.org/10.1038/
    srep10848

Li X, Ma L, Wang D, Zeng XC, Wu X, Yang J (2016b) Point defects in lines in single crystalline
    phosphorene: directional migration and tunable band gaps. Nanoscale 8:17801–17808. https://
    doi.org/10.1039/C6NR05414E

Li Z, Dharap P, Sharma P, Nagarajaiah S, Yakobson BI (2005b) Continuum field model of defect
    formation in carbon nanotubes. J Appl Phys 97:074303. https://doi.org/10.1063/1.1870102

Lin Y-C, Björkman T, Komsa H-P, Teng P-Y, Yeh C-H, Huang F-S, Lin K-H, Jadczak J, Huang
    Y-S, Chiu P-W, Krasheninnikov AV, Suenaga K (2015) Three-fold rotational defects in two-
    dimensional transition metal dichalcogenides. Nat Commun 6:6736. https://doi.org/10.1038/
    ncomms7736

Liu Y, Yakobson BI (2010) Cones, pringles, and grain boundary landscapes in graphene topology.
    Nano Lett 10:2178–2183. https://doi.org/10.1021/nl100988r

Liu Y, Zou X, Yakobson BI (2012) Dislocations and grain boundaries in two-dimensional boron
    nitride. ACS Nano 6:7053–7058. https://doi.org/10.1021/nn302099q

Liu Y, Xu F, Zhang Z, Penev ES, Yakobson BI (2014) Two-dimensional mono-elemental
    semiconductor with electronically inactive defects: the case of phosphorus. Nano Lett 14:6782–
    6786. https://doi.org/10.1021/nl5021393

López-Polín G, Gómez-Navarro C, Parente V, Guinea F, Katsnelson MI, Pérez-Murano F, Gómez-Herrero J (2014) Increasing the elastic modulus of graphene by controlled defect creation. Nat Phys 11:26–31. https://doi.org/10.1038/nphys3183

Ma J, Alfè D, Michaelides A, Wang E (2009) Stone-Wales defects in graphene and other planar s p2 -bonded materials. Phys Rev B 80:033407. https://doi.org/10.1103/PhysRevB.80.033407

Meng L, Jiang J, Wang J, Ding F (2014) Mechanism of metal catalyzed healing of large structural defects in graphene. J Phys Chem C 118:720–724. https://doi.org/10.1021/jp409471a

Mortazavi B, Ahzi S (2013) Thermal conductivity and tensile response of defective graphene: a molecular dynamics study. Carbon 63:460–470. https://doi.org/10.1016/j.carbon.2013.07.017

Mortazavi B, Cuniberti G (2014) Atomistic modeling of mechanical properties of polycrystalline graphene. Nanotechnology 25:215704. https://doi.org/10.1088/0957-4484/25/21/215704

Nascimento AJM, Nunes RW (2013) Lubrication of Stone-Wales transformations in graphene by hydrogen and hydroxyl functional groups. Nanotechnology 24:435707. https://doi.org/10.1088/0957-4484/24/43/435707

Okada S (2009) Atomic configurations and energetics of vacancies in hexagonal boron nitride: first-principles total-energy calculations. Phys Rev B 80:161404. https://doi.org/10.1103/PhysRevB.80.161404

Pakdel A, Zhi C, Bando Y, Golberg D (2012) Low-dimensional boron nitride nanomaterials. Mater Today 15:256–265. https://doi.org/10.1016/S1369-7021(12)70116-5

Rajasekaran G, Narayanan P, Parashar A (2015) Effect of point and line defects on mechanical and thermal properties of graphene: a review. Crit Rev Solid State Mater Sci 8436:1–25. https://doi.org/10.1080/10408436.2015.1068160

Ren Y, Cao G (2016) Effect of geometrical defects on the tensile properties of graphene. Carbon 103:125–133. https://doi.org/10.1016/j.carbon.2016.03.017

Rodrigo L, Pou P, Pérez R (2016) Graphene monovacancies: electronic and mechanical properties from large scale ab initio simulations. Carbon 103:200–208. https://doi.org/10.1016/j.carbon.2016.02.064

Samsonidze GG, Samsonidze GG, Yakobson BI (2002) Energetics of Stone-Wales defects in deformations of monoatomic hexagonal layers. Comput Mater Sci 23:62–72. https://doi.org/10.1016/S0927-0256(01)00220-8

Sensoy MG, Vinichenko D, Chen W, Friend CM, Kaxiras E (2017) Strain effects on the behavior of isolated and paired sulfur vacancy defects in monolayer MoS_2. Phys Rev B 95:014106. https://doi.org/10.1103/PhysRevB.95.014106

Sha ZD, Quek SS, Pei QX, Liu ZS, Wang TJ, Shenoy VB, Zhang YW (2014) Inverse pseudo Hall-Petch relation in polycrystalline graphene. Sci Rep 4:5991. https://doi.org/10.1038/srep05991

Sha Z-D, Pei Q-X, Ding Z, Jiang J-W, Zhang Y-W (2015) Mechanical properties and fracture behavior of single-layer phosphorene at finite temperatures. J Phys D Appl Phys 48:395303. https://doi.org/10.1088/0022-3727/48/39/395303

Sha Z-D, Pei QX, Zhang YY, Zhang YW (2016) Atomic vacancies significantly degrade the mechanical properties of phosphorene. Nanotechnology 27:315704. https://doi.org/10.1088/0957-4484/27/31/315704

Skowron ST, Lebedeva IV, Popov a M, Bichoutskaia E (2015) Energetics of atomic scale structure changes in graphene. Chem Soc Rev 44:3143–3176. https://doi.org/10.1039/c4cs00499j

Song Z, Artyukhov VI, Yakobson BI, Xu Z (2013) Pseudo hall-petch strength reduction in polycrystalline graphene. Nano Lett 13:1829–1833. https://doi.org/10.1021/nl400542n

Sorkin V, Zhang YW (2015) The deformation and failure behaviour of phosphorene nanoribbons under uniaxial tensile strain. 2D Mater 2:035007. https://doi.org/10.1088/2053-1583/2/3/035007

Sorkin V, Zhang YW (2016) Mechanical properties of phosphorene nanotubes: a density functional tight – binding study. Nanotechnology 27:395701. https://doi.org/10.1088/0957-4484/27/39/395701

Sorkin V, Zhang YW (2017) Mechanical properties and failure behavior of phosphorene with grain boundaries. Nanotechnology 28:123–128. https://doi.org/10.1088/1361-6528/aa537b

Sorkin V, Pan H, Shi H, Quek SY, Zhang YW (2014) Nanoscale transition metal dichalcogenides: structures, properties, and applications. Crit Rev Solid State Mater Sci 39:319–367. https://doi.org/10.1080/10408436.2013.863176

Sorkin V, Cai Y, Ong Z, Zhang G, Zhang YW (2016) Recent advances in the study of phosphorene and its nanostructures. Crit Rev Solid State Mater Sci 8436:1–82. https://doi.org/10.1080/10408436.2016.1182469

Stone AJ, Wales DJ (1986) Theoretical studies of icosahedral C_{60} and some related species. Chem Phys Lett 128:501–503. https://doi.org/10.1016/0009-2614(86)80661-3

Ulman K, Narasimhan S (2014) Point defects in twisted bilayer graphene: a density functional theory study. Phys Rev B 89:245429. https://doi.org/10.1103/PhysRevB.89.245429

Wang C, Ding Y (2013) Catalytically healing the Stone-Wales defects in graphene by carbon adatoms. J Mater Chem A 1:1885–1891. https://doi.org/10.1039/C2TA00736C

Wang MC, Yan C, Ma L, Hu N, Chen MW (2012) Effect of defects on fracture strength of graphene sheets. Comput Mater Sci 54:236–239. https://doi.org/10.1016/j.commatsci.2011.10.032

Wang R, Yang J, Wu X, Wang S (2016) Local charge states in hexagonal boron nitride with Stone-Wales defects. Nanoscale 8:8210–8219. https://doi.org/10.1039/C5NR09099G

Wang X, Tabarraei A, Spearot DE (2015) Fracture mechanics of monolayer molybdenum disulfide. Nanotechnology 26:175703. https://doi.org/10.1088/0957-4484/26/17/175703

Warner JH, Margine ER, Mukai M, Robertson AW, Giustino F, Kirkland AI (2012) Dislocation-driven deformations in graphene. Science 337:209–212. https://doi.org/10.1126/science.1217529

Wei Y, Wu J, Yin H, Shi X, Yang R, Dresselhaus M (2012) The nature of strength enhancement and weakening by pentagon–heptagon defects in graphene. Nat Mater 11:759–763. https://doi.org/10.1038/nmat3370

Xiong QL, Li ZH, Tian XG (2015) The defect-induced fracture behaviors of hexagonal boron-nitride monolayer nanosheets under uniaxial tension. J Phys D Appl Phys 48:375502. https://doi.org/10.1088/0022-3727/48/37/375502

Xu L, Wei N, Zheng Y (2013) Mechanical properties of highly defective graphene: from brittle rupture to ductile fracture. Nanotechnology 24:505703. https://doi.org/10.1088/0957-4484/24/50/505703

Yang Z, Huang Y, Ma F, Sun Y, Xu K, Chu PK (2015) Temperature and strain-rate effects on the deformation behaviors of nano-crystalline graphene sheets. Eur Phys J B 88:135. https://doi.org/10.1140/epjb/e2015-50850-x

Yanovsky YG, Nikitina EA, Karnet YN, Nikitin SM (2010) Simulation of deformation and fracture of graphene: effect of size, defects and surface modification. Phys Mesomech 13:329–336. https://doi.org/10.1016/j.physme.2010.11.013

Yazyev OV, Louie SG (2010) Topological defects in graphene: dislocations and grain boundaries. Phys Rev B Phys 81:195420. https://doi.org/10.1103/PhysRevB.81.195420

Yu N, Wang L, Li M, Sun X, Hou T, Li Y (2015a) Molybdenum disulfide as a highly efficient adsorbent for non-polar gases. Phys Chem Chem Phys 17:11700–11704. https://doi.org/10.1039/C5CP00161G

Yu ZG, Zhang YW, Yakobson BI (2015b) An anomalous formation pathway for dislocation-sulfur vacancy complexes in polycrystalline monolayer MoS2. Nano Lett 15:6855–6861. https://doi.org/10.1021/acs.nanolett.5b02769

Zhang J, Zhao J, Lu J (2012) Intrinsic strength and failure behaviors of graphene grain boundaries. ACS Nano 6:2704–2711. https://doi.org/10.1021/nn3001356

Zhou W, Zou X, Najmaei S, Liu Z, Shi Y, Kong J, Lou J, Ajayan PM, Yakobson BI, Idrobo JC (2013) Intrinsic structural defects in monolayer molybdenum disulfide. Nano Lett 13:2615–2622. https://doi.org/10.1021/nl4007479

Zobelli A, Ewels CP, Gloter A, Seifert G, Stephan O, Csillag S, Colliex C (2006) Defective structure of BN nanotubes: from single vacancies to dislocation lines. Nano Lett 6:1955–1960. https://doi.org/10.1021/nl0610811

Zobelli A, Ewels CP, Gloter A, Seifert G (2007) Vacancy migration in hexagonal boron nitride. Phys Rev B 75:094104. https://doi.org/10.1103/PhysRevB.75.094104

Zou X, Liu Y, Yakobson BI (2013) Predicting dislocations and grain boundaries in two-dimensional metal-disulfides from the first principles. Nano Lett 13:253–258. https://doi.org/10.1021/nl3040042

Mechanics and Electromechanics of Two-Dimensional Atomic Membranes

78

Harold S. Park

Contents

Abstract

Two-dimensional (2D) materials have, over the past decade, attracted significant amounts of research interest due to their exceptional and unique physical properties. Here, two areas of graphene mechanics are overviewed where computational techniques, both existing (classical molecular dynamics) and new (electromechanical coupling techniques), have enabled new insights. First, we discuss the prediction of and insights gained with regard to atomistic simulations of auxetic behavior in 2D materials. Second, new computational techniques are discussed that couple molecular dynamics, tight-binding, and quantum transport to examine how mechanical strain can impact, in interesting and often unanticipated fashion, the electronic properties of graphene kirigami.

H. S. Park (✉)
Department of Mechanical Engineering, Boston University, Boston, MA, USA
e-mail: parkhs@bu.edu

1   Introduction

Since the discovery of a facile method for its isolation, graphene (Novoselov et al. 2005), the simplest two-dimensional crystal, has attracted intense attention not only for its unusual physical properties (Geim et al. 2007; Castro Neto et al. 2009; Han et al. 2007; Seol et al. 2010; Akinwande et al. 2017) but also for its potential as the basic building block for a wealth of device applications (Portions of the introduction are reprinted with permission from Qi et al. (2014). Copyright 2014 by the American Physical Society.).

There exist key limitations that appear to restrict the application of graphene for all-carbon electronic circuits: one such limitation is that graphene, in its pristine form, is well known to be a semimetal with no bandgap (Castro Neto et al. 2009). A highly active field of study has recently emerged based on the idea of applying mechanical strain to modify the intrinsic response of electrons to external fields in graphene (Guinea et al. 2010a; Qi et al. 2013; Tomori et al. 2011). This includes the strain-induced generation of spectral (band) gaps and transport gaps, which suppress conduction at small densities. In this context, several groups (Guinea et al. 2008, 2010a, b; Pereira and Castro Neto 2009; Guinea and Low 2010; Abedpour et al. 2011; Kim et al. 2011; Yeh et al. 2011; Yang 2011; Kitt et al. 2012, 2013; Yue et al. 2012) have employed continuum mechanics coupled with effective models of the electronic dynamics to study the generation of pseudomagnetic fields (PMFs) in different graphene geometries and subject to different deformations. The potential impact of strain engineering beyond the generation of bandgaps has also attracted tremendous interest (Pereira and Castro Neto 2009; Pereira et al. 2010; Abanin and Pesin 2012; Wang et al. 2011).

Pereira and Castro Neto (Pereira et al. 2009) showed that a bandgap will not emerge under simple uniaxial strain unless the strain is larger than roughly 20%. This theoretical prediction, based on an effective tight-binding model for the electronic structure, has been subsequently confirmed by various more elaborate ab initio calculations (Ni et al. 2009; Farjam and Tabar 2009; Choi et al. 2010). The robustness of the gapless state arises because simple deformations of the lattice lead only to local changes of the *position* of the Dirac point with respect to the undeformed lattice configuration (Kane et al. 1997; Suzuura and Ando 2002) and to anisotropies in the Fermi surface and Fermi velocity (Pereira et al. 2010). The shift in the position of the Dirac point is captured, in the low-energy, two-valley, Dirac approximation, by a so-called pseudomagnetic vector potential and resulting pseudomagnetic field (PMF) that arises from the strain-induced perturbation of the tight-binding hoppings (Suzuura and Ando 2002). As a result, electrons react to mechanical deformations in a way that is analogous to their behavior under a real external magnetic field, except that overall time-reversal symmetry is preserved, since the PMF has opposite signs in the two time-reversal related valleys (Castro Neto et al. 2009). Guinea et al. (2010a) found that nearly homogeneous PMFs could be generated in graphene through triaxial stretching, but the resulting fields were found to be moderate, unless relatively large (i.e., >10%) tensile strains could be applied. Unfortunately, such large planar tensile strains have not been

experimentally realized in graphene to date. This is arguably attributed to the record-high tensile modulus of graphene and the unavoidable difficulty in effectively transferring the required stresses from substrates to this monolayer crystal (Gong et al. 2010).

An interesting and different direction has recently been pursued by employing lithography to extend the range of elastic deformation that can be sustained by a graphene by patterning nanomeshes out of large graphene sheets (Zhu et al. 2014) (Portions of the introduction are reprinted with permission from Bahamon et al. (2016). Copyright 2016 by the American Physical Society.). Such structures, inspired by the Japanese art of cutting paper called kirigami, have been designed and tested in recent experiments (Blees et al. 2015) that establish their mechanical robustness and extremely high elongation limits ($\sim$240%) compared to uncut graphene. In a previous study of the elastic and mechanical characteristics of graphene kirigami, some of us showed that their stretchability limit and effective Young's modulus can be characterized (and customized) in terms of two geometric parameters (Qi et al. 2014): α, the ratio of the overlapping cut length to the kirigami length, and β, the ratio of the overlapping width to the kirigami length.

The ability currently demonstrated to experimentally design graphene kirigami capable of ultrahigh elastic deformations raises the question of how the electronic states and the flow of current within the kirigami are, or can be, modified under deformation. In particular, since any kirigami always involves a number of bends, indentations, and narrow regions (henceforth, "constrictions"), it can naturally harbor a number of localized states at low energies (below the threshold for electronic transmission) as a result of either the geometry alone (González et al. 2011) or geometry combined with disorder (Ihnatsenka et al. 2009) and/or Coulomb interactions (Todd et al. 2008; Han et al. 2010). The experimental observation of Coulomb blockade in graphene nanoribbons and constrictions (Sols et al. 2007; Todd et al. 2008; Stampfer et al. 2009; Gallagher et al. 2010) indicates that such states are expected to be prevalent at low energies in "papercut" graphene devices and that disorder can promote or further stabilize them.

A segment of suitably patterned graphene hosting such states defines a quantum dot, and its periodic repetition would define an array of coupled quantum dots, analogously to what has been achieved in semiconductor structures through comb-shaped arrays of side gates (Kouwenhoven et al. 1990; Brum 1991; Ji et al. 1992). The fact that graphene kirigami of possibly any desired shape can sustain repeated stretching cycles (Blees et al. 2015) suggests that one might be able to design kirigami suitable for the study of coupled quantum dots, with the advantage that the inter-dot coupling responds directly to the deformed state of the structure. The vast range of stretchability limits and effective Young moduli that these structures can be designed with, combined with the proven mechanical resilience of graphene at the nanoscale, hint at the possibility of designing versatile electromechanical devices based on this concept of stretchable graphene quantum dot arrays whose current response can be strongly sensitive to the inter-dot coupling.

This is the first problem we address in this chapter, through a novel coupling of classical molecular dynamics, tight-binding, and quantum transport. We concentrate

only in the geometric aspects that contribute to localization of electronic states at certain portions of the kirigami. We confirm that such states define effective quantum dots, which result in a characteristic profile of the conductance at low energies defined by a sequence of resonant minibands and stop-gaps. This is analogous to the conductance of periodic split-gate semiconductor heterostructures (Kouwenhoven et al. 1990; Ulloa et al. 1990; Brum 1991), where each miniband consists of a set of resonant tunneling peaks. At low energies, the conductance profile is seen to be very sensitive to the deformation, and we analyze this response in terms of the interplay between variations in inter-dot coupling and strain barriers that develop under stretching.

The second problem we address in this chapter deals with recently discovered auxetic behavior in 2D materials (Jiang et al. 2016) (Portions of the introduction are reprinted with permission from Jiang et al. (2016). Copyright 2016 by the American Physical Society.). The Poisson's ratio, $\nu_{xy} = -\varepsilon_y/\varepsilon_x$, characterizes the resultant strain in the y-direction for a material under longitudinal deformation in the x-direction. The Poisson's ratio is typically a positive number and has a value around 0.3 for many engineering materials (i.e., steels). The value is positive when a material contracts in the transverse directions when stretched uniaxially. The Poisson's ratio in classical elasticity theory depends on the ratio between the bulk modulus and the shear modulus, e.g., $\nu = \frac{1}{2}\left(1 - 1/\left(\frac{K}{\mu} + \frac{1}{3}\right)\right)$ for three-dimensional isotropic materials. The Poisson's ratio is thus limited to the range $-1 < \nu < 0.5$ for three-dimensional isotropic materials within classical elasticity theory.

Materials are thus allowed to exhibit a negative Poisson's ratio (NPR) within classical elasticity theory. Materials with NPR were termed auxetic materials by Evans (1991). One way in which the impact of NPR can be gleaned is to note that there exist certain physical properties that are inversely proportional to $1 + \nu$ or $1 - \nu^2$, which implies that these properties become infinitely large in the limit of the Poisson's ratio $\nu \to -1$. For example, the sound of speed is proportional to $(1 + \nu)^{-1/2}$, and the material hardness is related to $(1 - \nu^2)^\gamma$, with γ as a constant. Hence, materials with NPR typically have novel properties such as enhanced toughness and enhanced sound and vibration absorption.

In 1987, Lakes performed seminal experiments to illustrate the NPR in a foam structure (Lakes 1987). Since then, many researchers have demonstrated that the NPR phenomenon is actually quite common both as an intrinsic material property and also in engineered structures (Rothenburg et al. 1991; Lakes 1993; Baughman and Galvao 1993; Evans and Alderson 2000; Yang et al. 2004; Ravirala et al. 2007; Lethbridge et al. 2010; Bertoldi et al. 2010; Alderson et al. 2012; Clausen et al. 2015). For example, 69% of the cubic elemental metals have intrinsic NPR along a non-axial direction, (Milstein and Huang 1979; Baughman et al. 1998) and the Poisson's ratio for FCC metals can be negative along some principal directions by proper control over the transverse loading (Ho et al. 2016).

Concurrently, nanomaterials, encompassing such well-known materials like buckyballs, carbon nanotubes, graphene, nanowires, black phosphorus, MoS_2, and others, have drawn significant interest within the past two decades. Within the last

3 years, the auxetic property has been found in some of these nanomaterials (Jiang and Park 2014), with the mechanisms underlying the auxetic properties often being due to specific nanoscale physical properties. In the present chapter, we discuss auxetic behavior in two forms of monolayer graphene, graphene nanoribbons (Jiang and Park 2016) and bulk graphene (Jiang et al. 2016), while discussing the distinct mechanics principles enabling the auxetic response in both of those forms.

2 Auxeticity in 2D Nanomaterials

2.1 Single-Layer Graphene

We begin our discussion of negative Poisson's ratio (NPR) in single-layer graphene (Jiang et al. 2016) (Section 2.1 reprinted (adapted) with permission from Jiang et al. (2016). Copyright (2016) American Chemical Society.). The crystal structure for single-layer graphene is shown in the inset of Fig. 1b. Periodic boundary conditions are applied in both x- and y-directions such that our studies,

Fig. 1 (Color online) Negative Poisson's ratio in graphene. (**a**) The resultant strain ε_y versus the applied strain ε_x. A robust valley point exists at $\varepsilon_x = 0.06$ in all curves for varying simulation parameters using both MD and MS approaches. (**b**) The Poisson's ratio extracted from the MS results in (a) through $\nu = -\partial \varepsilon_y / \partial \varepsilon_x$, which is negative for $\varepsilon_x > 0.06$. Image reprinted with permission from Jiang et al. (2016)
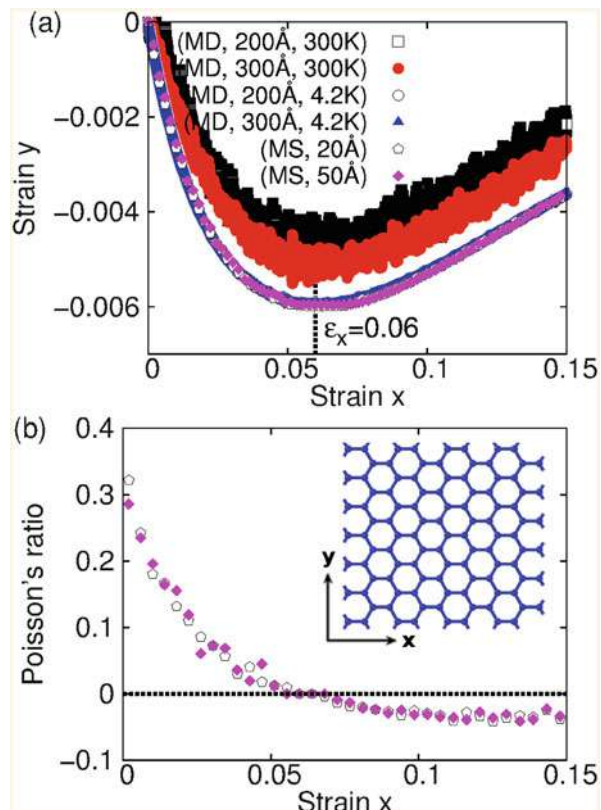
and the properties we report, represent those of bulk graphene without edge effects. The carbon-carbon interactions are described by the Brenner potential (Brenner et al. 2002), which has been widely used to study the mechanical response of graphene (Mo et al. 2009). The structure is stretched in the x (armchair)-direction, while graphene is allowed to be fully relaxed in the y (zigzag)-direction, using both molecular dynamics (MD) and molecular statics (MS) simulations. For the MD simulations, the standard Newton equations of motion are integrated in time using the velocity Verlet algorithm with a time step of 1 fs, which is small enough to maintain energy conservation during the MD simulations. This time step is also small enough to accurately discretize the atomic trajectory corresponding to the highest-frequency vibration modes in graphene (with frequency around (Jiang et al. 2008) 4.8×10^{13} Hz). For the MS simulations, the conjugate gradient algorithm is used for energy minimization. Simulations are performed using the publicly available simulation code LAMMPS (Plimpton 1995), while the OVITO package is used for visualization (Stukowski 2010).

Figure 1a shows the resultant strain ε_y in the y-direction in graphene that is stretched by ε_x in the x-direction. The x-axis is along the horizontal direction, while the y-axis is in the vertical direction as shown in the inset of Fig. 1b. The resultant strain in the y-direction is computed by $\varepsilon_y = \frac{L_y - L_{y0}}{L_{y0}}$ with L_{y0} and L_y as the initial and deformed lengths in the y-direction. We simulate the tensile deformation of graphene using both MD and MS simulations. MD simulations are carried out for square graphene of size $L = 200$ and 300 Å at 4.2 and 300 K. MS simulations are performed for graphene of dimension $L = 20$ and 50 Å.

As shown in Fig. 1a, there is a robust valley point around $\varepsilon_x = 0.06$ (6%) in all of these curves. This valley point indicates that the Poisson's ratio, calculated by (Ho et al. 2014) $\nu = -\partial \varepsilon_y / \partial \varepsilon_x$, is positive for $\varepsilon_x < 0.06$ but becomes negative for $\varepsilon_x > 0.06$. These results demonstrate that the NPR is robust, as it is observed for both low and room temperature conditions, as well as for all structural sizes we have considered. We note that graphene is highly stretchable and has been stretched in a wide strain range experimentally. A strain up to 0.15 has been applied on graphene to measure the nonlinear stiffness (Lee et al. 2008) or manipulate its rippled structure (Bao et al. 2009). Furthermore, recent experiments reported a uniaxial strain up to 0.1, which can be controlled in a reversible and nondestructive manner in graphene (Garza et al. 2014). Thus, the critical strain of $\varepsilon_x = 0.06$ we find has been achieved in contemporary experiments on graphene, so that theoretical results in the present work are experimentally verifiable. The Poisson's ratio is about 0.3 at $\varepsilon_x \approx 0$, which agrees with a recent numerical result with the realistic interatomic potential LCBOPII (Los et al. 2006).

Figure 1a shows that results from MD simulations at 4.2 K coincide with the results from MS simulations, so we will concentrate on the MS simulation results for the rest of this chapter. Figure 1b shows the strain dependence for the Poisson's ratio extracted from these two curves from MS simulations in panel (a). It explicitly shows that the Poisson's ratio is negative for $\varepsilon_x > 0.06$.

We note that there have been previous reports of NPR in graphene. Specifically, large numbers of vacancy defects (Grima et al. 2015) or patterning periodic porous structures (Ho et al. 2016) for bulk graphene or compressive edge stress-induced warping in graphene ribbons are three different mechanisms to achieve the NPR in graphene (Jiang et al. 2016). The Poisson's ratio for graphene can also be driven into the negative regime by thermally induced ripples at high temperatures above 1700 K (Zakharchenko et al. 2009). In contrast, the NPR revealed in the present work represents an intrinsic material property for single-layer graphene.

Discussion To explore the underlying mechanism for the intrinsic NPR, we first illustrate two major deformation modes for the tensile deformation of graphene in Fig. 2. These two deformation modes are fundamental deformation modes corresponding to the bond stretching and angle bending interactions (Chang et al. 2003); i.e., $V_b = \frac{K_b}{2} (b - b_0)^2$ and $V_\theta = \frac{K_\theta}{2} (\theta - \theta_0)^2$, where b is the bond length and θ is the angle in the deformed graphene and $\theta_0 = 120°$ and $b_0 = 1.42\,\text{Å}$ are material constants related to undeformed graphene. K_b is the force constant that characterizes the resistance to bond stretching, and so a larger value of K_b indicates a stiffer bond. K_θ characterizes the resistance to angle bending, and so a larger value of K_θ means a larger resistance to angle bending deformations. The values of these force constants can be obtained by using the value of the Young's modulus and the Poisson's ratio of graphene (Chang et al. 2003). The bond stretching and angle bending are two major interaction terms in graphene,
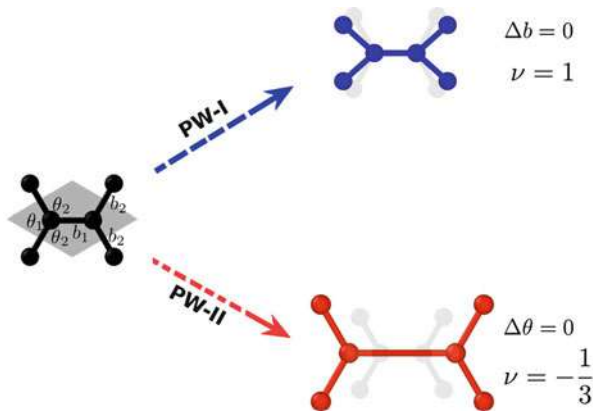
Fig. 2 (Color online) Two typical ideal deformation pathways during the tensile deformation of graphene. The left atom cluster (black online) is stretched along the horizontal x-direction. The parallelogram gray area indicates the unit cell. PW-I (blue online): carbon-carbon bond lengths remain constant ($\Delta b = 0$), while angles are altered to accommodate the external strain, which results in a Poisson's ratio of $\nu = 1$. PW-II (red online): angles are unchanged, and bond lengths are elongated to accommodate the external tension, resulting in a NPR of $\nu = -1/3$. The lighter shades show the undeformed structure. Image reprinted with permission from Jiang et al. (2016)

especially in graphene without out-of-plane deformation. We note that the bond stretching and angle bending interactions were used to derived analytic expressions for the Poisson's ratio in carbon nanotubes (Chang et al. 2003, 2005; Shen et al. 2004; Yao et al. 2008). The analytic expressions illustrate the dependence of the Poisson's ratio on geometrical parameters and force constants. For example, Yao et al. performed a speculative investigation on the possibility of NPR for carbon nanotubes by varying one parameter (or ratio of parameters) while holding all other parameters unchanged (Yao et al. 2008). However, it was determined that significant, and likely nonphysical changes to the geometry or material constants would be necessary for the NPR to appear in CNTs.

The overall deformation process for graphene depends on the competition between the bond stretching and angle bending interactions. For $V_b \gg V_\theta$, the bonds are too stiff to be stretched ($\Delta b \approx 0$), so only the bond angles will change during the tension of graphene. This type of deformation will be referred to as the PW-I deformation mode, as shown in the top (blue online) of Fig. 2. For $V_b \ll V_\theta$, the bond angles cannot be changed ($\Delta \theta \approx 0$), so the bond lengths will be stretched to accommodate the applied tension. This type of deformation will be referred to as the PW-II deformation mode, as shown in the bottom (red online) of Fig. 2.

The Poisson's ratio corresponding to the PW-I and PW-II deformation modes can be derived by simple algebra. The unit cell is displayed by the parallelogram gray area in the left configuration in Fig. 2. The size of the cell in Fig. 2 in the x- and y-directions are

$$L_x = 2 \left(b_1 + b_2 \cos \frac{\theta_1}{2} \right) \tag{1}$$

$$L_y = 2 b_2 \sin \frac{\theta_1}{2}, \tag{2}$$

which yields

$$dL_x = 2 \left(db_1 + db_2 \cos \frac{\theta_1}{2} - \frac{b_2}{2} \sin \frac{\theta_1}{2} d\theta_1 \right) \tag{3}$$

$$dL_y = 2 \left(db_2 \sin \frac{\theta_1}{2} + \frac{b_2}{2} \cos \frac{\theta_1}{2} d\theta_1 \right). \tag{4}$$

As a result, the Poisson's ratio is

$$
\begin{aligned}
\nu &= -\frac{\varepsilon_y}{\varepsilon_x} = -\frac{dL_y/L_y}{dL_x/L_x} \\
&= -\frac{b_1 + b_2 \cos \frac{\theta_1}{2}}{b_2 \sin \frac{\theta_1}{2}} \times \frac{db_2 \sin \frac{\theta_1}{2} + \frac{b_2}{2} \cos \frac{\theta_1}{2} d\theta_1}{db_1 + db_2 \cos \frac{\theta_1}{2} - \frac{b_2}{2} \sin \frac{\theta_1}{2} d\theta_1},
\end{aligned}
\tag{5}
$$

where $b_1 = b_2 = b_0$ and $\theta_1 = \theta_2 = \theta_0$ for undeformed graphene. We note that, for small strains, the definition of the Poisson's ratio in Eq. (5) is consistent with the

numerical formula used in the above to extract the Poisson's ratio in Fig. 1b, because ε_y and ε_x have a linear relationship for small strains.

For the PW-I mode, we have $\Delta b_1 \approx 0$, $\Delta b_2 \approx 0$, and $\Delta \theta_1 \neq 0$. As a result, we obtain the Poisson's ratio for PW-I mode as $\nu = 1$ from Eq. (5). For the PW-II mode, we have $\Delta \theta_1 \approx 0$, and the force equilibrium condition leads to Chang et al. (2003) $db_2 = \frac{1}{2} db_1 \cos \frac{\theta_1}{2}$. Hence, the Poisson's ratio for the PW-II mode is $\nu = -1/3$ from Eq. (5). It is interesting to note that the Poisson's ratio ($\nu = -1/3$) for the PW-II mode coincides with the expectations of the self-consistent screening approximation (Los et al. 2006).

Figure 3a illustrates the interplay between the PW-I and PW-II deformation modes during the tensile deformation of graphene. Specifically, it shows the absolute value of the relative variation of the angles θ_1 and θ_2 and bond lengths b_1 and b_2; we note that the change in angle θ_1 is negative in stretched graphene. For $\varepsilon_x < 0.035$, the variations of angles θ_1 and θ_2 are larger than the variations of bonds b_1 and b_2, respectively, which indicates PW-I to be the dominant deformation mode for graphene subject to small uniaxial tensile strains. For $0.035 < \varepsilon_x < 0.085$, the variation of angle θ_1 becomes less than the variation of bond b_1, while the

Fig. 3 (Color online) Pathway energy criteria for PW-I and PW-II deformation modes. (**a**) The variation of key geometrical parameters (angles θ and bond lengths b) in graphene. The y-axis shows the relative variation, i.e., $\Delta b / b_0$ or $\Delta \theta / \theta_0$. (**b**) Pathway energy curve for PW-I and PW-II deformation modes. The curves show a crossover at $\varepsilon_x = 0.06$, which predicts a transition from PW-I mode (positive Poisson's ratio) to PW-II mode (negative Poisson's ratio) during the tensile deformation of graphene. Left bottom inset (black online) shows the undeformed structure. Top inset (blue online) displays the PW-I deformed structure. Right inset (red online) is the PW-II deformed structure. Image reprinted with permission from Jiang et al. (2016)
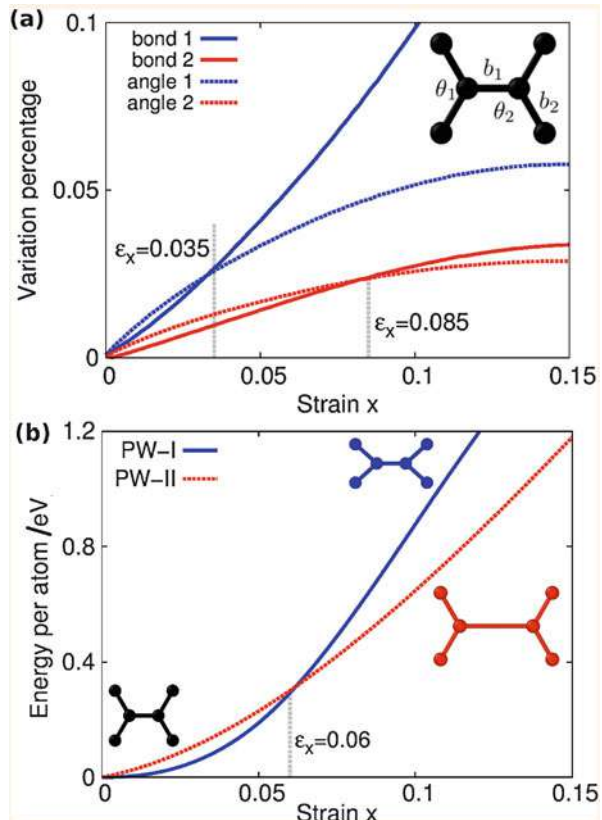
variation of angle θ_2 is still larger than the variation of bond b_2, which implies a competition between the PW-I and PW-II deformation modes for moderate strains. For $\varepsilon_x > 0.085$, variations for both bonds are larger than the variations of angles, so PW-II overcomes PW-I to be the dominant deformation mode for large tensile strains. Hence, the value of the Poisson's ratio will decrease with increasing strain and will become negative at some critical strain between [0.035, 0.085] when PW-II dominates the deformation process. The critical strain of 0.06 for the NPR in Fig. 1 falls in this strain range.

We perform an energy-based analysis, shown in Fig. 3b, of the PW-I and PW-II deformation modes to gain further insight into the interplay between these two deformation modes governing the transition from positive to negative Poisson's ratio at $\varepsilon_x = 0.06$. The energy curve is computed as follows. For PW-I, the structure is manually deformed corresponding to the PW-I mode shown in the top of Fig. 2. We then calculate the potential energy of this deformed structure, which is higher than the potential energy of undeformed graphene. The energy curve shown in Fig. 3b is the difference between the potential energy per atom of the deformed and undeformed graphene structures. The energy curve for PW-II is computed similarly, where angular distortions are allowed while the bond lengths are kept constant. Figure 3b clearly shows a crossover around $\varepsilon_x = 0.06$ between the energy curves for PW-I and PW-II modes.

We thus propose a pathway energy-based criteria: the tensile deformation process for graphene is governed by the deformation mode with lower pathway energy. According to this criteria, PW-I mode will be the major deformation mode for $\varepsilon_x < 0.06$, in which the pathway energy for PW-I mode is lower than the pathway energy for PW-II mode. Similarly, the pathway energy criteria predicts the PW-II to be the major deformation mode for $\varepsilon_x > 0.06$, in which PW-II has lower pathway energy. We showed in Fig. 2 that the PW-I mode has a positive Poisson's ratio, while the PW-II mode has a NPR. As a consequence, the Poisson's ratio is positive for $\varepsilon_x < 0.06$ and will turn negative for $\varepsilon_x > 0.06$. This prediction is in excellent agreement with the numerical results in Fig. 1, where the Poisson's ratio changes from positive to negative at $\varepsilon_x = 0.06$. We note that, to our knowledge, it is the first time the pathway energy criteria is proposed, which is based on the energetic competition for the two major in-plane deformation pathways. This criteria may be useful for future investigations into the mechanical properties of nanomaterials similar as graphene.

2.2   Single-Layer Graphene Ribbons

While the previous section discussed auxetic behavior in bulk, monolayer graphene, we now demonstrate that graphene ribbons can also exhibit NPR (Jiang and Park 2016) (Section 2.2 reprinted (adapted) with permission from Jiang and Park (2014). Copyright (2016) American Chemical Society.). A characteristic feature for free edges in graphene is the warped configuration that is induced by the compressive edge stress as shown in Fig. 4a. The warped structure can be described by the surface

Fig. 4 (Color online) Warped free edges in graphene. Warping surface is described by $z(x, y) = Ae^{-y/l_c}\sin(\pi x/\lambda)$, with $A = 2.26\,\text{Å}$, $l_c = 8.55\,\text{Å}$, and $\lambda = 32.01\,\text{Å}$. (**a**) Top view of graphene with free boundary conditions in the y-direction, while periodic boundary conditions are applied in the x-direction. Color is with respect to the atomic z-coordinate. (**b**) The z-coordinates for atoms in the horizontal red box. (**c**) The z-coordinates for atoms in the vertical blue box. Image reprinted with permission from Jiang and Park (2016)
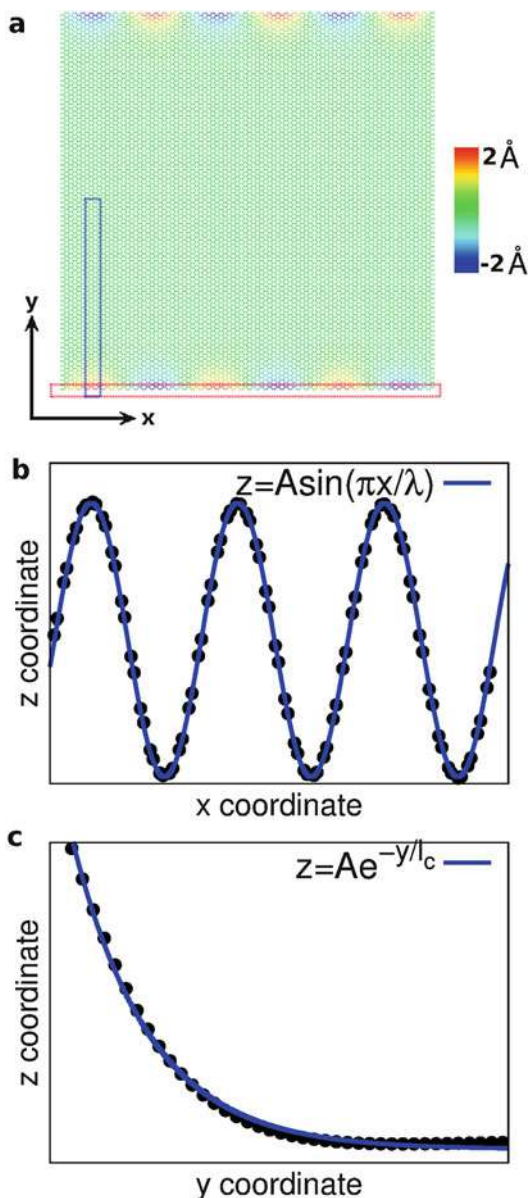
function (Shenoy and Reddy 2008) $z(x, y) = Ae^{-y/l_c}\sin(\pi x/\lambda)$, where $\lambda = L/n$, with L being the length of the graphene ribbon and n being the warping number. The graphene ribbon shown in Fig. 4a has dimensions of $195.96 \times 199.22\,\text{Å}$, resulting in the fitting parameters for the warped free edges as the warping amplitude $A = 2.26\,\text{Å}$, penetration depth $l_c = 8.55\,\text{Å}$, and half wave length $\lambda = 32.01\,\text{Å}$. We note that λ is about one sixth of the length L, i.e., $\lambda = L/6$.

We study five sets of graphene structures with free boundary conditions in the y-direction. Set I: graphene is 195.96 Å in length, and the warped edge has a warping number $n = 6$. Set II: graphene is 195.96 Å in length, and the warped edge has a warping number $n = 8$. Set III: graphene is 195.96 Å in length, and the warped edge has a warping number $n = 10$. Set IV: graphene is 195.96 Å in length, and the warped edge has a warping number $n = 12$. Set V: graphene is 97.98 Å in length, and the warped edge has a warping number $n = 2$. For each simulation set, we consider eight different widths of 29.51, 39.35, 49.19, 59.03, 78.70, 98.38, 147.57, and 199.22 Å. We will demonstrate that the NPR phenomena we report is robust and is observed for different warping periodicities.

Figure 5a shows the strain dependence for the Poisson's ratio of Set I, where the width of the graphene ribbon increases for data from the bottom to the top in the figure. The occurrence of a NPR is clearly observed for small strains and for the narrower width ribbons. Furthermore, the Poisson's ratio changes from negative to positive at some critical strain ε_c. This critical strain is more clearly illustrated in the inset, which shows a critical strain of $\varepsilon_c = 0.005$ in the ε_y-ε_x relation for graphene with width 29.51 Å. For $\varepsilon_x < \varepsilon_c$, graphene expands in the y-direction when it is stretched in the x-direction; i.e., the NPR phenomenon occurs.

The critical strain ε_c represents a structural transition for the warped edge. To reveal this structural transition, we show in Fig. 5b the z position of two atoms from different warped edge regions. One atom is at the peak of the warped edge (shown by the red arrow in the inset), while the other atom is at the valley of the warped edge (shown by blue arrow). Figure 5b clearly shows that both atoms fall into the xy plane at the critical strain $\varepsilon_x = \varepsilon_c$. In other words, the warped edge transitions at the critical point from a three-dimensional, out-of-plane warping configuration into a two-dimensional planar configuration due to the externally applied tensile strain. The connection of the critical strain in the disappearance of the NPR in Fig. 5a and the transition to the two-dimensional planar configuration in Fig. 5b implies that the NPR is connected to the flattening of the warped edges, as the Poisson's ratio becomes positive after the structural transition of the warped edge. We note that the z-coordinates of the atoms in the warped edges in Fig. 5b can be well fitted to the functions $z = \pm b_0 \sin(\theta_0(1 - \varepsilon/\varepsilon_c))$ for $\varepsilon < \varepsilon_c$.

The width dependence for the Poisson's ratio is displayed in Fig. 5c. The Poisson's ratio is strain dependent as shown in Fig. 5a, so we compute an averaged Poisson's ratio using data in the strain range $[0, \varepsilon_c]$, which is equivalent to extracting the Poisson's ratio value by a linear fitting for the ε_y-ε_x relation in $[0, \varepsilon_c]$. Figure 5c shows this averaged Poisson's ratio value for graphene with different widths. We note that the critical strain varies for graphene with different width as indicated by the inset in Fig. 5c, where the critical strain is fitted to the function $\varepsilon_c = 0.0082 - 0.092/W$. The critical strain is smaller in narrower graphene, because some interactions occur between the free ($\pm$y) edges for narrower widths. These edge interactions enable the tension-induced structural transition of the warped edges for narrower ribbons to occur at lower values of tensile strains, because the warping directions for the free ($\pm$y) edges are different, forming a seesaw-like configuration. The saturation value $\varepsilon_c = 0.0082$ at $W \to \infty$ can

Fig. 5 (Color online) Poisson's ratio for graphene from Set I. (**a**) Strain dependence for Poisson's ratio. Inset displays the strain-strain relation for graphene with width 29.51 Å, in which graphene expands in the y-direction when it is stretched in the x-direction by strain smaller than $\varepsilon_c = 0.005$, indicating the NPR effect. (**b**) The strain dependence for the z positions of two atoms in the peak (valley) of the warped edge in graphene. Data are fitted to functions $z = \pm b_0 \sin(\theta_0(1 - \varepsilon/\varepsilon_c))$ with constraint $b_0 = z_0/\sin\theta_0$. Inset displays these two atoms; i.e., the atom on the peak (red arrow) and valley (blue arrow) of the warped edge. Both atoms fall into the graphene plane for strain larger than ε_c. (**c**) Width dependence for the Poisson's ratio. Inset shows the critical strain in graphene ribbons of different width. Image reprinted with permission from Jiang and Park (2016)
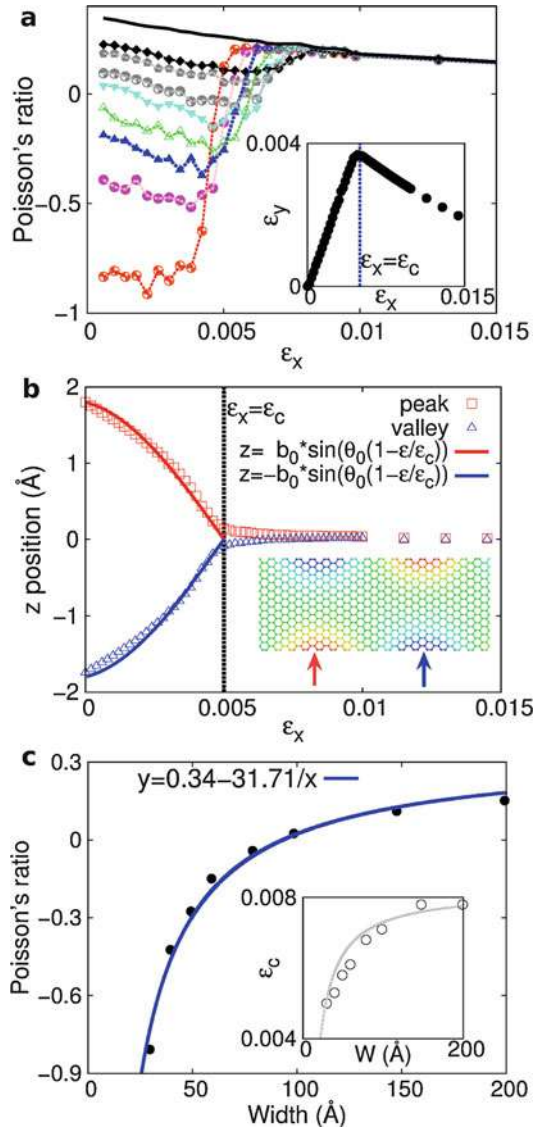
be regarded as the actual value of the critical strain for an isolated warped free edge.

The Poisson's ratio in Fig. 5c increases with increasing width (W) and can be fitted to the function $\nu = 0.34 - 31.71/W$. According to this result, graphene can be regarded as the integration of one central region (with bulk Poisson's ratio ν_0) and two edge regions (with Poisson's ratio ν_e). The size of each edge region is l_c, which is the penetration depth of the warped configuration in Fig. 4. The size of the remaining central region is $W - 2l_c$. Simple algebra gives the effective Poisson's

ratio for the graphene ribbon as

$$\nu = \nu_0 - \frac{2l_c}{W}(\nu_0 - \nu_e).$$ (6)

Comparing Eq. (6) with the fitting function in Fig. 5c, we get $\nu_0 = 0.34$ and $\nu_e = -1.51$.

   This result shows that for the Set I ribbon geometries, in the limit of an ultra-narrow, edge-dominated graphene ribbon, the value of the NPR can be as large as -1.51. Perhaps more importantly, according to Eq. (6), the NPR phenomenon can be observed in graphene sheets with widths up to about 10 nm. Such width nanoribbons are not small and are regularly studied experimentally (Jiao et al. 2009). Furthermore, Fig. 5a shows that the NPR phenomenon is most significant for tensile strains smaller than about 0.5%. These strain values are important as they fall within the strain range of [0, 0.8%] that has already been achievable in many experimental strain engineering investigations (Ni et al. 2008; Mohiuddin et al. 2009). We thus expect that this NPR phenomenon can readily be observed experimentally in the near future.

Discussion   We have demonstrated in the above discussion the connection between the NPR phenomenon and the warped free edges in graphene. We now present an analytic model to describe the relationship between the NPR phenomenon and the warped free edges. In Fig. 6a, the warped edge is represented by an inclined plate (IP) (gray area). During the tensile deformation, the IP falls into the xy plane. The side view in the dashed ellipse illustrates the mechanism enabling the NPR clearly. Specifically, it shows that the projection (b_y) of the IP on the y-axis increases during the falling down of the IP, resulting in the NPR phenomenon.

   This IP model is inspired by the strain-dependent z (out-of-plane)-coordinates of atoms in the warped edges shown in Fig. 5b, where the z-coordinates are fitted to the functions $z = \pm b_0 \sin(\theta_0(1 - \varepsilon/\varepsilon_c))$, in which the parameters are restricted by $b_0 = z_0/\sin\theta_0$. This function describes exactly the trajectory of the tip of the IP (displayed by blue arrow in Fig. 6a) during its falling down process in Fig. 5b. This function also indicates that the IP's tilting angle θ is a linear function of the applied tensile strain ε:

$$\theta = \theta_0\left(1 - \frac{\varepsilon}{\varepsilon_c}\right),$$ (7)

where θ_0 is the initial tilting angle. This expression gives $\theta = 0$ at the critical strain $\varepsilon = \varepsilon_c$, as required by the definition of the critical strain in Fig. 5b. For the applied tensile strain ε in the x-direction, the resulting strain in the y-direction is

$$\varepsilon_y = \frac{\cos\theta - \cos\theta_0}{\cos\theta_0} \approx \frac{\theta_0^2}{\varepsilon_c}\varepsilon,$$
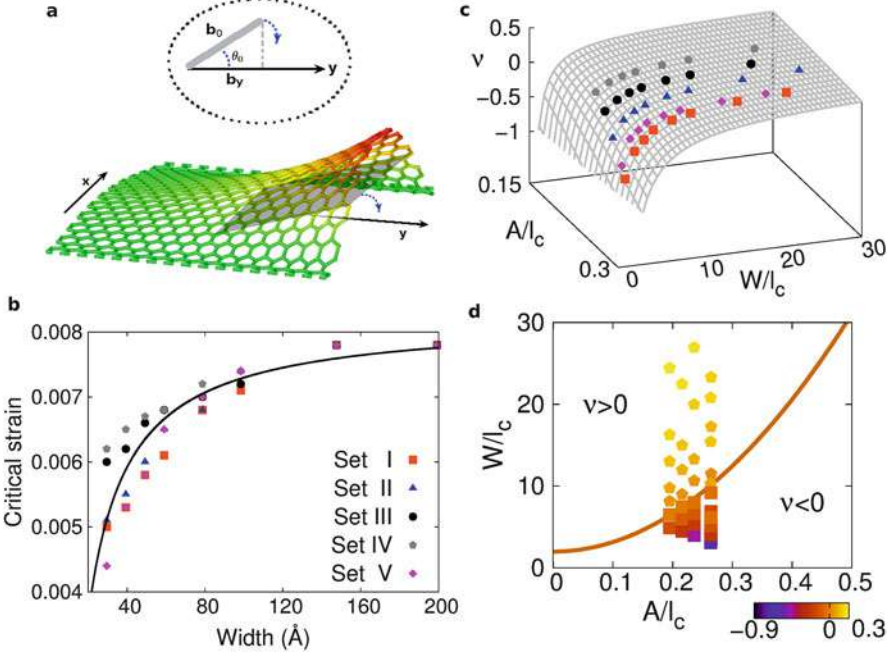
Fig. 6 (Color online) IP model for warped edge induced NPR. (**a**) The warped free edge is represented by the IP (in gray). During the tensile deformation of graphene, the IP falls down, which leads to the increase of its projection along the y-direction, resulting in the NPR effect. (**b**) Critical strain versus width for all of the five simulation Sets. The solid line is the fitting function $\varepsilon_c = 0.0082 - 0.092/W$. (**c**) Comparison of analytic IP model for Poisson's ratio as function of graphene width $\tilde{W} = W/l_c$ and warping amplitude $\tilde{A} = A/l_c$ according to Eq. (10) for simulation sets I, II, III, IV, and V. (**d**) Phase diagram for positive or negative Poisson's ratio in the $\tilde{W}$ and $\tilde{A}$ parameter space, according to Eq. (11), using all data from all of the five simulation Sets. Color bar is for the value of the Poisson's ratio. Image reprinted with permission from Jiang and Park (2016)

yielding the Poisson's ratio of the edge

$$\nu_e = -\frac{\varepsilon_y}{\varepsilon_x} = -\frac{\theta_0^2}{\varepsilon_c}. \qquad (8)$$

We now determine the initial tilting angle θ_0 for the IP. The tilting angle with respect to the y-axis for the tangent plane at point (x, y, z) on the warped surface is

$$\phi(x, y) = \tan^{-1}\left(\frac{\partial w}{\partial y}\right) \approx \frac{A}{l_c} e^{-y/l_c} \sin\frac{\pi x}{\lambda},$$

in which the tilting angle is assumed to be small. This assumption is reasonable as will be shown below. The average tilting angle for the warping area $x \in [0, \lambda]$ and $y \in [0, l_c]$ is

$$\bar{\phi} = \frac{1}{\lambda l_c} \int_0^\lambda dx \int_0^{l_c} dy \phi(x, y) = \frac{A}{l_c} \frac{2}{\pi} \left(1 - \frac{1}{e}\right). \tag{9}$$

Inserting the value of A and l_c from Fig. 4, we get an average tilting angle $\bar{\phi} = 0.106$. We use this average tilting angle as the initial tilting angle for the IP, i.e., $\theta_0 = \bar{\phi} = 0.106$. As a result, we obtain the Poisson's ratio for the warped edge:

$$\nu_e = -\frac{\theta_0^2}{\varepsilon_c} = -\left(\frac{A}{l_c} \frac{2}{\pi} \left(1 - \frac{1}{e}\right)\right)^2 / \varepsilon_c = -1.37$$

in which the critical strain $\varepsilon_c = 0.0082$ is the saturation value from the inset of Fig. 5c, which should be used here in the discussion of the Poisson's ratio for an isolated warped edge. The Poisson's ratio of -1.37 for the Set I geometries that is obtained in the limit of an ultra-narrow graphene ribbon using the IP model is very close to the value of -1.51 obtained via the molecular statics calculations in Fig. 5c.

We find in Fig. 6b that the critical strains for Sets I–V obey the same relation $\varepsilon_c = 0.0082 - 0.092/W$, where again we use the saturation value of $\varepsilon_c = 0.0082$ for the calculation of the effective Poisson's ratio of graphene ribbons with warped edges. The critical strain is directly related to the compressive edge stress, which generates compressive strain (ε_e) in the edge region and thus various warped configurations (with different local minimum potential energies). From Eq. (7), the magnitude of the compressive edge strains equals the critical strain ε_c when the warped edges have been completely flattened into a planar structure, i.e., $\varepsilon_e = \varepsilon_c = 0.0082$. Using the calculated values for the Young's modulus of the edge region as $E_e = 11.158 \text{ eVÅ}^{-2}$, while the compressive edge stress density is Shenoy and Reddy (2008) $\sigma_e = 1.05 \text{ eVÅ}^{-1}$, we can estimate the width (W_e) of the region that will be compressed by the edge strain according to $W_e = \sigma_e/(E_e \varepsilon_e) = 11.48 \text{ Å}$. Taking a representative value of the penetration depth $l_c = 8.55 \text{ Å}$ from simulation Set I, we find that the penetration depth of the resultant warped configuration is close to W_e. These analysis illustrate that the saturation value of the critical strain is an intrinsic property for the free edge, which should not depend on the detailed warping configuration.

From Eqs. (6), (8), and (9), we can thus obtain the following general formula for the Poisson's ratio in graphene ribbons of arbitrary width:

$$\nu = \nu_0 - \frac{2}{\tilde{W}} \left(\nu_0 + \frac{1}{\varepsilon_c} \tilde{A}^2 C_0^2\right), \tag{10}$$

where $C_0 = \frac{2}{\pi} \left(1 - \frac{1}{e}\right)$ is a universal constant and $\nu_0 = 0.34$ is the Poisson's ratio for bulk graphene. The dimensionless quantity $\tilde{W} = W/l_c$ is the width with reference to the penetration depth l_c. This is a bulk-related quantity, and a larger $\tilde{W}$ correlates with moving the Poisson's ratio in the positive direction. The other dimensionless quantity $\tilde{A} = A/l_c$ is the warping amplitude with reference to

the penetration depth. This quantity is an edge-related property, which tunes the Poisson's ratio in the negative direction. The sign and also the value of the Poisson's ratio are determined by the competition between these two effects, while the length of graphene has no effect. Figure 6c shows a three-dimensional plot for the Poisson's ratio predicted by the IP model based on Eq. (10). The numerical results for all of the five simulation Sets are also shown in the figure, which agree quite well with the IP model.

We note that the effective Poisson's ratio defined in Eq. (10) is intrinsically width-dependent, due to the width dependence of the warping amplitude A and the penetration depth l_c. Thus, by using the dimensionless quantities $\tilde{W}$ and $\tilde{A}$, the effect of the warping number is intrinsically included in the expression for the effective Poisson's ratio. Furthermore, the resultant Eq. (10) is a general expression for the Poisson's ratio, which is an explicit function of the ribbon geometry. Hence, Eq. (10) could be readily extended to describe the Poisson's ratio in other similar atomic-thick materials.

By equating Eq. (10) to be zero, we get

$$\tilde{W} = 2 + \frac{2}{v_0} \frac{C_0^2}{\varepsilon_c} \tilde{A}^2. \tag{11}$$

This function is plotted in Fig. 6d and serves to delineate the positive and negative Poisson's ratio regions. This figure serves as a phase diagram for NPR phenomenon in the parameter space of $\tilde{A}$ and $\tilde{W}$. The NPR phenomenon occurs for graphene with parameters in the region below the curve. In particular, if the width $\tilde{W} < 2$, then the Poisson's ratio is negative irrespective of the value for the other parameter $\tilde{A}$. The numerical results for all of the five simulation Sets can be correctly categorized into positive or negative regions in this phase diagram, further validating the analytical IP model we have presented.

3   Computational Electromechanical Coupling for Graphene Kirigami

Having discussed how MD simulations can be used to predict unique mechanical behavior in 2D materials, we now focus on the insights that can be gained through unique computational electromechanical coupling approaches (Portions of Section 3, including Figures 7–16, are reprinted with permission from Bahamon et al. (2016). Copyright 2016 by the American Physical Society.). We illustrate this through the study of graphene kirigami. Specifically, we combine information of the local atomic displacements obtained from MD simulations of deformed kirigami with quantum transport calculations to assess the conductance characteristics at different stages of deformation (Bahamon et al. 2015). At the structural level, the initial stage (stage 1 in Fig. 7) of the end-to-end longitudinal deformation is characterized by a very small effective Young's modulus and an essentially negligible average stretching of the interatomic bonds because, in this early stage, elongation occurs
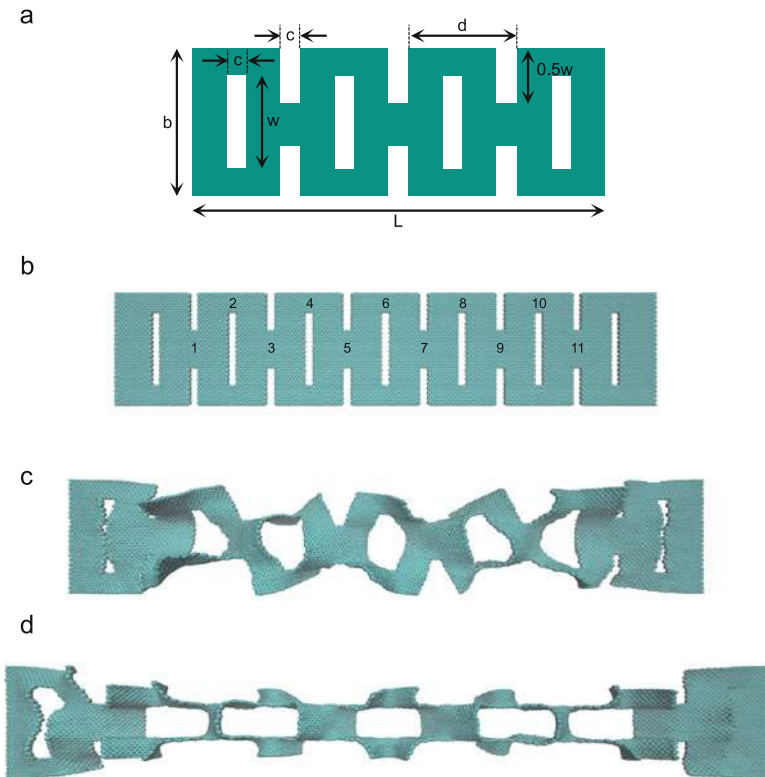
Fig. 7 (**a**) Pattern schematics of the graphene kirigami indicating the most relevant geometric parameters. Snapshots of the actual kirigami used in the calculations for deformations of (**b**) 0%, (**c**) 15.5%, and (**d**) 34.7%. The figures were generated by VMD (Humphrey et al. 1996). Image reprinted with permission from Bahamon et al. (2016)

mostly through bending and twisting of the structure in three-dimensional space (Qi et al. 2014). Despite the low in-plane stretching that occurs during this stage, the conductance and the current are significantly reduced, and the resonant features disappear. However, further elongation into a second stage of deformation where there is substantial bond stretching leads to the revival of the resonant features and the restoration of the magnitudes of conductance to the same levels observed in the undistorted kirigami. These changes are consistently understood as a consequence of transport at low energies being dominated by resonant tunneling between states localized at specific portions of each periodic unit of the kirigami: elongation during stage 1 of deformation perturbs the coupling (overlap) between these states and considerably degrades the conductance as a result of hopping disorder that weakens the super-periodicity of the kirigami structure; further stretching into stage 2 results in strong and localized strain barriers that re-establish the super-periodicity, favor electron localization, and hence revive resonant transmission. The

terminology "strain barrier" is employed here deliberately to emphasize the direct impact that deformation hot-spots have in the electronic behavior of the system at high elongations: as discussed below, the development of localized and periodic regions with extreme bond stretching reinforces the geometry-induced confinement of electronic states within each repeating unit of the kirigami.

3.1   Methodology

Our representative kirigami is obtained by cutting out a graphene nanoribbon of width b according to the pattern shown in Fig. 7a. The rectangular interior holes are defined by the height w and width c, while the connecting necks have length c and a cut starting from the outer edge of length 0.5 w. The longitudinal period is d, and L defines the total length of the system. We shall concentrate our discussion in the kirigami shown schematically in Fig. 7b: our actual system contains 11408 atoms with the geometrical parameters $b \approx 10$ nm, $w \approx 6.8$ nm, $c \approx 0.7$ nm, $d = 4.8$ nm, and $L \approx 34$ nm. For definiteness, we base our discussion on this specific structure where the graphene lattice is oriented so that all horizontal edges are zigzag edges. We also do not include any disorder or edge roughness at this stage. The results, however, are general and should hold when these two restrictions are relaxed because *the only key physical ingredient is the existence of localized states defining a local quantum dot* at specific constrictions or bends, and these can be stabilized by different geometries, with or without disorder (Ihnatsenka et al. 2009). For illustration, we show that explicitly by analyzing a different lattice orientation in the discussion session.

   This study is divided in three stages that aim to determine the electronic transport properties of the kirigami under longitudinal tension. The deformed structures are first obtained with molecular dynamics (MD) simulations of the finite kirigami. We used the Sandia open-source code LAMMPS (Lammps 2018; Plimpton 1995) with the AIREBO (Zhao and Aluru 2010; Stuart et al. 2000; Qi et al. 2010) potential to describe the C-C interactions: the cutoff radii are 2 Å for the REBO term and 6.8 Å for the Lennard-Jones term in the AIREBO potential. The explicit position of each carbon atom in the deformed structure is then used to build a π-band tight-binding Hamiltonian, $H = \sum_{<i,j>} t_{ij}(c_i^\dagger c_j + c_j^\dagger c_i)$, for the distorted kirigami. In this Hamiltonian c_i denotes the annihilation operator on site i, and t_{ij} represents the hopping amplitude between nearest neighbor sites i and j. The stretching, the compression, and the rotation of the C-C bonds created by the tensile load are taken into account in this effective Hamiltonian by means of a position-dependent hopping parameter that reflects the overlap between two arbitrarily oriented p_z orbitals (Bahamon et al. 2015):

$$t_{ij}(d) = V_{pp\pi}(d_{ij})\,\hat{n}_i \cdot \hat{n}_j$$
$$+ \left[V_{pp\sigma}(d_{ij}) - V_{pp\pi}(d_{ij}) \right] \frac{(\hat{n}_i \cdot \mathbf{d}_{ij})(\hat{n}_j \cdot \mathbf{d}_{ij})}{d_{ij}^2}. \tag{12}$$

Here, $\hat{n}_i$ is the unit normal vector to the surface at site i, $\mathbf{d}_{ij}$ is the vector connecting sites i and j, and $V_{pp\sigma}(d)$ and $V_{pp\pi}(d)$ are the Slater-Koster integrals. The presence of the local normals accounts for the relative rotation of neighboring orbitals, whereas the effect of bond stretching is captured by the distance dependence of the Slater-Koster parameters, which we assumed to vary as (Pereira et al. 2009; Qi et al. 2014)

$$V_{pp\pi}(d_{ij}) = t_0\, e^{-\beta(d_{ij}/a-1)}, \tag{13}$$

$$V_{pp\sigma}(d_{ij}) = 1.7\, V_{pp\pi}(d_{ij}), \tag{14}$$

where $t_0 = 2.7\,\text{eV}$, $a \simeq 1.42\,\text{Å}$ is the equilibrium C-C bond length in graphene, and $\beta = 3.37$ captures the exponential decrease in the hopping with interatomic distance. All these quantities are easily calculated using the atomic positions provided by the MD simulations.

Subsequently, in order to inject charge into the system, we couple two undeformed semi-infinite graphene nanoribbons to the left and right edges of the kirigami, the "contacts" being of the same width as the kirigami. To guarantee that the properties observed are those of the kirigami and not of the deformed contact or the contact-kirigami interface, we keep the contacts undeformed during all stages of deformation. In terms of Green's functions, the conductance in the Landauer-Büttiker formalism can be written as Caroli et al. (1971), Datta (1995), Haug et al. (2008)

$$G = \frac{2e^2}{h} T(E) = \frac{2e^2}{h} \text{Tr}[\Gamma_R G^r \Gamma_L G^a], \tag{15}$$

where $G^r = [G^a]^\dagger = [E + i\eta - H - \Sigma_L - \Sigma_R]^{-1}$ is the retarded (advanced) Green's function, the coupling between the contacts and the central region is represented by $\Gamma_{L(R)} = i[\Sigma_{L(R)} - \Sigma_{L(R)}^\dagger]$, and $\Sigma_{L(R)}$ is the self-energy of left(right) contact. To correlate the conductance features with the real-space distribution of the electronic states, we map the local density of states (LDOS) at a given site i directly from the local Green's function according to the identity $\rho_{ii} = -\text{Im}[G^r(\mathbf{r}_i, \mathbf{r}_i, E)]/\pi$. The $I - V$ curves of the device, where I is the total current as a function of the applied bias voltage V, are calculated from the transmission function $T(E, V)$ as (Datta 1995)

$$I(V) = \frac{2e}{h} \int_{-\infty}^{+\infty} T(E, V)\,[f_L(E) - f_R(E)]\,dE, \tag{16}$$

where $f_{L(R)}(E)$ is the Fermi distribution of the left (right) contact.

3.2   Conductance Under Deformation

For a representative illustration showing the behavior in the two stages of deformation, we chose the kirigami structure shown in Fig. 7b that has a small number of cuts per repeating unit and a very high stretchability, up to 65 %. In order to facilitate the presence of localized states within each segment of the kirigami even in the absence of any deformation, we chose to orient the underlying graphene lattice so that the longitudinal cuts are along a zigzag direction, as the internal mini-zigzag edges so defined are expected to naturally support localized states (Fujita et al. 1996) (as pointed out above, this is not a limitation).

   A kirigami such as this one has four deformation stages (Qi et al. 2014): (i) elongation with bending and twisting, but very little in-plane stress, (ii) elongation with stress, (iii) yielding, and (iv) fracture. We restrict our analysis to the first two, where deformations are elastic and reversible and whose stress-strain characteristic is shown in Fig. 8. During the first stage (in this particular structure that corresponds to total deformations below $\approx$20%), horizontal and vertical segments twist and rotate, and, as a result of this excursion of the graphene sheet into the third dimension, the kirigami elongates without significant modification of the average C-C bond length, except for very localized strain hot-spots at the corners of the connecting elements (Qi et al. 2014). A representative sample of a kirigami in this stage is shown in Fig. 7c for a total deformation of 15.5%. With further increase in the tensile load, the kirigami is not capable of accommodating higher elongations
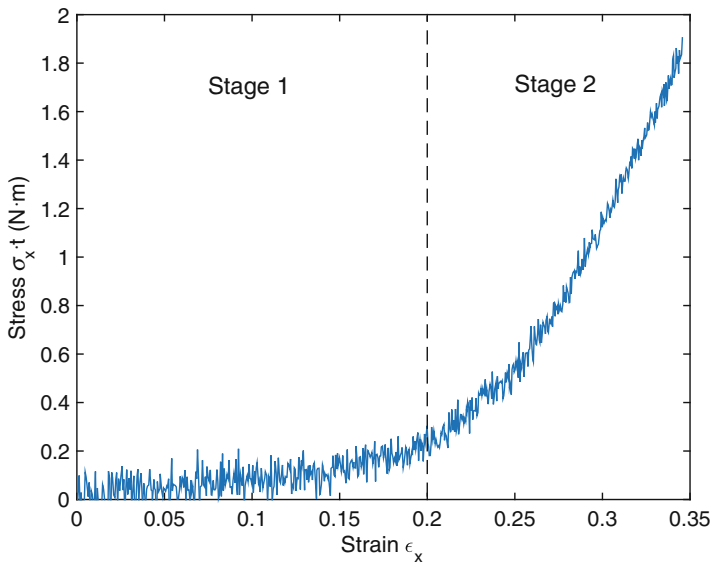
Fig. 8 Stress-strain curve for the kirigami, showing the first two stages corresponding to the elastic and reversible deformations. Image reprinted with permission from Bahamon et al. (2016)

only by twisting, and this triggers the onset of stage 2 (here in the range 20–40%), where further deformation occurs through stretching of the carbon bonds. One important consequence of the existence of these two regimes is that the effective Young's modulus is much lower in stage 1, $E \simeq 0.69$ N/m, (where it is essentially determined by the very small bending stiffness of monolayer graphene) than in stage 2 $E \simeq 15.08$ N/m. This makes mechanical manipulation of kirigami structures experimentally possible and easy in stage 1 (Blees et al. 2015).

We followed the evolution of conductance at low energies for different strain values within the reversible and elastic region and summarize the results in Fig. 9a
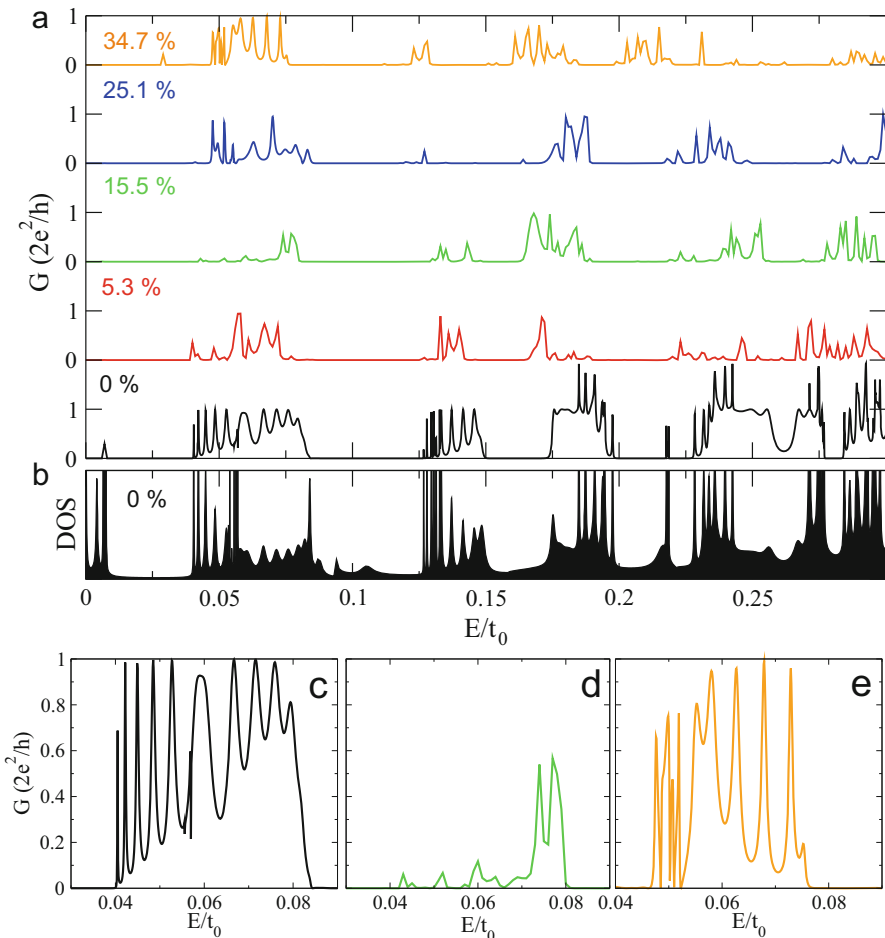
Fig. 9 (**a**) Conductance of the kirigami for different values of deformation. (**b**) DOS of the undeformed structure. The lower row of panels contains a magnification of the energy range comprising the first group of resonances ($0.03 < E/t_0 < 0.09$) for the undeformed kirigami (**c**), kirigami deformed by 15.5% (**d**), and deformed by 34.7% (**e**). Image reprinted with permission from Bahamon et al. (2016)

for deformations of 0%, 5.3%, 15.5%, 25.1%, and 34.7%. The conductance profile of the reference (undeformed) structure displays resonant transmission within well-defined bands (minibands) of width $\approx 0.05\, t_0$. We can see in the DOS plotted in Fig. 9b for the undeformed structure that each miniband arises from the clustering of a finite number of states. At those energies transmission occurs through resonant tunneling and is entirely suppressed otherwise, strikingly different from the conductance profile expected for a graphene nanoribbon (Ulloa et al. 1990; González et al. 2011). In an infinite kirigami, the stop-gaps that separate the minibands arise from the folding of the 1D Brillouin zone of the ideal (pre-cut) graphene ribbon as a result of the new superlattice defined by the kirigami, with the expected opening of spectral gaps at the edge of the reduced zone. In our finite kirigami, Fig. 9b shows that the DOS is likewise strongly suppressed at the miniband edges but does not reach zero in the stop-gaps. This contrasts with the sharp changes in the conductance (Fig. 9a): finite (resonant) within a miniband and clearly zero in the stop-gaps. This is a clear sign that states contributing to the DOS in the stop-gap regions are spatially localized and unable to hybridize to form a tunneling pathway that spans the entire length of the system. Therefore, the robust stop-gap structure in the transmission is not a result of spectral characteristics of the kirigami alone.

To be specific, let us analyze the closeup plot shown in Fig. 9c that captures the energy interval $0.03 < E/t_0 < 0.09$. A mapping of the LDOS at any of the resonances shown in this panel confirms that the electronic density associated with a resonance is strongly localized around the internal longitudinal strips. Figure 10a shows a representative example of the "LDOS hot-spots" (in red) appearing at those regions and a good overlap/coupling with the external contacts. Another snapshot of the LDOS at a different transmission resonance is shown in Fig. 11. The existence of the minibands of resonant transmission is therefore due to the presence of localized states within each unit of the kirigami, which are localized both transversely and longitudinally by the combined effect of the internal zigzag edges and the finite extent of each segment of the kirigami (González et al. 2011; Zhang et al. 2008; Wu et al. 2010; Faria et al. 2015; Kouwenhoven et al. 1990; Ulloa et al. 1990; Brum 1991). At the lowest energies, transmission across the entire system is assisted by tunneling through these localized states that, hence, play a role similar to that of an array of coupled quantum dots that can enable resonant transmission across the entire system at (and only at) well-defined energies. The width of each miniband is determined by the overlap, γ, between localized states, and the number of resonant peaks in a miniband, N, counts the number of isolated quantum dots (Ulloa et al. 1990; Ji et al. 1992). Our geometry contains the 11 internal channels labeled 1–11 in Fig. 7b that define 11 effective quantum dots. Correspondingly, each transmission miniband has a fine structure consisting of 11 peaks as can be seen in Fig. 9c (the outermost two directly in contact with the metallic leads do not define quantum dots because backscattered electrons at these sections are completely absorbed by the contacts).

A quantitative estimate of the inter-dot overlap can be obtained assuming a one-dimensional tight-binding model, according to which the energy levels within one miniband should appear at positions

Fig. 10 Normalized LDOS for the kirigami (**a**) undeformed at $E = 0.067t_0$, (**b**) deformed by 15.5% at $E = 0.074t_0$, and (**c**) deformed by 34.7% at $E = 0.073t_0$. Image reprinted with permission from Bahamon et al. ([2016](#))
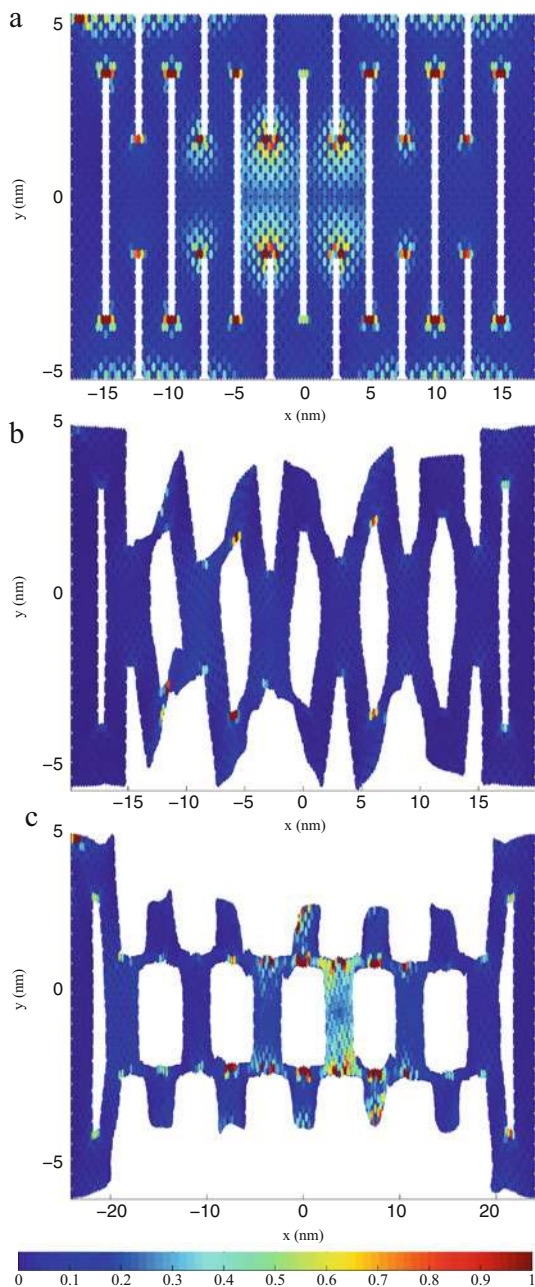
Fig. 11 LDOS at one of the transmission peaks ($E = 0.044t_0$) that contributes to the lowest transmission miniband of the undeformed kirigami, as discussed in the main text. The corresponding kirigami structure is shown underneath the density plot. Image reprinted with permission from Bahamon et al. (2016)
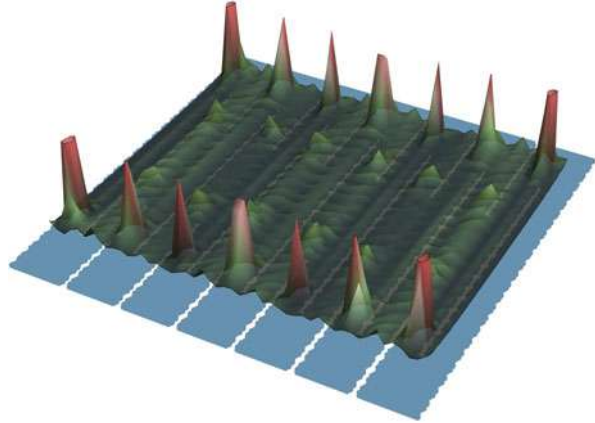
$$E_n = -2\gamma \, \cos\left(\frac{n\pi}{N+1}\right), \qquad n = 1, 2, \ldots, N, \qquad (17)$$

relative to the center of the miniband. The width of a miniband is given by $\Delta E = E_N - E_1$. The data in Fig. 9c allow us to estimate $\Delta E \approx 0.04t_0$ for the undeformed kirigami, which corresponds to $\gamma \approx 0.01\,t_0$. Recalling that t_0 is the hopping between nearest neighboring carbon atoms in undeformed graphene [13], this shows an interdot overlap two orders of magnitude smaller than the hopping in the underlying graphene lattice. Such a small value of γ is natural given the large spatial separation ($\sim d/2 = 2.4$ nm, cf. Fig. 7a) between each pair of hybridized localized states that we see in Fig. 11.

Under tensile load in stage 1, the kirigami elongates and becomes distorted, which significantly perturbs the overlap between the localized states. That is the conclusion that follows from the progressive disappearance of the resonant minibands under deformation that we can see in Fig. 9a for deformations of 5.3 and 15.5%. It is notable that, even though in stage 1 there is very little change in the *average* C-C bond length, the regions that are most affected by that, and which are more strongly bent, are those in the vicinity of the localized states of the undeformed kirigami. It is, therefore, not surprising that the conductance assisted by resonant tunneling through these states can be significantly modified in this stage since the perturbations to the hopping can be significant in precisely the regions more critical for the overlap between neighboring localized states. For an elongation of 15.5%, Fig. 9a shows that the miniband structure disappears and the system behaves largely as an insulator through most of the energy range shown. The LDOS associated with the few remaining weak conductance peaks reflects the less effective coupling between the localized states (see Fig. 10a). In the analogy introduced above, in this stage of deformation, we have the equivalent of a device consisting of distinct and asymmetrically coupled quantum dots, for which the energy of transmission resonances is predominantly determined by the energy levels and coupling of individual dots (Joe et al. 2000; Chen et al. 1994).

A simple extrapolation of this picture would forecast a progressively more substantial degradation of the conductance with further stretching. However, inspection of Fig. 9a reveals otherwise: resonant transmission is revived beyond a geometry-controllable threshold when the system enters stage 2 of deformation. At 25.1% strain, resonant transmission is recovered at particular energies within the lowest miniband, and further stretching to 34.7% strain, rather than degrading, further improves this situation with resonant peaks reappearing again throughout the entire lowest miniband as in the undeformed device. As expected, the width of each miniband is now smaller than in the undeformed state, a combined effect of the increase in inter-dot distance and the decrease of electronic hoppings in the carbon lattice. Figure 10c exhibits the LDOS for one of the revived conductance peaks of the same system that shows a strong similarity with the undeformed counterpart in Fig. 10a.

It is clear from both the conductance and the LDOS that deforming the system well inside stage 2 restores the coupling between the localized states trapped at the internal mini-edges that is strongly affected during stage 1. As pointed out earlier, the major difference between stages 1 and 2 is that, in the latter, further elongation proceeds through deformation of the C-C bonds and therein lies the mechanism that promotes the enhanced coupling and the revival of resonant transmission. As one intuitively anticipates, and can be explicitly seen in Fig. 12b, the deformations in question are borne almost entirely by the bonds along the segments that link neighboring units of the kirigami. A comparison of the bond length distribution in stage 1 (Fig. 12a) versus stage 2 (Fig. 12b) indeed shows that the first regime is characterized by small and seemingly random variations, whereas the latter features a clear and periodic pattern of deformation hot-spots where the C-C bond is stretched in excess of 20%.

Variations in bond length imply perturbations to the electronic hopping [t_{ij}, (12)], which is the quantity that directly affects the transport. However, it is important to mention that the C-C bond map excludes bending contributions that should dominate the hopping modifications in stage 1. To clarify this point, we plot in Fig. 13 the corresponding hopping distribution maps. The variations are of the order of $\sim \pm 10\%$ and $\sim \pm 20\%$ at 15.5% and 34.7% elongation, respectively. More than the magnitude, the key difference lies in the sign and spatial pattern of these hopping variations: in the first case, they appear randomly distributed and one sees places of both enhanced and reduced hopping compared to the undeformed structure; in the second, the pattern is clearly periodic and dominated by reduced hopping ($\sim 20\%$) spatially correlated with the strain hot-spots of Fig. 12b. These electronic "weak links" reinforce the super-periodicity of the structure, while simultaneously contributing to localize further the electronic states within each segment, thereby promoting the revival of the miniband and stop-gap structure in the conductance.

In summary, the profile of the conductance at low energies in these graphene structures is typical of a system of weakly coupled quantum dots: minibands of resonant transmission alternating with stop-gaps due to the superlattice periodicity along the longitudinal direction. Direct inspection of the LDOS at any of the
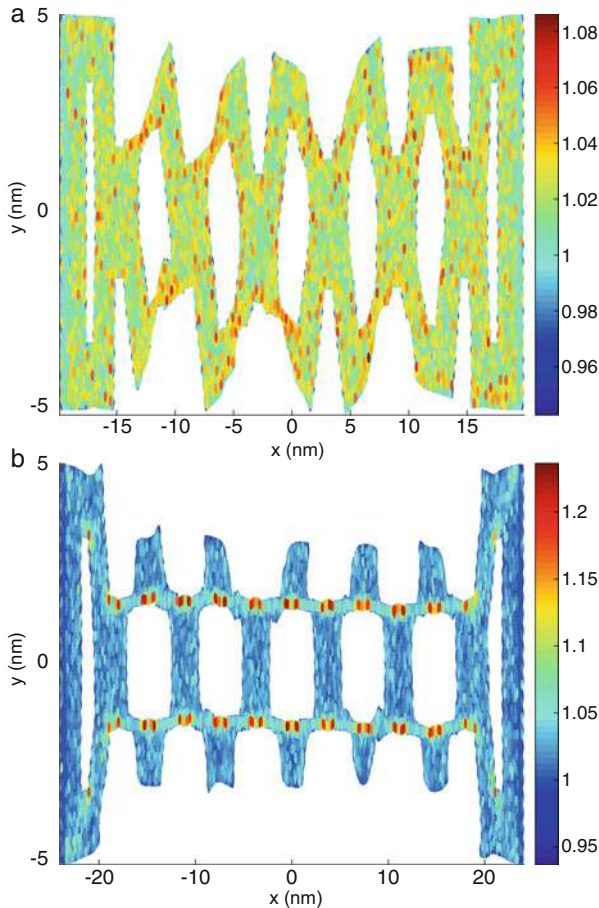
Fig. 12 Normalized C-C bond length for the kirigami deformed by 15.5% (panel a, stage 1) and 34.7% (panel b, stage 2), where the normalization is by the undeformed C-C bond length. Image reprinted with permission from Bahamon et al. (2016)

resonant energies provides a direct confirmation of this in real space. Analysis of the evolution of the conductance and the LDOS with deformation shows that the resonant tunneling assisted by localized states is strongly suppressed during stage 1 of deformation. The local perturbations to the electronic hopping in this regime are sufficient to disturb the overlap between the localized states. This effect is reversed in stage 2 by the appearance of strong strain barriers spanning the length of the system along the links between individual elements of the periodic structure. These have a confining nature from the electronic point of view, which consistently explains the reinforcement of overlap between localized states and the revival of resonant transmission, at the highest deformations.

Fig. 13 The distribution of the nearest-neighbor hopping amplitudes t_{ij}/t_0 calculated according to (12) for the same structures represented in Fig. 12. Image reprinted with permission from Bahamon et al. (2016)
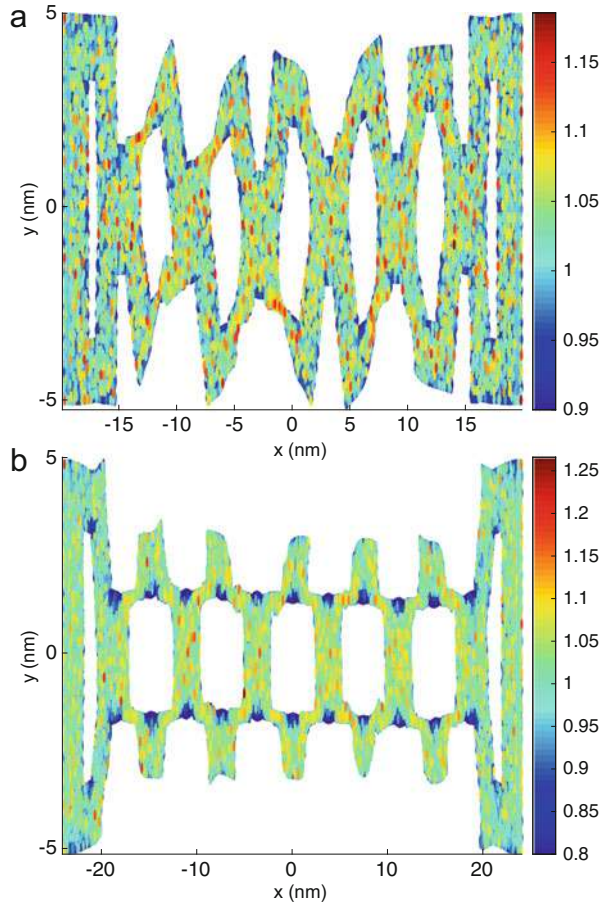
3.3   I-V Characteristics and Negative Differential Resistance

In principle, to fully examine the current-voltage characteristics of these kirigami, we should calculate the transmission function for different bias and gate voltages, according to (16). However, similar to the experimental setup used in Blees et al. (2015) where the kirigami was immersed in an electrochemically controlled liquid gate with a fixed bias voltage $V_{SD} = 100$ meV, we restrict our analysis to the electric current as a function of gate voltage for a fixed bias. Figure 14 displays current for kirigami deformed by 0%, 15.5%, and 34.7% under a fixed $V_{SD} = 100$ meV and at $T = 300$ K.

One can see a clear oscillation of the current with increasing V_g, a behavior that traces back to the underlying miniband structure of the conductance discussed in the context of Fig. 9. The fact that below $V_g = 0.1\,t_0$ the current magnitude first decreases with deformation, but then increases again in stage 2, is a natural

Fig. 14 Current *vs* gate
voltage for the kirigami
deformed by 0%, 15.5%, and
34.7%. We used a bias
voltage of $V_{SD} = 100\,\text{meV}$,
the current is expressed in
units of $I_0 = 2e/h$ and the
gate voltage in units of t_0.
Image reprinted with
permission from Bahamon
et al. (2016)
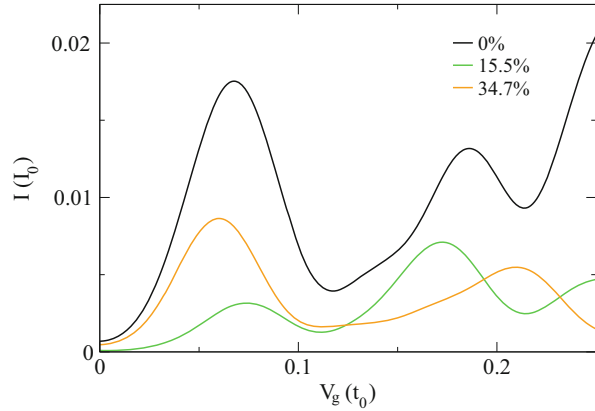
consequence of the revival of resonant transmission at low energies promoted by
the localized strain that emerges in stage 2. The most interesting aspect of the
I-V characteristics shown in Fig. 14 is that the current oscillation immediately
implies that these kirigami have negative differential resistance over reasonably
large intervals of V_g, which is a much sought-after property for new classes of
nonlinear electronic devices (Esaki and Tsu 1970; Capasso et al. 1986).

3.4   Introducing Dephasing

So far, our electronic transport model considers coherent transport, including
only edges and mechanical deformations as source of scattering. To address the
robustness of the strain modulation of the conductance in a more realistic scenario,
we now turn to effects of random scatterers modeled within the dephasing Büttiker-
probe model (Büttiker 1986). To achieve that, we distribute voltage probes as
phase-breaking scatterers all over the kirigami (except the left and right edges) and
set the current at each probe to zero. Given that we are not interested in the particular
physical mechanism (impurities, phonons, or electrons) behind the destruction of
quantum coherence, we fix the self-energy of each Büttiker-probe to $\Sigma_\phi = -i\eta$
(Golizadeh-Mojarad and Datta 2007), which is related to the phase relaxation time
by $\tau_\phi = \hbar/2\Gamma_\phi = \hbar/4\eta$ (Datta 1995; Pastawski and Medina 2001). In Fig. 15, we
plot the effective conductance that obtains for weak ($\eta = 0.01t_0$, $\tau_\phi \approx 6.1 \times 10^{-15}$s)
and strong ($\eta = 0.5t_0$, $\tau_\phi \approx 1.2 \times 10^{-16}$s) phase-breaking under various
deformations. Inspecting Fig. 15a, we can see that weak dephasing is sufficient
to destroy the resonant transmission and miniband structure in the transmission,
irrespective of the amount (or absence) of deformation. Strong phase-breaking
processes (Fig. 15b) completely destroy any fine structure in the conductance
associated with the particular geometry of the device that are still visible in Fig. 15a.
In this case, there are no differences among the effective conductances of unstrained
and deformed kirigami, all of which display the step-wise behavior characteristic
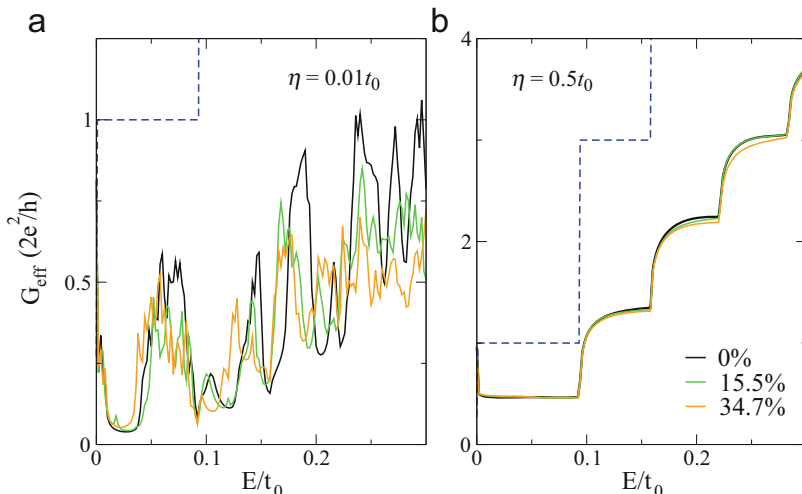
Fig. 15 Effective conductance in the presence of (**a**) weak ($\eta = 0.01t_0$, $\tau_\phi \approx 6.1 \times 10^{-15}$s) and (**b**) strong ($\eta = 0.5t_0$, $\tau_\phi \approx 1.2 \times 10^{-16}$s) phase-breaking scatterers for different deformations. The dashed blue line corresponds to the conductance of a pristine zigzag nanoribbon with the same width of our contacts. Be aware of the different scales in the vertical axes. Image reprinted with permission from Bahamon et al. (2016)

of a pristine graphene nanoribbon, with the steps appearing at energies that are determined by the overall width of the device (parameter b in Fig. 7a).

To illustrate the distinction between the weak and strong dephasing regimes used in Fig. 15, it is instructive to compute the associated phase-relaxation lengths, $l_\phi = \sqrt{D\tau_\phi}$ (Datta 1995) using the representative carrier diffusion coefficient $D = 0.3$ m^2/s (Berger et al. 2006). In the case of weak dephasing discussed above, the phase-relaxation length is larger than the device size, $l_\phi \approx 42$ nm $> L$, whereas that value for strong dephasing is similar to the periodicity of the kirigami, $l_\phi \approx 6$ nm $\approx w$. Coherent effects in the conductance are observed when the typical device dimensions are smaller than the phase relaxation length (Datta 1995; Golizadeh-Mojarad and Datta 2007; Berger et al. 2006). In our case, the formation of gaps and minibands is a coherent effect created by destructive and constructive interference of electron waves at the periodically repeating segments of the kirigami (the localized states arise as the electron bounces back and forth without losing its phase within different sections of the kirigami). Figure 15b shows a conductance that has become insensitive to any effect associated with strain or even the geometry. This confirms that, in order to observe efficient mechanical control over the conductance, the phase-relaxation length must be larger than the segments of the kirigami that harbor the localized states that support the resonant tunneling at low energy. Measurements of the phase-relaxation length at low temperature in graphene have found $l_\phi \approx 1$ μm (Berger et al. 2006), which is much smaller than the size of the typical internal lengths used in the kirigami experimentally probed by Blees and collaborators

(Blees et al. 2015). This might partially explain why deformed kirigami used in that experiment did not show noticeable sensitivity of the conductance to the elongation.

3.5   Final Remarks

Our study shows that, when the typical feature sizes of a graphene-based kirigami are in the nanoscale, the electric conductance at low energies might be governed by resonant tunneling through states that are localized by the specific local geometry of each repeating element of the kirigami. The longitudinal periodicity of the structure results in an efficient coupling (strong overlap) between these states in the undeformed configuration. This state of affairs is, however, strongly sensitive to stage 1 deformations, and the conductance easily degrades in that regime where the overall elongation is a result of the structural twisting and bending, rather than extensive stretching of the C-C bonds. The electronic overlap is reinforced for stage 2 deformations as a result of the confining nature of the localized strain barriers that set in during this stage, and resonant transmission is hence revived at high overall elongations. The regime of resonant tunneling exists within well-defined energy minibands isolated from each other by sizable transmission stop-gaps. Their existence results in a strong oscillation of the I-V characteristic as a function of gate voltage, and opens the possibility of driving the system from a conventional resistive regime to one of negative differential resistance, by simple electrostatic gating.

To explain the evolution of the conductance profile with elongation, we analyzed directly the bond stretching and the perturbations to the nearest-neighbor hopping. It is customary in discussions of strained graphene to introduce the concept of pseudo-magnetic field (PMF) (Kane et al. 1997; Suzuura and Ando 2002; Vozmediano et al. 2010) and map strain fields to PMFs in order to, for example, obtain a semiclassical intuition about how certain strain patterns disturb the motion of electrons. In the present case, however, the characteristic dimensions of our device are small, and, in addition, the strain distribution displays sharp variations within these small scales. This restricts the usefulness of the PMF concept and any semiclassical picture to interpret the effects of strain. We therefore focused directly in the changes affecting the nearest-neighbor hopping amplitudes since these are the quantities that more direct and fundamentally determine the conductance.

Finally, in our main discussion, we focused on a kirigami whose horizontal segments consisted of zigzag strips of graphene. Not surprisingly, the mini zigzag edges parallel to the transverse direction play an important role in stabilizing the localized states in the undeformed structure that are key for the conductance profile at low energy. It is important to reiterate that this is not a limitation, for *the key physical ingredient is the existence of localized states defining local quantum dots* and the interplay between the strain-induced changes of the C-C hopping and effective inter-dot coupling upon mechanical stretching. To be specific, in Fig. 16a we show that rotating the underlying lattice by 90 degrees in the same kirigami leads to a similar profile of miniband and stop-gap low-energy conductance and that it is equally sensitive to deformation: for example, at about 16% elongation, the mini-
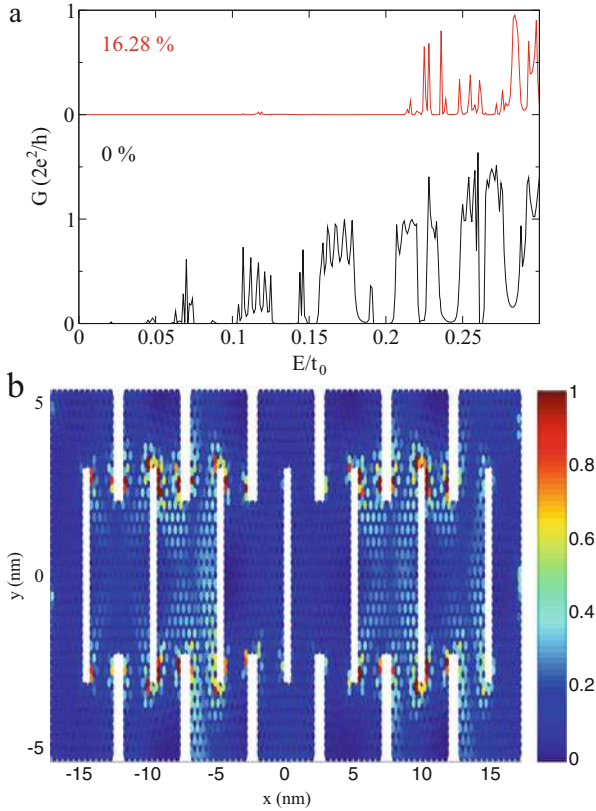
Fig. 16 (**a**) Conductance of the armchair kirigami deformed by 0% and 16.28%. (**b**) LDOS for the undeformed armchair kirigami at $E = 0.121t_0$. Image reprinted with permission from Bahamon et al. (2016)

band structure is entirely suppressed. In a future study, we shall undertake the impact of edge roughness. We can, however, anticipate that edge roughness is expected to *improve* the scenario we describe here because this type of disorder promotes further localization of low-energy states in narrow graphene structures (Todd et al. 2008; Han et al. 2010), and that explains the experimentally observed gaps and Coulomb blockade in rough (lithographically patterned) graphene nanoribbons (Sols et al. 2007; Todd et al. 2008; Stampfer et al. 2009; Gallagher et al. 2010). We therefore predict that, in a realistic scenario, individual segments of a kirigami can behave as true quantum dots with a coupling (and, consequently, an overall transmission) amenable to modulation through the same type of deformation discussed here. The conditions for such behavior are expected to be very encompassing, depending only on choosing appropriate combinations of scales and geometries capable of hosting localized quantum dot states, supported either by the geometry, disorder, or interactions.

4   Conclusions

We have demonstrated that both existing computational techniques, like classical molecular dynamics, and new methodologies, such as combinations of molecular dynamics, and electronic structure tight-binding, and quantum transport, can both be used to develop new insights into the mechanics and electromechanics of graphene. There is significant potential for these computational tools to enable novel contributions to electromechanical coupling in 2D materials and also for 2D auxetics. For example, most reports of auxetic behavior in 2D materials have been for the out-of-plane direction, i.e. due to rippling of 2D materials. In contrast, very few examples of in-plane auxeticity have been reported. Similarly, other forms of electromechanical coupling in 2D materials, like piezoelectricity and flexoelectricity, have recently emerged – computational modeling will surely be needed to elucidate the behavior and properties of 2D piezo and flexoelectric materials (Akinwande et al. 2017). Finally, while we have focused on graphene in the present work, similar ideas and methodologies may be developed and used in the future to examine the physical properties of the other 2D materials families, like the transition metal dichalcogenides, black phosphorus, the monochalcogenides, boron nitride, and many others.

References

Abanin DA, Pesin DA (2012) Interaction-induced topological insulator states in strained graphene. Phys Rev Lett 109:066802

Abedpour N, Asgari R, Guinea F, Strains and pseudomagnetic fields in circular graphene rings. Phys Rev B 84:(2011)115437

Akinwande D, Brennan CJ, Bunch JS, Egberts P, Felts JR, Gao H, Huang R, Kim J-S, Li T, Li Y, Liechti KM, Lu N, Park HS, Reed EJ, Wang P, Yakobson BI, Zhang T, Zhang Y-W, Zhou Y, Zhu Y (2017) A review on mechanics and mechanical properties of 2D materials – graphene and beyond. Extreme Mech Lett 13:42–72

Alderson K, Alderson A, Anand S, Simkins V, Nazare S, Ravirala N (2012) Auxetic warp knit textile structures. Phys Status Solidi B 249(7):1322–1329

Bahamon DA, Qi Z, Park HS, Pereira VM, Campbell DK (2015) Conductance signatures of electron confinement induced by strained nanobubbles in graphene. Nanoscale 7:15300–15309

Bahamon DA, Qi Z, Park HS, Pereira VM, Campbell DK (2016) Graphene kirigami as a platform for stretchable and tunable quantum dot arrays. Phys Rev B 95:235408

Bao W, Miao F, Chen Z, Zhang H, Jang W, Dames C, Lau CN (2009) Controlled ripple texturing of suspended graphene and ultrathin graphite membranes. Nat Nanotechnol 4:562–566

Baughman RH, Galvao DS (1993) Crystalline networks with unusual predicted mechanical and thermal properties. Nature 365:735

Baughman RH, Shacklette JM, Zakhidov AA, Stafstrom S (1998) Negative Poisson's ratios as a common feature of cubic metals. Nature 392:362–365

Berger C, Song Z, Li X, Wu X, Brown N, Naud C, Mayou D, Li T, Hass J, Marchenkov AN, Conrad EH, First PN, de Heer WA (2006) Electronic confinement and coherence in patterned epitaxial graphene. Science 312(5777):1191–1196

Bertoldi K, Reis PM, Willshaw S, Mullin T (2010) Negative Poisson's ratio behavior induced by an elastic instability. Adv Mater 22:361–366

Blees MK, Barnard AW, Rose PA, Roberts SP, McGill KL, Huang PY, Ruyack AR, Kevek JW, Kobrin B, Muller DA, McEuen PL (2015) Graphene kirigami. Nature 524(7564): 204–207

Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys Condens Matter 14:783–802

Brum JA (1991) Electronic properties of quantum-dot superlattices. Phys Rev B 43:12082–12085

Büttiker M (1986) Four-terminal phase-coherent conductance. Phys Rev Lett 57:1761–1764

Capasso F, Mohammed K, Cho AY (1986) Resonant tunneling through double barriers, perpendicular quantum transport phenomena in superlattices, and their device applications. IEEE J Quantum Electron 22(9):1853–1869

Caroli C, Combescot R, Nozieres P, Saint-James D (1971) Direct calculation of the tunneling current. J Phys C Solid State Phys 4(8):916

Castro Neto AH, Guinea F, Peres NMR, Novoselov KS, Geim AK (2009) The electronic properties of graphene. Rev Modern Phys 81:109–162

Chang T, Gao H (2003) Size-dependent elastic properties of a single-walled carbon nanotube via a molecular mechanics model. J Mech Phys Solids 51:1059–1074

Chang T, Geng J, Guo X (2005) Chirality- and size-dependent elastic properties of single-walled carbon nanotubes. Appl Phys Lett 87(25):251929

Chen X, Xiong S, Wang G (1994) Tunneling in quantum-wire superlattices with random layer thicknesses. Phys Rev B 49:14736–14739

Choi S-M, Jhi S-H, Son Y-W (2010) Effects of strain on electronic properties of graphene. Phys Rev B 81:081407(R)+

Clausen A, Wang F, Jensen JS, Sigmund O, Lewis JA (2015) Topology optimized architectures with programmable Poisson's ratios over large deformations. Adv Mater 27:5523–5527

Datta S (1995) Electronic transport in mesoscopic systems. Cambridge University Press, Cambridge

Esaki L, Tsu R (1970) Superlattice and negative differential conductivity in semiconductors. IBM J Res Dev 14(1):61–65

Evans KE (1991) Auxetic polymers: a new range of materials. Endeavour 15(4):170–174

Evans KE, Alderson A (2000) Auxetic materials: functional materials and structures from lateral thinking! Adv Mater 12(9):617

Faria D, Carrillo-Bastos R, Sandler N, Latgé A (2015) Fano resonances in hexagonal zigzag graphene rings under external magnetic flux. J Phys Condens Matter 27(17):175301

Farjam M, Rafii-Tabar H (2009) Comment on "band structure engineering of graphene by strain: first-principles calculations". Phys Rev B 80:167401

Fujita M, Wakabayashi K, Nakada K, Kusakabe K (1996) Peculiar localized state at zigzag graphite edge. J Phys Soc Jpn 65(7):1920–1923

Gallagher P, Todd K, Gordon DG (2010) Disorder-induced gap behavior in graphene nanoribbons. Phys Rev B 81:115409

Garza HHP, Kievit EW, Schneider GF, Staufer U (2014) Controlled, reversible, and nondestructive generation of uniaxial extreme strains (>10%) in graphene. Nano Lett 14(7):4107–4113

Geim AK, Novoselov KS (2007) The rise of graphene. Nat Mater 6:183–191

Golizadeh-Mojarad R, Datta S (2007) Nonequilibrium green's function based models for dephasing in quantum transport. Phys Rev B 75:081301

Gong L, Kinloch IA, Young RJ, Riaz I, Jalil R, Novoselov KS (2010) Interfacial stress transfer in a graphene monolayer nanocomposite. Adv Matter 22:2694

González JW, Pacheco M, Rosales L, Orellana PA (2011) Transport properties of graphene quantum dots. Phys Rev B 83:155450

Grima JN, Winczewski S, Mizzi L, Grech MC, Cauchi R, Gatt R, Attard D, Wojciechowski KW, Rybicki J (2015) Tailoring graphene to achieve negative Poisson's ratio properties. Adv Mater 27:1455–1459

Guinea F, Low T (2010) Band structure and gaps of triangular graphene superlattices. Philos Trans R Soc A (Math Phys Eng Sci) 368(1932):5391–5402

Guinea F, Horovitz B, Le Doussal P (2008) Gauge field induced by ripples in graphene. Phys Rev B 77:205421

Guinea F, Katsnelson MI, Geim AK (2010a) Energy gaps and a zero-field quantum hall effect in graphene by strain engineering. Nat Phys 6(1):30–33

Guinea F, Geim AK, Katsnelson MI, Novoselov KS (2010b) Generating quantizing pseudomagnetic fields by bending graphene ribbons. Phys Rev B 81(3):035408

Han MY, Ozyilmaz B, Zhang Y, Kim P (2007) Energy band-gap engineering of graphene nanoribbons. Phys Rev Lett 98:206805

Han MY, Brant JC, Kim P (2010) Electron transport in disordered graphene nanoribbons. Phys Rev Lett 104:056801

Haug H, Jauho A-P (2008) Transport in mesoscopic semiconductor structures. In: Quantum kinetics in transport and optics of semiconductors. Solid-state sciences, vol 123. Springer, Berlin/Heidelberg, pp 181–212

Ho DT, Park S-D, Kwon S-Y, Park K, Kim SY (2014) Negative Poisson's ratios in metal nanoplates. Nat Commun 5:3255

Ho VH, Ho DT, Kwon S-Y, Kim SY (2016) Negative Poisson's ratio in cubic materials along principal directions. Phys Status Solidi B. https://doi.org/10.1002/pssb.201600017

Ho VH, Ho DT, Kwon S-Y, Kim SY (2016) Negative Poisson's ratio in periodic porous graphene structures. Phys Status Solidi B. https://doi.org/10.1002/pssb.201600061

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38

Ihnatsenka S, Zozoulenko IV, Kirczenow G (2009) Band-gap engineering and ballistic transport in edge-corrugated graphene nanoribbons. Phys Rev B 80:155415

Jiang J-W, Park HS (2014) Negative Poisson's ratio in single-layer black phosphorus. Nat Commun 5:4727

Jiang J-W, Park HS (2016) Negative Poisson's ratio in single-layer graphene ribbons. Nano Lett 16:2657–2662

Jiang J-W, Tang H, Wang B-S, Su Z-B (2008) Raman and infrared properties and layer dependence of the phonon dispersions in multilayered graphene. Phys Rev B 77(23):235421

Jiang J-W, Kim SY, Park HS (2016) Auxetic nanomaterials: recent progress and future directions. Appl Phys Rev 3:041101

Jiang J-W, Chang T, Guo X, Park HS (2016) Intrinsic negative Poisson's ratio for single-layer graphene. Nano Lett 16:5286–5290

Jiang J-W, Park HS (2016) Negative Poisson's ratio in single-layer graphene ribbons. Nano Lett 16:2657–2662

Jiao L, Zhang L, Wang X, Diankov G, Dai H (2009) Narrow graphene nanoribbons from carbon nanotubes. Nature 458:877

Ji Z-L, Berggren K-F (1992) Quantum bound states in narrow ballistic channels with intersections. Phys Rev B 45:6652–6658

Joe YS, Ikeler DS, Cosby RM, Satanin AM, Kim CS (2000) Characteristics of transmission resonance in a quantum-dot superlattice. J Appl Phys 88(5):2704–2708

Kane CL, Mele EJ (1997) Size, shape, and low energy electronic structure of carbon nanotubes. Phys Rev Lett 78:1932

Kim K-J, Blanter YM, Ahn K-H (2011) Interplay between real and pseudomagnetic field in graphene with strain. Phys Rev B 84(8):081401

Kitt AL, Pereira VM, Swan AK, Goldberg BB (2012) Lattice-corrected strain-induced vector potentials in graphene. Phys Rev B 85(11):115432

Kitt AL, Pereira VM, Swan AK, Goldberg BB (2013) Erratum: lattice-corrected strain-induced vector potentials in graphene. Phys Rev B 87:159909(E); Phys Rev B 85:115432 (2012)

Kouwenhoven LP, Hekking FWJ, van Wees BJ, Harmans CJPM, Timmering CE, Foxon CT (1990) Transport through a finite one-dimensional crystal. Phys Rev Lett 65:361–364

Lakes RS (1987) Foam structures with a negative poisson's ratio. Science 235:1038–1040

Lakes R (1993) Advances in negative Poisson's ratio materials. Adv Mater 5:293–296

Lammps (2018) http://lammps.sandia.gov/

Lee C, Wei X, Kysar JW, Hone J (2008) Measurement of the elastic properties and intrinsic strength of monolayer graphene. Science 321:385

Lethbridge ZA, Walton RI, Marmier AS, Smith CW, Evans KE (2010) Elastic anisotropy and extreme poisson's ratios in single crystals. Acta Mater 58:6444–6451

Los JH, Fasolino A, Katsnelson MI (2006) Scaling behavior and strain dependence of in-plane elastic properties of graphene. Phys Rev Lett 116:015901

Milstein F, Huang K (1979) Existence of a negative Poisson ratio in fcc crystals. Phys Rev B 19(4):2030

Mo Y, Turner KT, Szlufarska I (2009) Friction laws at the nanoscale. Nature 457:1116

Mohiuddin TMG, Lombardo A, Nair RR, Bonetti A, Savini G, Jalil R, Bonini N, Basko DM, Galiotis C, Marzari N, Novoselov KS, Geim AK, Ferrari AC (2009) Uniaxial strain in graphene by Raman spectroscopy: G peak splitting, Gruneisen parameters, and sample orientation. Phys Rev B 79(20):205433

Ni ZH, Yu T, Lu YH, Wang YY, Feng YP, Shen ZX (2008) Uniaxial strain on graphene: Raman spectroscopy study and band-gap opening. ACS Nano 2(11):2301–2305

Ni ZH, Yu T, Lu YH, Wang YY, Feng YP, Shen ZX (2009) Uniaxial strain on graphene: Raman spectroscopy study and band-gap opening. ACS Nano 3:483

Novoselov KS, Geim AK, Morozov SV, Jiang D, Katsnelson MI, Grigorieva IV, Dubonos SV, Firsov AA (2005) Two-dimensional gas of massless dirac fermions in graphene. Nature 438:197–200

Pastawski HM, Medina E (2001) "Tight binding" methods in quantum transport through molecules and small devices from the coherent to the decoherent description. Rev Mex de Física 47(S1):1–23

Pereira VM, Castro Neto AH (2009) Strain engineering of graphene's electronic structure. Phys Rev Lett 103(4):4

Pereira VM, Castro Neto AH, Peres NMR (2009) Tight-binding approach to uniaxial strain in graphene. Phys Rev B 80:045401

Pereira VM, Castro Neto AH, Liang HY, Mahadevan L (2010) Geometry, mechanics, and electronics of singular structures and wrinkles in graphene. Phys Rev Lett 105:156603

Pereira VM, Ribeiro RM, Peres NMR, Castro Neto AH (2010) Optical properties of strained graphene. Eur Phys Lett 92:67001

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19

Qi Z, Zhao F, Zhou X, Sun Z, Park HS, Wu H (2010) A molecular simulation analysis of producing monatomic carbon chains by stretching ultranarrow graphene nanoribbons. Nanotechnology 21(26):265702

Qi Z, Bahamon DA, Pereira VM, Park HS, Campbell DK, Castro Neto AH (2013) Resonant tunneling in graphene pseudomagnetic quantum dots. Nano Lett 13:2692

Qi Z, Kitt AL, Park HS, Pereira VM, Campbell DK, Castro Neto AH (2014) Pseudomagnetic fields in graphene nanobubbles of constrained geometry: a molecular dynamics study. Phys Rev B 90:125419

Qi Z, Campbell DK, Park HS (2014) Atomistic simulations of tension-induced large deformation and stretchability in graphene kirigami. Phys Rev B 90:245437

Ravirala N, Alderson A, Alderson KL (2007) Interlocking hexagons model for auxetic behaviour. J Mater Sci 42:7433–7445

Rothenburg L, Berlint AA, Bathurst RJ (1991) Microstructure of isotropic materials with negative poisson's ratio. Nature 354:470

Seol JH, Jo I, Moore AL, Lindsay L, Aitken ZH, Pettes MT, Li X, Yao Z, Huang R, Broido D, Mingo N, Ruoff RS, Shi L (2010) Two-dimensional phonon transport in supported graphene. Science 328:213–216

Shen L, Li J (2004) Transversely isotropic elastic properties of single-walled carbon nanotubes. Phys Rev B 69:045414

Shenoy VB, Reddy CD, Ramasubramaniam A, Zhang YW (2008) Edge-stress-induced warping of graphene sheets and nanoribbons. Phys Rev Lett 101(24):245501

Sols F, Guinea F, Neto AHC (2007) Coulomb blockade in graphene nanoribbons. Phys Rev Lett 99:166803

Stampfer C, Güttinger J, Hellmüller S, Molitor F, Ensslin K, Ihn T (2009) Energy gaps in etched graphene nanoribbons. Phys Rev Lett 102:056403

Stuart SJ, Tutein AB, Harrison JA (2000) A reactive potential for hydrocarbons with intermolecular interactions. J Chem Phys 112(14):6472–6486

Stukowski A (2010) Visualization and analysis of atomistic simulation data with OVITO – the open visualization tool. Model Simul Mater Sci Eng 18:015012

Suzuura H, Ando T (2002) Phonons and electron-phonon scattering in carbon nanotubes. Phys Rev B 65:235412

Todd K, Chou H-T, Amasha S, Goldhaber-Gordon D (2008) Quantum dot behavior in graphene nanoconstrictions. Nano Lett 9:416

Tomori H, Kanda A, Goto H, Ootuka Y, Tsukagoshi K, Moriyama S, Watanabe E, Tsuya D (2011) Introducing nonuniform strain to graphene using dielectric nanopillars. Appl Phys Express 4(7):3

Ulloa SE, Castao E, Kirczenow G (1990) Ballistic transport in a novel one-dimensional superlattice. Phys Rev B 41:12350–12353

Vozmediano MAH, Katsnelson MI, Guinea F (2010) Gauge fields in graphene. Phys Rep 496:109

Wang ZF, Zhang Y, Liu F (2011) Formation of hydrogenated graphene nanoripples by strain engineering and directed surface self-assembly. Phys Rev B 83:041403

Wu Z, Zhang ZZ, Chang K, Peeters FM (2010) Quantum tunneling through graphene nanorings. Nanotechnology 21(18):185201

Yang HT (2011) Strain induced shift of dirac points and the pseudo-magnetic field in graphene. J Phys Condens Matter 23(50):505502

Yang W, Li Z-M, Shi W, Xie B-H, Yang M-B (2004) Review on auxetic materials. J Mater Sci 39:3269–3279

Yao YT, Alderson A, Alderson KL (2008) Can nanotubes display auxetic behaviour? Phys Status Solidi B 245(11):2373–2382

Yeh NC, Teague ML, Yeom S, Standley BL, Wu RTP, Boyd DA, Bockrath MW (2011) Strain-induced pseudo-magnetic fields and charging effects on CVD-grown graphene. Surf Sci 605(17–18):1649–1656

Yue K, Gao W, Huang R, Liechti KM (2012) Analytical methods for the mechanics of graphene bubbles. J Appl Phys 112(8):083512

Zakharchenko KV, Katsnelson MI, Fasolino A (2009) Finite temperature lattice properties of graphene beyond the quasiharmonic approximation. Phys Rev Lett 102(4):046808

Zhang ZZ, Chang K, Chan KS (2008) Resonant tunneling through double-bended graphene nanoribbons. Appl Phys Lett 93(6):062106. https://doi.org/10.1063/1.2970957

Zhao H, Aluru NR (2010) Temperature and strain-rate dependent fracture strength of graphene. J Appl Phys 108(6):064321

Zhu S, Huang Y, Li T (2014) Extremely compliant and highly stretchable patterned graphene. Appl Phys Lett 104(17):173103

Surface Energy and Nanoscale Mechanics **79**

Kosar Mozaffari, Shengyou Yang, and Pradeep Sharma

Contents

Abstract

The mechanical response of nanostructures, or materials with characteristic features at the nanoscale, differs from their coarser counterparts. An important physical reason for this size-dependent phenomenology is that surface or interface properties are different than those of the bulk material and acquire significant prominence due to an increased surface-to-volume ratio at the nanoscale. In

K. Mozaffari · S. Yang · P. Sharma (✉)
Department of Mechanical Engineering, Cullen College of Engineering, University of Houston, Houston, TX, USA
e-mail: mozaffari.kosar@gmail.com; uhsyyang@gmail.com

this chapter, we provide an introductory tutorial on the continuum approach to incorporate the effect of surface energy, stress, and elasticity and address the size-dependent elastic response at the nanoscale. We present some simple illustrative examples that underscore both the physics underpinning the capillary phenomenon in solids as well as a guide to the use of the continuum theory of surface energy.

1   Introduction

In a manner of speaking, a free surface or an interface is a "defect." We will use the word "surface" to imply both a free surface as well as an interface separating two materials. It represents a drastic interruption in the symmetry of the material – much like the more conventionally known defects such as dislocations. Imagine the atoms on a free surface. They have a different coordination number, charge distribution, possible dangling bonds, and many other attributes that distinguish them from atoms further away in the bulk of the material (Ibach 1997; Cammarata 2009). It is therefore hardly surprising that the surface of a material should have mechanical (or in general other physical) properties that differ from the bulk of the material. In other words, as much as conventional defects (such as dislocations) impact the physical response of materials, so do surfaces. However, for coarse-sized structures, the surface-to-volume ratio is negligible, and so even though surfaces do have different properties, they hardly matter in terms of the overall response of the structure. This situation changes dramatically at the nanoscale. As a rather extreme example, 2 nm cube of Copper has nearly 50% of its atoms on the surface. What length scale is "small enough" for surface effects to become noticeable arguably depends on the strength of the surface properties. For hard crystalline materials, this length scale is certainly below 50 nm and often only of significant importance below 10 nm (Miller and Shenoy 2000). Ultra-soft materials (with elastic modulus in the 1–5 kPa range) are an interesting exception where even at micron scale, surface energy-related size effects may be observable (Style et al. 2013).

In a continuum field setting, the role of surfaces may be captured by assuming that they are zero-thickness entities and possess a nontrivial *excess* energy that is distinct from the bulk. The surface energy concept for solids encompasses the fact that surfaces appear to possess a residual "surface tension"-like effect known as surface stress and also an elastic response, termed "surface or superficial elasticity." In the context of fluids, "capillarity" has long been studied, and the concept of surface tension is well-known (see Fig. 1). The situation for solid surfaces is somewhat more subtle than fluids in many ways since deformation is a rather important contributor to surface energy – which is not the case for simple liquids. For a simple liquid, surface energy, the so-called surface stress and surface tension are the same concept. This is not the case for solid surfaces. For further details, see a recent review article by Style et al. (2017). Given the existence of an extensive body of work on this subject, we avoid a detailed literature review and simply point
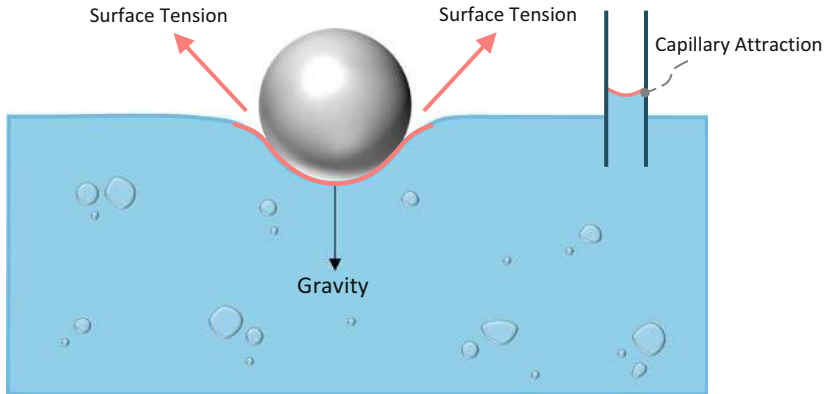
Fig. 1 The ramifications of surface energy – in the form of surface tension – are well-known from our daily lives and are illustrated in this schematic. Surface tension forces, for instance, assist in overcoming gravity and enable the floatation of a ball. The ability of a water spider to walk on the water surface is another example. Also shown is a schematic of a capillary tube, frequently used by biologists and chemists, to highlight how surface tension causes water to be drawn up the tube

to the following overview articles that the reader may consult (and the references therein): (Javili et al. 2013; Duan et al. 2009; Wang et al. 2010, 2011; Li and Wang 2008; Cammarata 2009; Ibach 1997; Müller and Saúl 2004). We will primarily follow the approach pioneered by Gurtin, Murdoch, Fried, and Huang (Gurtin and Murdoch 1975b; Biria et al. 2013; Huang and Sun 2007) and attempt to present a simplified tutorial on the continuum theory for surface energy. For the sake of brevity and with the stipulation that this chapter is merely meant to be a first step to understand surface elasticity, we avoid several complexities and subtleties that exist on this topic such as choice of reference state in the development of the continuum theory (Huang and Wang 2013; Javili et al. 2018), consistent linearization from a nonlinear framework and differences in the various linearized theories (Javili et al. 2018; Liu et al. 2017), curvature dependence of surface energy (Steigmann and Ogden 1997, 1999; Chhapadia et al. 2011; Fried and Todres 2005), and generalized interface models that may allow greater degree of freedom than just a "no slip" surface c.f. (Gurtin et al. 1998; Chatzigeorgiou et al. 2017). We also present three simple case studies or illustrative examples that both highlight the use of the theory and the physical consequence of surface energy effects. While the focus of the chapter is primarily on mechanics and, specifically, elasticity, the framework used in this chapter can be used as a starting point for applications outside mechanics. Indeed, surface energy effects are of significant interest to a variety of disciplines and permeate topics as diverse as fluid mechanics (de Gennes et al. 2004), sensors and resonators (Park 2008), catalysis (Müller and Saúl 2004; Haiss 2001; Pala and Liu 2004), self-assembly (Suo and Lu 2000), phase transformations (Fischer et al. 2008), biology (Liu et al. 2017), and composites (Duan et al. 2005), among others.

2 Preliminary Concepts

2.1 The Need for Surface Tensors

In the study of mechanics of surfaces, we have to contend with tensor fields that "live" on a surface. For example, we will need to define the strain field experienced by a surface. The general machinery of curvilinear tensor calculus then becomes necessary to describe surface mechanics which (at least for our taste) becomes somewhat cumbersome. A direct notation was developed by Gurtin and Murdoch (1975b) which we (prefer and) briefly motivate in this section.

To establish the basic idea, for now, consider just a flat surface shown in Fig. 2 with outward unit normal $n = e_3$. On physical grounds, assuming that there is no "slip" between the surface and the underlying bulk material, the surface strain is simply the strain field of the bulk material *at the spatial position of the surface*. In addition, intuitively, the normal components of the strain tensor ought not to exist for the zero-thickness surface, i.e., $E_s(x) = E^{bulk}(x)$ projected on the tangent plane of the surface where x is on the surface. To make this idea more concrete, assume a body occupies a domain Ω_0 experiencing a strain $E : \Omega_0 \to \mathbb{R}^{3\times3}_{sym}$. We then expect the surface strain to be simply

$$E_s := \begin{bmatrix} E_{11} & E_{12} & 0 \\ E_{21} & E_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{1}$$

Our intuition works well for flat surfaces, but we must generalize the mathematical framework to contend with general curved surfaces. This brings to fore the question of how we define surface tensors on arbitrarily curved surfaces. To understand this, we define next the surface projection tensor. More details on the mathematical preliminaries can be found in the work of Gurtin, Murdoch, and co-workers (Gurtin and Murdoch 1975b, 1978; Gurtin et al. 1998).
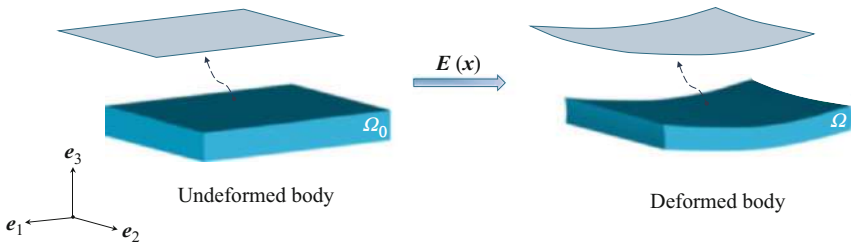
Fig. 2 Schematic of the deformation of a body that occupies the domain Ω_0. The upper surface is artificially separated from the underlying bulk to highlight the surface deformation $E(x)$. Cartesian coordinates with a positively oriented orthonormal basis $\{e_1, e_2, e_3\}$ are shown here and the outward unit normal to the upper surface is $n(x) = e_3$

2.1.1   Surface Projection Tensor

As discussed in the preceding section, we will often need to project second-order tensors on to a curved surface. To that end, the surface projection tensor for the tangent surface with outward unit normal $n(x)$ is defined as

$$\mathbb{P}(x) = \mathbf{I} - n(x) \otimes n(x). \tag{2}$$

Here $\mathbf{I}$ is the second-order identity tensor, and "$\otimes$" denotes the tensor product (or the dyadic product).

If we choose the flat upper surface shown in Fig. 2 as an example, the normal to the upper surface is $n = e_3$ in Cartesian coordinates with a positively oriented orthonormal basis $\{e_1, e_2, e_3\}$. With the definition in (2), the projection tensor on the flat upper surface admits the form

$$\mathbb{P}(x) = \mathbf{I} - e_3 \otimes e_3 = e_1 \otimes e_1 + e_2 \otimes e_2 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{\{e_1, e_2, e_3\}}. \tag{3a}$$

Another example is the projection tensor on a spherical surface with radius R that can be represented by $S = \{x \in \mathbb{R}^3 : x \cdot x - R^2 = 0\}$. The outward unit normal to this surface is $n = e_r$ in spherical coordinates $\{r, \theta, \phi\}$ with basis $\{e_r, e_\theta, e_\phi\}$. From the definition of the projection tensor in (2), we have

$$\mathbb{P}(x) = \mathbf{I} - e_r \otimes e_r = e_\theta \otimes e_\theta + e_\phi \otimes e_\phi := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{\{e_r, e_\theta, e_\phi\}}. \tag{3b}$$

The projection tensor in (3b) can also be easily expressed in the Cartesian coordinates by using the identities $e_\theta = \cos\theta \cos\phi e_1 + \cos\theta \sin\phi e_2 - \sin\theta e_3$ and $e_\phi = -\sin\phi e_1 + \cos\phi e_2$ as well as the tensor product between two vectors.

2.1.2   Surface Vector and Tensor Fields

Let $v(x)$ be a smooth vector field and $T(x)$ be a smooth second-order tensor. Their projections $v_s(x)$ and $T_s(x)$ on a smooth surface with outward unit normal $n(x)$, respectively, are

$$v_s = \mathbb{P}v \quad \text{and} \quad T_s = \mathbb{P}T\mathbb{P}, \tag{4}$$

where the projection tensor $\mathbb{P}$ is defined in (2).

For example, consider a vector field $v = v_i e_i$ and a second-order tensor field $T = T_{ij} e_i \otimes e_j$ in Cartesian coordinates with a positively oriented orthonormal basis $\{e_1, e_2, e_3\}$. Take the upper surface with outward unit normal $n(x) = e_3$. The projection tensor in (2) becomes $\mathbb{P}(x) = \mathbf{I} - e_3 \otimes e_3 = e_1 \otimes e_1 + e_2 \otimes e_2$. Thus, with (4), the projected vector and tensor on the upper surface, respectively, are

$$\boldsymbol{v}_s = \mathbb{P}\boldsymbol{v} = (\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2)v_i\boldsymbol{e}_i = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 \tag{5a}$$

and

$$
\begin{aligned}
\boldsymbol{T}_s = \mathbb{P}\boldsymbol{T}\mathbb{P} &= (\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2) \cdot T_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j \cdot (\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2) \\
&= T_{11}\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + T_{12}\boldsymbol{e}_1 \otimes \boldsymbol{e}_2 + T_{21}\boldsymbol{e}_2 \otimes \boldsymbol{e}_1 + T_{22}\boldsymbol{e}_2 \otimes \boldsymbol{e}_2 \\
&:= \begin{bmatrix} T_{11} & T_{12} & 0 \\ T_{21} & T_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\end{aligned}
\tag{5b}
$$

Equation (5b) makes it evident how the surface strain in (1), written earlier by intuition, may be formally derived from the bulk strain field.

By $(4)_2$, the projection of the identity tensor $\mathbf{I}$ on the surface with outward unit normal $\boldsymbol{n}(\boldsymbol{x})$ can be obtained as

$$\mathbb{I} = \mathbb{P}\mathbf{I}\mathbb{P}, \tag{6}$$

which is called the *surface identity tensor* on the surface and is often used in the context of second-order tensors.

2.2 Differentiation and Integration on a Surface

2.2.1 Surface Gradient, Normal Derivative, and Curvature Tensor

Consider a smooth scalar field $\phi(\boldsymbol{x}) : \Omega_0 \to \mathbb{R}$ and a smooth vector field $\boldsymbol{v}(\boldsymbol{x}) : \Omega_0 \to \mathbb{R}^3$ over the domain Ω_0. By using the gradient operator ∇, their (three-dimensional) gradients are represented by $\nabla\phi(\boldsymbol{x})$ and $\nabla\boldsymbol{v}(\boldsymbol{x})$, respectively. In contrast, the surface gradient operator on a surface with unit normal $\boldsymbol{n}$ is denoted by ∇_s, and together with the projection tensor $\mathbb{P}$ in (2), the *surface gradients* of the two fields can be represented by (Gurtin et al. 1998)

$$\nabla_s\phi = \mathbb{P}\nabla\phi \quad \text{and} \quad \nabla_s\boldsymbol{v} = (\nabla\boldsymbol{v})\mathbb{P}. \tag{7}$$

We now define the *normal derivative* of fields on a surface with outward unit normal $\boldsymbol{n}$. Following the scalar field ϕ and the vector field $\boldsymbol{v}$ in (7), we define their normal derivatives as

$$\frac{\partial\phi}{\partial n} = \boldsymbol{n} \cdot \nabla\phi \quad \text{and} \quad \frac{\partial\boldsymbol{v}}{\partial n} = (\nabla\boldsymbol{v})\boldsymbol{n}. \tag{8}$$

The normal derivative $\dfrac{\partial\phi}{\partial n}$ can be regarded as the rate of change of ϕ in the direction $\boldsymbol{n}$. By (2) and (8), the surface derivatives (7) can be recast as

$$\nabla_s\phi = \nabla\phi - \left(\frac{\partial\phi}{\partial n}\right)\boldsymbol{n} \quad \text{and} \quad \nabla_s\boldsymbol{v} = \nabla\boldsymbol{v} - \left(\frac{\partial\boldsymbol{v}}{\partial n}\right) \otimes \boldsymbol{n}. \tag{9}$$

Equation (9) represents the relation between the surface gradient, the gradient, and the normal derivative. We note that Eqs. (7) and (9) are two (alternative but equivalent) ways to represent the surface gradient.

The *curvature tensor* of a surface with outward unit normal $\boldsymbol{n}$ is then defined as (Gurtin et al. 1998)

$$\mathbb{L} = -\nabla_s \boldsymbol{n}, \tag{10}$$

which is symmetric and hence tangential, that is, $\mathbb{L} = \mathbb{L}^T$ and $\mathbb{L}^T \boldsymbol{n} = \boldsymbol{0}$.

In order to enhance the understanding of these definitions, we now consider some simple examples by referring to Fig. 2. In Cartesian coordinates, the gradients of the scalar and vector fields, respectively, are $\nabla \phi = \frac{\partial \phi}{\partial x_i} \boldsymbol{e}_i$ and $\nabla \boldsymbol{v} = \frac{\partial \boldsymbol{v}}{\partial x_j} \otimes \boldsymbol{e}_j = \frac{\partial v_i}{\partial x_j} \boldsymbol{e}_i \otimes \boldsymbol{e}_j$. The outward unit normal to the flat upper surface is $\boldsymbol{n} = \boldsymbol{e}_3$, and then its projection tensor is $\mathbb{P} = \boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2$ which was mentioned in (3a).

By (7) and the Kronecker delta, the surface gradients on the flat surface in Fig. 2 are

$$\nabla_s \phi = (\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2) \frac{\partial \phi}{\partial x_i} \boldsymbol{e}_i = \frac{\partial \phi}{\partial x_1} \boldsymbol{e}_1 + \frac{\partial \phi}{\partial x_2} \boldsymbol{e}_2 \tag{11a}$$

and

$$\nabla_s \boldsymbol{v} = \left(\frac{\partial \boldsymbol{v}}{\partial x_j} \otimes \boldsymbol{e}_j \right) (\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2) = \frac{\partial \boldsymbol{v}}{\partial x_1} \otimes \boldsymbol{e}_1 + \frac{\partial \boldsymbol{v}}{\partial x_2} \otimes \boldsymbol{e}_2. \tag{11b}$$

By (8), the normal derivatives on the surface in the example are

$$\frac{\partial \phi}{\partial n} = \boldsymbol{e}_3 \cdot \frac{\partial \phi}{\partial x_i} \boldsymbol{e}_i = \frac{\partial \phi}{\partial x_3} \quad \text{and} \quad \frac{\partial \boldsymbol{v}}{\partial n} = \left(\frac{\partial \boldsymbol{v}}{\partial x_j} \otimes \boldsymbol{e}_j \right) \boldsymbol{e}_3 = \frac{\partial \boldsymbol{v}}{\partial x_3}. \tag{12}$$

Using (9), together with (12), we can obtain the same surface gradients as (11). Similarly, using (10), together with (11b), we find the obvious answer that the curvature tensor $\mathbb{L}$ of the flat upper surface in Fig. 2 is

$$\mathbb{L} = -\nabla_s \boldsymbol{e}_3 = -\frac{\partial \boldsymbol{e}_3}{\partial x_1} \otimes \boldsymbol{e}_1 - \frac{\partial \boldsymbol{e}_3}{\partial x_2} \otimes \boldsymbol{e}_2 = \boldsymbol{0}. \tag{13}$$

As another illustration, we may consider the curvature tensor for a spherical surface of radius R. In spherical coordinates $\{r, \theta, \phi\}$ with orthonormal basis $\{\boldsymbol{e}_r, \boldsymbol{e}_\theta, \boldsymbol{e}_\phi\}$, the outward unit normal to the spherical surface is $\boldsymbol{n} = \boldsymbol{e}_r$, whose gradient is $\nabla \boldsymbol{e}_r = r^{-1}(\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta + \boldsymbol{e}_\phi \otimes \boldsymbol{e}_\phi)$. By (7) and (10), the curvature tensor in this example is

$$\mathbb{L} = -\nabla_s \boldsymbol{e}_r = -(\nabla \boldsymbol{e}_r)\mathbb{P} = -\frac{1}{R}(\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta + \boldsymbol{e}_\phi \otimes \boldsymbol{e}_\phi), \tag{14}$$

which is symmetric $\mathbb{L} = \mathbb{L}^T$ and tangential $\mathbb{L}^T \boldsymbol{e}_r = \boldsymbol{0}$.

2.2.2 Surface Divergence, Trace, and Mean Curvature

The *surface divergences* of a vector field $\boldsymbol{v}(\boldsymbol{x})$ and a tensor field $\boldsymbol{T}(\boldsymbol{x})$ are defined as follows:

$$\mathrm{div}_s\boldsymbol{v} = \mathrm{tr}(\nabla_s\boldsymbol{v}) \quad \text{and} \quad \boldsymbol{a} \cdot \mathrm{div}_s\boldsymbol{T} = \mathrm{div}_s(\boldsymbol{T}^T\boldsymbol{a}), \tag{15}$$

where "tr," here and henceforth, denotes the *trace* and $\boldsymbol{a} \in \mathbb{R}^3$ is an arbitrary constant vector. An important identity related to the surface divergence operator is

$$\mathrm{div}_s(\boldsymbol{T}^T\boldsymbol{n}) = \boldsymbol{n} \cdot \mathrm{div}_s\boldsymbol{T} + \boldsymbol{T} \cdot \nabla_s\boldsymbol{n}. \tag{16}$$

For further useful surface identities, the reader is referred to the work (Gurtin and Murdoch 1975b).

The *mean curvature* is then simply

$$\kappa = \frac{1}{2}\mathrm{tr}(\mathbb{L}) = -\frac{1}{2}\mathrm{tr}(\nabla_s\boldsymbol{n}) = -\frac{1}{2}\mathrm{div}_s\boldsymbol{n}. \tag{17}$$

For the same example in Eq. (11b), the corresponding surface divergence is

$$\mathrm{div}_s\boldsymbol{v} = \mathrm{tr}(\nabla_s\boldsymbol{v}) = \mathrm{tr}\left(\frac{\partial(v_i\boldsymbol{e}_i)}{\partial x_1} \otimes \boldsymbol{e}_1 + \frac{\partial(v_i\boldsymbol{e}_i)}{\partial x_2} \otimes \boldsymbol{e}_2\right) = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2}. \tag{18}$$

Similarly, for a second-order tensor $\boldsymbol{T} = T_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j$, its surface gradient is a vector, and its k-th component, by $(15)_2$, can be represented by

$$(\mathrm{div}_s\boldsymbol{T})_k = \boldsymbol{e}_k \cdot \mathrm{div}_s\boldsymbol{T} = \mathrm{div}_s(\boldsymbol{T}^T\boldsymbol{e}_k) = \frac{\partial T_{k1}}{\partial x_1} + \frac{\partial T_{k2}}{\partial x_2}, \tag{19}$$

where the identity $\boldsymbol{T}^T\boldsymbol{e}_k = T_{ki}\boldsymbol{e}_i$ and the example result in (18) are used. Regarding the examples of the curvature tensors (13) and (14), their mean curvatures are 0 and $-R^{-1}$, respectively.

2.2.3 Divergence Theorem for Surfaces

Consider a surface $\mathsf{S}_0 \subset \partial\Omega_0$ with a smooth boundary curve $\partial\mathsf{S}_0$. For a smooth vector $\boldsymbol{u}$ that is tangential on the surface S_0 and a smooth tensor field $\boldsymbol{T}$, the divergence theorem is defined by Gurtin and Murdoch (1975b) and Gurtin et al. (1998)

$$\int_{\mathsf{S}_0} \mathrm{div}_s\boldsymbol{u} = \int_{\partial\mathsf{S}_0} \boldsymbol{u} \cdot \boldsymbol{v}, \qquad \int_{\mathsf{S}_0} \mathrm{div}_s\boldsymbol{T} = \int_{\partial\mathsf{S}_0} \boldsymbol{T}\boldsymbol{v}, \tag{20}$$

where $\boldsymbol{v}$ is the outward unit normal to the boundary curve $\partial\mathsf{S}_0$.

3   Theoretical Framework for Surface Mechanics

With the mathematical preliminaries necessary for surface mechanics described earlier, we can now proceed in a rather standard manner to derive the pertinent governing equations. The original theory by Gurtin and Murdoch (1975b) was derived by employing stress as the primitive concept. We (following Huang and co-workers Huang and Wang 2006; Huang and Sun 2007) favor a variational approach where we take the surface energy as the primitive concept.

3.1   Kinematics

Assume a deformable solid that occupies the domain Ω_0 in the reference configuration shown in Fig. 3. The boundary of the domain is denoted by $\partial\Omega_0$, which can be divided into two parts: the displacement boundary $\partial\Omega_0^u$ and the traction boundary $\partial\Omega_0^t$. Mathematically, $\partial\Omega_0^u \cup \partial\Omega_0^t = \partial\Omega_0$ and $\partial\Omega_0^u \cap \partial\Omega_0^t = \varnothing$.

A material point is denoted by $x \in \Omega_0$. Consider a smooth mapping $y : \Omega_0 \to \mathbb{R}^3$, that is, $y(x) = x + u(x)$. Here u is the displacement vector. The deformation gradient is defined as $F = \nabla x$. The displacement and traction boundary conditions are

$$u = u_0 \quad \text{on } \partial\Omega_0^u \quad \text{and} \quad t = t_0 \quad \text{on } \partial\Omega_0^t. \tag{21}$$
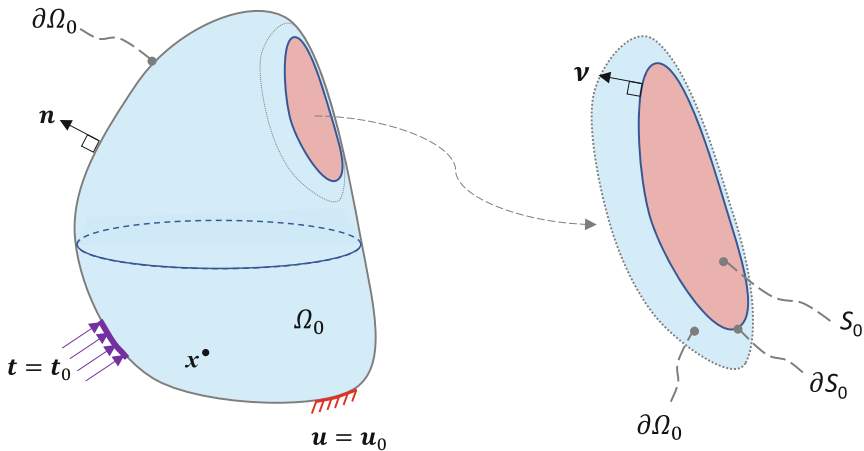
Fig. 3 Schematic of a deformable body that occupies the domain Ω_0 with the boundary $\partial\Omega_0$. The material point is denoted by $x \in \Omega_0$, and the outward unit normal to $\partial\Omega_0$ is represented by n. The displacement boundary is $u = u_0$ while the traction boundary is $t = t_0$. In particular, the surface effect is shown on the surface $S_0 \subset \partial\Omega_0$. On the boundary curve ∂S_0 of the surface S_0, the outward unit normal to ∂S_0 is denoted by v

3.2 Energy Variation and Equations of Equilibrium

We explicitly assume that the surface or the interface does not "slip" from the rest of the material. Gurtin et al. (1998) have dealt with the special case where interfaces may not be coherent and admit jumps in displacements. We do not consider that complexity here. The energy functional of the general system shown in Fig. 3 is then:

$$\mathscr{F}[\boldsymbol{u}; \boldsymbol{x}] = \int_{\Omega_0} \Psi(\nabla \boldsymbol{u}) + \int_{\mathsf{S}_0} \Gamma_s(\nabla \boldsymbol{u}) - \int_{\partial \Omega_0^t} \boldsymbol{t}_0 \cdot \boldsymbol{u}, \qquad (22)$$

where Ψ is the strain energy function per unit volume of the bulk, Γ_s is the surface energy function per unit area of the surface $\mathsf{S}_0 \subset \partial \Omega_0$, and $\boldsymbol{t}_0$ is the dead load applied on the traction boundary $\partial \Omega_0^t$ in the reference configuration. The reference configuration taken here is actually the initial configuration, which is neither subjected to any body force nor tractions. We note that in the reference configuration, there exists the surface stress, which is regarded as "residual." The residual stress field can be described according to the surface energy function Γ_s. To further clarify the surface effects on the energy functional, Huang and Wang (2006, 2013) proposed an extra configuration, a "fictitious stress-free configuration." For further details on such subtleties, the reader is referred to their work.

We now invoke the principle of minimum energy to seek the equilibrium state of the deformed body:

$$\min\{\mathscr{F}[\boldsymbol{u}] : \boldsymbol{u} \in \mathscr{S}\}. \qquad (23)$$

Here the set $\mathscr{S}$ denotes the smooth function space over the domain Ω_0, and the displacement field must satisfy $\boldsymbol{u} = \boldsymbol{u}_0$ on $\partial \Omega_0^u$ in $(21)_1$.

If we assume the state $\boldsymbol{u}$ to be the minimizer of the energy functional in (22), then by the principle of energy minimization (23), we have

$$\mathscr{F}[\boldsymbol{u}] \leq \mathscr{F}[\boldsymbol{u} + \epsilon \boldsymbol{u}_1], \qquad (24)$$

where $\epsilon \in \mathbb{R}$ and $\boldsymbol{u} + \epsilon \boldsymbol{u}_1$ belongs to the set of all kinematically admissible deformations in the neighborhood of the deformation $\boldsymbol{u}$. We only consider small perturbations, so the norm $\|\epsilon \boldsymbol{u}_1\| \ll 1$. By the displacement boundary $(21)_1$, the variation $\boldsymbol{u}_1$ satisfies

$$\boldsymbol{u}_1 = \boldsymbol{0} \quad \text{on } \partial \Omega_0^u. \qquad (25)$$

The inequality (24) leads to the following first and second variation conditions

$$\delta \mathscr{F}[\boldsymbol{u}] = 0 \quad \text{and} \quad \delta^2 \mathscr{F}[\boldsymbol{u}] \geq 0. \qquad (26)$$

In this book chapter, we only limit our attention to the first variation that leads to the equilibrium equations and the natural boundary conditions. The first variation is written as

$$\delta \mathcal{F}[\boldsymbol{u}] := \left. \frac{\mathrm{d}\mathcal{F}[\boldsymbol{u} + \epsilon \boldsymbol{u}_1]}{\mathrm{d}\epsilon} \right|_{\epsilon=0}. \tag{27}$$

Using (27) and the chain rule, the first variation of (22) reads

$$\delta \mathcal{F}[\boldsymbol{u}] = \int_{\Omega_0} \boldsymbol{S} \cdot \nabla \boldsymbol{u}_1 + \int_{\mathsf{S}_0} \mathbb{S} \cdot \nabla \boldsymbol{u}_1 - \int_{\partial \Omega_0^t} \boldsymbol{t}_0 \cdot \boldsymbol{u}_1, \tag{28}$$

where

$$\boldsymbol{S} = \frac{\partial \Psi}{\partial \nabla \boldsymbol{u}} \quad \text{and} \quad \mathbb{S} = \frac{\partial \Gamma_s}{\partial \nabla \boldsymbol{u}}. \tag{29}$$

Here $\mathbb{S}$ in (29) is the first Piola-Kirchhoff surface stress tensor (Gurtin et al. 1998), and $\boldsymbol{S}$ is the first Piola-Kirchhoff bulk stress tensor.

Employing the surface gradient of a vector in $(9)_2$ as well as the property of the surface tensor $\mathbb{S}$, that is, $\mathbb{S}\boldsymbol{n} = \boldsymbol{0}$, the integrand $\mathbb{S} \cdot \nabla \boldsymbol{u}_1$ in (28) can be recast as

$$\mathbb{S} \cdot \nabla \boldsymbol{u}_1 = \mathbb{S} \cdot \nabla_s \boldsymbol{u}_1 + \frac{\partial \boldsymbol{u}_1}{\partial n} \cdot \mathbb{S}\boldsymbol{n} = \mathbb{S} \cdot \nabla_s \boldsymbol{u}_1. \tag{30}$$

In the work by Gurtin et al. (1998), they derived the identity $\mathbb{S}\boldsymbol{n} = \boldsymbol{0}$ on the surface S_0 (see their Eq.(37)) by assuming an arbitrary function $\frac{\partial \boldsymbol{u}_1}{\partial n}$ on the surface S_0. Here we obtain this identity through the definition of a surface tensor $(4)_2$, that is, for an arbitrary surface tensor $\boldsymbol{T}_s = \mathbb{P}\boldsymbol{T}\mathbb{P}$, we have $\boldsymbol{T}_s \boldsymbol{n} = \mathbb{P}\boldsymbol{T}\mathbb{P}\boldsymbol{n} = \mathbb{P}\boldsymbol{T}\boldsymbol{0} = \boldsymbol{0}$ since $\mathbb{P}\boldsymbol{n} = (\boldsymbol{I} - \boldsymbol{n} \otimes \boldsymbol{n})\boldsymbol{n} = \boldsymbol{0}$.

By the identity (16), we further have

$$\mathbb{S} \cdot \nabla_s \boldsymbol{u}_1 = \mathrm{div}_s(\mathbb{S}^T \boldsymbol{u}_1) - \boldsymbol{u}_1 \cdot \mathrm{div}_s\mathbb{S}. \tag{31}$$

Thus, by (31) and the divergence theorem (20), the second integral in (28) becomes

$$\int_{\mathsf{S}_0} \mathbb{S} \cdot \nabla \boldsymbol{u}_1 = \int_{\mathsf{S}_0} \left[\mathrm{div}_s(\mathbb{S}^T \boldsymbol{u}_1) - \boldsymbol{u}_1 \cdot \mathrm{div}_s\mathbb{S} \right] = \int_{\partial \mathsf{S}_0} \boldsymbol{u}_1 \cdot \mathbb{S}\boldsymbol{v} - \int_{\mathsf{S}_0} \boldsymbol{u}_1 \cdot \mathrm{div}_s\mathbb{S}. \tag{32}$$

Similarly, by using the divergence theorem in volume, we obtain

$$\int_{\Omega_0} \boldsymbol{S} \cdot \nabla \boldsymbol{u}_1 = \int_{\Omega_0} \left[\mathrm{div}(\boldsymbol{S}^T \boldsymbol{u}_1) - \boldsymbol{u}_1 \cdot \mathrm{div}\boldsymbol{S} \right] = \int_{\partial \Omega_0} \boldsymbol{u}_1 \cdot \boldsymbol{S}\boldsymbol{n} - \int_{\Omega_0} \boldsymbol{u}_1 \cdot \mathrm{div}\boldsymbol{S}. \tag{33}$$

Substituting (32) and (33) into (28), we have

$$\delta\mathscr{F}[\boldsymbol{u}] = -\int_{\Omega_0} \boldsymbol{u}_1 \cdot \mathrm{div}\boldsymbol{S} + \int_{\partial\Omega_0\backslash\mathsf{S}_0} \boldsymbol{u}_1 \cdot \boldsymbol{S}\boldsymbol{n} + \int_{\mathsf{S}_0} \boldsymbol{u}_1 \cdot (\boldsymbol{S}\boldsymbol{n} - \mathrm{div}_s\mathbb{S})$$
$$+ \int_{\partial\mathsf{S}_0} \boldsymbol{u}_1 \cdot \mathbb{S}\boldsymbol{v} - \int_{\partial\Omega_0^t} \boldsymbol{t}_0 \cdot \boldsymbol{u}_1. \tag{34}$$

It is known that $\partial\Omega_0^u \cup \partial\Omega_0^t = \partial\Omega_0$ and $\partial\Omega_0^u \cap \partial\Omega_0^t = \varnothing$ as well as $\mathsf{S}_0 \subset \partial\Omega_0$. However, the relation between S_0 and $\partial\Omega_0^t$ (or $\partial\Omega_0^u$) is not given before. To simplify the discussion, we regard S_0 as a subset of $\partial\Omega_0^t$, that is, $\mathsf{S}_0 \subset \partial\Omega_0^t$. Thus, the last integral in (34) can be reformulated as

$$-\int_{\partial\Omega_0^t} \boldsymbol{t}_0 \cdot \boldsymbol{u}_1 = -\int_{\partial\Omega_0^t\backslash\mathsf{S}_0} \boldsymbol{t}_0 \cdot \boldsymbol{u}_1 - \int_{\mathsf{S}_0} \boldsymbol{t}_0 \cdot \boldsymbol{u}_1. \tag{35}$$

Also, with $\boldsymbol{u}_1 = \boldsymbol{0}$ on $\partial\Omega_0^u$ in (25), the second integral in (34) reduces to

$$\int_{\partial\Omega_0\backslash\mathsf{S}_0} \boldsymbol{u}_1 \cdot \boldsymbol{S}\boldsymbol{n} = \int_{\partial\Omega_0^t\backslash\mathsf{S}_0} \boldsymbol{u}_1 \cdot \boldsymbol{S}\boldsymbol{n}. \tag{36}$$

Using (35) and (36), the first variation finally becomes

$$\delta\mathscr{F}[\boldsymbol{u}] = -\int_{\Omega_0} \boldsymbol{u}_1 \cdot \mathrm{div}\boldsymbol{S} + \int_{\partial\Omega_0^t\backslash\mathsf{S}_0} \boldsymbol{u}_1 \cdot (\boldsymbol{S}\boldsymbol{n} - \boldsymbol{t}_0) + \int_{\mathsf{S}_0} \boldsymbol{u}_1 \cdot (\boldsymbol{S}\boldsymbol{n} - \mathrm{div}_s\mathbb{S} - \boldsymbol{t}_0)$$
$$+ \int_{\partial\mathsf{S}_0} \boldsymbol{u}_1 \cdot \mathbb{S}\boldsymbol{v}. \tag{37}$$

Since the variation $\boldsymbol{u}_1$ in (37) is arbitrary, the vanishing of the first variation $\delta\mathscr{F}[\boldsymbol{u}] = 0$ and the fundamental lemma of calculus of variations (Courant and Hilbert 1953) leads us to the following set of governing equations

$$\left.\begin{aligned}
\mathrm{div}\boldsymbol{S} &= \boldsymbol{0} & \text{in } \Omega_0, \\
\boldsymbol{S}\boldsymbol{n} &= \boldsymbol{t}_0 & \text{on } \partial\Omega_0^t \setminus \mathsf{S}_0, \\
\boldsymbol{S}\boldsymbol{n} - \mathrm{div}_s\mathbb{S} = \boldsymbol{t}_0, \quad \mathbb{S}\boldsymbol{n} &= \boldsymbol{0} & \text{on } \mathsf{S}_0, \\
\mathbb{S}\boldsymbol{v} &= \boldsymbol{0} & \text{on } \partial\mathsf{S}_0.
\end{aligned}\right\} \tag{38}$$

Here we rewrite the equation $\mathbb{S}\boldsymbol{n} = \boldsymbol{0}$ on S_0 (see the statement above (30)) in (38). Equation (38), together with $(21)_1$ and (29), forms a well-defined boundary value problem. For the readers convenience, we reiterate the notations here: $\boldsymbol{S}$ denotes the first Piola-Kirchhoff stress, $\mathbb{S}$ the first surface Piola-Kirchhoff stress, $\boldsymbol{n}$ the outward unit normal to the surface, $\boldsymbol{v}$ the outward unit normal to the boundary curve, and $\boldsymbol{t}_0$ the applied dead load.

3.3   Constitutive Equations and Elastic Stress Tensors

In (29), we have defined the first Piola-Kirchhoff bulk stress tensor S and the first Piola-Kirchhoff surface stress tensor $\mathbb{S}$ through the partial derivative with respect to the displacement gradient ∇u. By the chain rule, these two first Piola-Kirchhoff stresses can also be defined as

$$S = \frac{\partial \bar{\Psi}(F)}{\partial F} \quad \text{and} \quad \mathbb{S} = \frac{\partial \bar{\Gamma}_s(F)}{\partial F}, \tag{39}$$

where $F = \nabla(x + u) = I + \nabla u$ and

$$\bar{\Psi}(F) = \Psi(\nabla u)\big|_{\nabla u = F - I} \quad \text{and} \quad \bar{\Gamma}_s(F) = \Gamma_s(\nabla u)\big|_{\nabla u = F - I}. \tag{40}$$

By the *frame indifference* in the strain energy functions (Gurtin et al. 2010) and the polar decomposition $F = RU$, $R \in \text{Orth}^+ = \{\text{all rotations}\}$, and U is the right stretch tensor, we have

$$\begin{aligned}
\bar{\Psi}(F) &= \bar{\Psi}(R^T F) = \bar{\Psi}(R^T R U) = \bar{\Psi}(U), \\
\bar{\Gamma}_s(F) &= \bar{\Gamma}_s(R^T F) = \bar{\Gamma}_s(R^T R U) = \bar{\Gamma}_s(U).
\end{aligned} \tag{41}$$

Using the relation $U = \sqrt{U^T U} = \sqrt{U^T R^T R U} = \sqrt{F^T F} = \sqrt{C}$, we can introduce strain energy functions $\hat{\Psi}(C)$ and $\hat{\Gamma}_s(C)$, such that

$$\begin{aligned}
\hat{\Psi}(C) &= \bar{\Psi}(\sqrt{C}) = \bar{\Psi}(U) = \bar{\Psi}(F), \\
\hat{\Gamma}_s(C) &= \bar{\Gamma}_s(\sqrt{C}) = \bar{\Gamma}_s(U) = \bar{\Gamma}_s(F).
\end{aligned} \tag{42}$$

Similarly, by the definition of the Green strain tensor $E = \frac{1}{2}(C - I)$, we can express strain energy functions $\Psi^*(E)$ and $\Gamma_s^*(E)$, such that

$$\begin{aligned}
\Psi^*(E) &= \hat{\Psi}(C) = \bar{\Psi}(U) = \bar{\Psi}(F), \\
\Gamma_s^*(E) &= \hat{\Gamma}_s(C) = \bar{\Gamma}_s(U) = \bar{\Gamma}_s(F).
\end{aligned} \tag{43}$$

In contrast to the two first Piola-Kirchhoff stresses (39), by (43) and $E = \frac{1}{2}(C - I)$, the two second Piola-Kirchhoff stresses are defined as

$$T = 2\frac{\partial \hat{\Psi}(C)}{\partial C} = \frac{\partial \Psi^*(E)}{\partial E} \quad \text{and} \quad \mathbb{T} = 2\frac{\partial \hat{\Gamma}_s(C)}{\partial C} = \frac{\partial \Gamma_s^*(E)}{\partial E}. \tag{44}$$

By (39), (43), (44), and the chain rule, the relations between the first P-K stresses (S, $\mathbb{S}$) and the second P-K stresses (T, $\mathbb{T}$) are (P-K is the abbreviation for the Piola-Kirchhoff stress.)

$$S = FT \quad \text{and} \quad \mathbb{S} = F\mathbb{T}. \tag{45}$$

The relations between the first P-K stresses $(\boldsymbol{S}, \mathbb{S})$ and the Cauchy stresses $(\boldsymbol{\sigma}, \sigma^s)$ are

$$\boldsymbol{S} = (\det \boldsymbol{F})\boldsymbol{\sigma}\,\boldsymbol{F}^{-T} \quad \text{and} \quad \mathbb{S} = (\det \boldsymbol{F})\sigma^s\boldsymbol{F}^{-T}. \tag{46}$$

3.4 Linearized Bulk and Surface Stresses and Constitutive Choice

A peculiarity of considering surface effects is the perceived presence of residual stresses – the surface tension-like quantity in solids is precisely a residual stress state. This becomes evident if we linearize the energy functions around a reference configuration which is not stress-free.

For small deformation, the Cauchy-Green tensor $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F} = (\mathbf{I} + \nabla \boldsymbol{u})^T(\mathbf{I} + \nabla \boldsymbol{u})$ can be reduced to $\boldsymbol{C} = \mathbf{I} + \nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T$, $|\nabla \boldsymbol{u}| \ll 1$, by dropping the higher-order terms $o(|\nabla \boldsymbol{u}|)$. Thus, the strain tensor $\boldsymbol{E} = \frac{1}{2}(\boldsymbol{C} - \mathbf{I})$ in (44) may be approximated by the infinitesimal strain:

$$\boldsymbol{E} = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T), \tag{47}$$

Using Taylor series, $(44)_1$ gives

$$\boldsymbol{T} = \left.\frac{\partial \Psi^*(\boldsymbol{E})}{\partial \boldsymbol{E}}\right|_{\boldsymbol{E}=0} + \left.\frac{\partial^2 \Psi^*(\boldsymbol{E})}{\partial \boldsymbol{E}^2}\right|_{\boldsymbol{E}=0} \cdot \boldsymbol{E} + o(|\boldsymbol{E}|), \tag{48}$$

where $o(|\boldsymbol{E}|)$ denotes the higher-order terms.

Since the reference configuration is assumed to be stress-free for the bulk portion of the material, it follows that $\left.\frac{\partial \Psi^*(\boldsymbol{E})}{\partial \boldsymbol{E}}\right|_{\boldsymbol{E}=0} = \boldsymbol{0}$ and the elasticity tensor is

$$\mathbb{C} := \left.\frac{\partial^2 \Psi^*(\boldsymbol{E})}{\partial \boldsymbol{E}^2}\right|_{\boldsymbol{E}=0}. \tag{49}$$

Thus, as usual, the linearized small-deformation stress-strain relation of bulk portion of the material is

$$\boldsymbol{T} = \mathbb{C}\boldsymbol{E}. \tag{50}$$

Consider the first P-K stresses $\boldsymbol{S} = \boldsymbol{F}\boldsymbol{T}$ in $(45)_1$ and the Cauchy stresses $\boldsymbol{\sigma} = (\det \boldsymbol{F})^{-1}\boldsymbol{S}\boldsymbol{F}^T$ in $(46)_1$ of bulk materials at small deformations. By the relation (50) and the infinitesimal strain (47), it is easy to show that to a first order

$$\boldsymbol{S} = \boldsymbol{T} = \boldsymbol{\sigma}. \tag{51}$$

The equivalence of the first P-K, the second P-K, and the Cauchy stresses in (51), however, does not hold for surface stresses due to the existence of the residual stress. This is rather important to note. In linearized elasticity (without residual stresses), it is usual to ignore the distinction between the various stress measures. For surface

elasticity, even in the linearized case, we must take cognizance of the different interpretations of the various stress measures. The first P-K stress measure is the most useful since it represents the force per unit *referential* area and is likely to be the quantity controlled in traction-controlled experiments (as opposed to Cauchy traction).

From previous works (see Eq. (40) in the work of Gurtin et al. 1998 and Eq. (7.9) in the work of Gurtin and Murdoch 1975b), we can show that the second P-K surface stress $\mathbb{T}$ in $(44)_2$ and the first P-K surface stress $\mathbb{S}$ in $(45)_2$ can be recast as

$$\mathbb{T} = \frac{\partial \Gamma_s{}^*(\boldsymbol{E}_s)}{\partial \boldsymbol{E}_s}, \quad \mathbb{S} = (\mathbb{I} + \mathbb{P}\nabla_s \boldsymbol{u})\mathbb{T}, \tag{52}$$

where $\boldsymbol{E}_s$, in contrast to the infinitesimal strain (47), is the infinitesimal surface strain

$$\boldsymbol{E}_s = \mathbb{P}\boldsymbol{E}\mathbb{P} = \frac{1}{2}[\mathbb{P}\nabla_s \boldsymbol{u} + (\mathbb{P}\nabla_s \boldsymbol{u})^T]. \tag{53}$$

Using the Taylor series for $(52)_1$, for small deformation, we have

$$\mathbb{T}(\boldsymbol{E}_s) = \left.\frac{\partial \Gamma_s{}^*(\boldsymbol{E}_s)}{\partial \boldsymbol{E}_s}\right|_{\boldsymbol{E}_s=\boldsymbol{0}} + \left.\frac{\partial^2 \Gamma_s{}^*(\boldsymbol{E}_s)}{\partial \boldsymbol{E}_s^2}\right|_{\boldsymbol{E}_s=\boldsymbol{0}} \cdot \boldsymbol{E}_s + o(|\boldsymbol{E}_s|). \tag{54}$$

With an appropriate choice of an elastic constitutive law (i.e., specification of $\Gamma_s{}^*$ and Ψ^*), we now have all the governing equations and can solve the pertinent boundary value problems of physical interest. Analytical solutions are rather hard to come by for the anisotropic case, and by far, most problems solved in the literature have been restricted to isotropic continua.

For isotropic linear elastic materials incorporating the residual surface stress, we have (Gurtin and Murdoch 1975a, b)

$$\tau_0\mathbb{I} := \left.\frac{\partial \Gamma_s{}^*(\boldsymbol{E}_s)}{\partial \boldsymbol{E}_s}\right|_{\boldsymbol{E}_s=\boldsymbol{0}}, \quad \mathbb{C}_s := \left.\frac{\partial^2 \Gamma_s{}^*(\boldsymbol{E}_s)}{\partial \boldsymbol{E}_s^2}\right|_{\boldsymbol{E}_s=\boldsymbol{0}}, \tag{55}$$

where $\tau_0\mathbb{I}$ is the residual stress and the surface elasticity tensor $\mathbb{C}_s$ gives

$$\mathbb{C}_s[\boldsymbol{E}_s] = \lambda_0 \mathrm{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s \tag{56}$$

with surface elastic moduli λ_0 and μ_0.

By (55) and (56), a general linearized constitutive law for isotropic linear materials in terms of the second P-K surface stress $\mathbb{T}$ in (54) can be written as

$$\mathbb{T}(\boldsymbol{E}_s) = \tau_0\mathbb{I} + \lambda_0 \mathrm{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s. \tag{57}$$

For solving actual boundary value problems (including the ones we present later in this chapter), we will need the first P-K stress tensor. To do so, we take cognizance

of the relation between the two P-K stresses. By $(52)_2$ and (57), the first P-K surface stress $\mathbb{S}$ becomes

$$\mathbb{S}(\nabla_s \boldsymbol{u}) = \tau_0 \mathbb{I} + \lambda_0 \text{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s + \tau_0 \nabla_s \boldsymbol{u}, \tag{58}$$

where the higher-order terms $|(\nabla_s \boldsymbol{u})\text{tr}(\boldsymbol{E}_s)|$ and $|(\nabla_s \boldsymbol{u})\boldsymbol{E}_s|$ are omitted. Here and henceforth, the difference between $\nabla_s \boldsymbol{u}$ and $\mathbb{P}\nabla_s \boldsymbol{u}$ is not specified for simplicity.

By $(46)_2$ and a similar argument as (52), the Cauchy surface stress σ^s and the first P-K stress $\mathbb{S}$ in (58) have the relation

$$\sigma^s = [\det(\mathbb{I} + \nabla_s \boldsymbol{u})]^{-1}\mathbb{S}(\mathbb{I} + \nabla_s \boldsymbol{u})^T. \tag{59}$$

For small deformation, $[\det(\mathbb{I} + \nabla_s \boldsymbol{u})]^{-1} = 1 - \text{tr}(\nabla_s \boldsymbol{u}) = 1 - \text{tr}(\boldsymbol{E}_s)$; thus, σ^s in (59) can be recast as

$$\begin{aligned}
\sigma^s &= [1 - \text{tr}(\boldsymbol{E}_s)][\tau_0 \mathbb{I} + \lambda_0 \text{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s + \tau_0 \nabla_s \boldsymbol{u}](\mathbb{I} + \nabla_s \boldsymbol{u}^T) \\
&= [\tau_0 \mathbb{I} + \lambda_0 \text{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s + \tau_0 \nabla_s \boldsymbol{u} - \tau_0 \text{tr}(\boldsymbol{E}_s)\mathbb{I}](\mathbb{I} + \nabla_s \boldsymbol{u}^T) \\
&= \tau_0 \mathbb{I} + \lambda_0 \text{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s + \tau_0 \nabla_s \boldsymbol{u} - \tau_0 \text{tr}(\boldsymbol{E}_s)\mathbb{I} + \tau_0 \nabla_s \boldsymbol{u}^T \\
&= \tau_0 \mathbb{I} + (\lambda_0 - \tau_0)\text{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s + \tau_0(\nabla_s \boldsymbol{u} + \nabla_s \boldsymbol{u}^T)
\end{aligned} \tag{60}$$

by dropping the higher-order terms related to the product between these terms $\text{tr}(\boldsymbol{E}_s)$, $\boldsymbol{E}_s$, $\nabla_s \boldsymbol{u}$, and $\nabla_s \boldsymbol{u}^T$.

Using the definition of surface strain in (53) and (60) finally becomes

$$\sigma^s = \tau_0 \mathbb{I} + (\lambda_0 - \tau_0)\text{tr}(\boldsymbol{E}_s)\mathbb{I} + 2(\mu_0 + \tau_0)\boldsymbol{E}_s = \tau_0 \mathbb{I} + \lambda_s \text{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_s \boldsymbol{E}_s. \tag{61}$$

Here

$$\lambda_s = \lambda_0 - \tau_0 \quad \text{and} \quad \mu_s = \mu_0 + \tau_0 \tag{62}$$

are the Lamé constants of the surface.

In contrast to the equivalence (51), $\mathbb{S}$ in (58), $\mathbb{T}$ in (57), and σ^s in (61) are not equivalent if the residual stress is nonzero $\tau_0 \neq 0$, namely:

$$\mathbb{S} \neq \mathbb{T} \neq \sigma^s \qquad \text{for} \quad \tau_0 \neq 0. \tag{63}$$

In the literature, sometimes (including the work by the corresponding author), the distinction between the various forms of the surface stress has often been blurred. For the most part, this does not lead to qualitative differences, but care must be exercised when solving surface elasticity problem to ascertain which precise stress measure and constitutive parameters are being deployed. As self-evident, there is a distinction between (λ_s, μ_s) and (λ_0, μ_0).

If we set the surface Lamé constants (λ_s, μ_s) to zero in (62), we obtain $\lambda_0 = \tau_0$ and $\mu_0 = -\tau_0$, and then the result in (58) is what is often called the *surface tension* (in the reference configuration) (Gurtin and Murdoch 1975a)

$$\mathbb{S} = \tau_0[1 + \text{tr}(\boldsymbol{E}_s)]\mathbb{I} - 2\tau_0\boldsymbol{E}_s + \tau_0\nabla_s\boldsymbol{u}. \tag{64}$$

This is to be contrasted with the expression in the current configuration where the Cauchy surface tension from (61) by setting (λ_s, μ_s) to zero will be in the form of isotropic "pressure"

$$\sigma^s = \tau_0\mathbb{I}. \tag{65}$$

4   Illustrative Examples

In this section we choose three illustrative examples that highlight both the use of the surface elasticity theory as well as provide insights into the physical consequences of surface energy at the nanoscale. These examples are inspired from Altenbach et al. (2013), Murdoch (2005), and Sharma et al. (2003) although, to be consistent with our own style (presented in the preceding sections), we have modified them slightly. Germane to the study of analytical study of nanostructures, we also note parallel developments in the literature on the so-called surface Cauchy-Born rule and numerical methods (Park et al. 2006).

4.1   Young's Modulus of a Nano-rod Considering Surface Effects

This example addresses how the elastic modulus of an isotropic nano-rod alters due to the influence of surface energy. This particular problem is inspired from the work by Altenbach et al. (2013) although there are some minor differences in our solution. Consider a circular cylinder (rod) with radius R_0 whose axis coincides with the $\boldsymbol{e}_z$ direction. To interrogate the elastic response, we assume that the rod is under uniaxial tension and the coordinate system used for this problem is cylindrical coordinate with basis $(\boldsymbol{e}_r, \boldsymbol{e}_\theta, \boldsymbol{e}_z)$ as shown in Fig. 4.

Fig. 4 A schematic of a nano-rod under a uniaxial tension. Surface effects are only considered on the radial surface $\partial\Omega_0$, and edge effects are ignored. Force $\boldsymbol{F}$ is exerted in the $\boldsymbol{e}_z$ direction
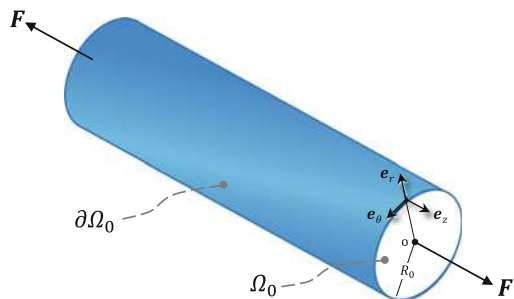
For isotropic elastic materials without surface effects, the uniaxial tension along the axial direction admits a homogeneous deformation. To simplify the discussion, we assume the deformation here incorporating the surface effect is also homogeneous, namely:

$$\boldsymbol{u} = Ar\boldsymbol{e}_r + Bz\boldsymbol{e}_z, \tag{66}$$

where A and B are constants. The assumed displacement (66) can be used for either small or finite deformation.

For the radial surface with outward unit normal $\boldsymbol{n} = \boldsymbol{e}_r$, the projection tensor is $\mathbb{P} = \boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta + \boldsymbol{e}_z \otimes \boldsymbol{e}_z$. By (66), the displacement gradient $\nabla \boldsymbol{u}$, the strain tensor $\boldsymbol{E} = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T)$ for small deformation, the surface displacement gradient $\nabla_s \boldsymbol{u} = (\nabla \boldsymbol{u})\mathbb{P}$, and the surface strain tensor $\boldsymbol{E}_s = \mathbb{P}\boldsymbol{E}\mathbb{P}$ for small deformation become

$$\begin{aligned}
\nabla \boldsymbol{u} &= \boldsymbol{E} = A(\boldsymbol{e}_r \otimes \boldsymbol{e}_r + \boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta) + B\boldsymbol{e}_z \otimes \boldsymbol{e}_z, \\
\nabla_s \boldsymbol{u} &= \boldsymbol{E}_s = A\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta + B\boldsymbol{e}_z \otimes \boldsymbol{e}_z.
\end{aligned} \tag{67}$$

By the constitutive law of isotropic linear materials and $(67)_1$, we have

$$\boldsymbol{S} = \lambda \mathrm{tr}(\boldsymbol{E})\mathbf{I} + 2\mu \boldsymbol{E} := \begin{bmatrix} S_{rr} & 0 & 0 \\ 0 & S_{\theta\theta} & 0 \\ 0 & 0 & S_{zz} \end{bmatrix}, \tag{68a}$$

where λ and μ are Lamé constants of bulk materials and

$$S_{rr} = S_{\theta\theta} = \lambda(2A + B) + 2\mu A, \quad S_{zz} = \lambda(2A + B) + 2\mu B. \tag{68b}$$

In addition, by the constitutive law (58) and by setting $\tau_0 = 0$, together with $(67)_2$, we have

$$\mathbb{S} = \lambda_0 \mathrm{tr}(\boldsymbol{E}_s)\mathbb{I} + 2\mu_0 \boldsymbol{E}_s := \begin{bmatrix} \mathbb{S}_{\theta\theta} & 0 \\ 0 & \mathbb{S}_{zz} \end{bmatrix}, \tag{69a}$$

where

$$\mathbb{S}_{\theta\theta} = \lambda_0(A + B) + 2\mu_0 A, \quad \mathbb{S}_{zz} = \lambda_0(A + B) + 2\mu_0 B. \tag{69b}$$

Using the Young-Laplace equation $(38)_3$, without external load $\boldsymbol{t}_0 = \boldsymbol{0}$ on the radial surface ($r = R_0$) with unit normal $\boldsymbol{e}_r$, we have the equality $\boldsymbol{S}\boldsymbol{e}_r = \mathrm{div}_s \mathbb{S}$. By (68a), the identity (16), and $\mathbb{S}^T \boldsymbol{e}_r = \boldsymbol{0}$ in $(38)_3$, we can obtain

$$S_{rr} = \boldsymbol{e}_r \cdot \boldsymbol{S}\boldsymbol{e}_r = \boldsymbol{e}_r \cdot \mathrm{div}_s \mathbb{S} = -\mathbb{S} \cdot \nabla_s \boldsymbol{e}_r. \tag{70}$$

Since $\nabla_s \boldsymbol{e}_r = (\nabla \boldsymbol{e}_r)\mathbb{P} = \left(\frac{1}{r}\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta\right)(\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta + \boldsymbol{e}_z \otimes \boldsymbol{e}_z) = \frac{1}{r}\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta$, (70) implies

$$S_{rr} = -\frac{\mathbb{S}_{\theta\theta}}{R_0}. \tag{71}$$

By (68b), (69b), and (71) be written as

$$\lambda(2A + B) + 2\mu A = -\frac{\lambda_0(A + B) + 2\mu_0 A}{R_0}, \tag{72}$$

which gives the ratio

$$\frac{A}{B} = -\frac{\lambda + \lambda_0/R_0}{2\lambda + 2\mu + (\lambda_0 + 2\mu_0)/R_0}. \tag{73}$$

By (67) and (68b), the Young modulus for uniaxial tension is

$$E^{\text{rod}} = \frac{S_{zz}}{E_{zz}} = \frac{\lambda(2A + B) + 2\mu B}{B} = \lambda\left(1 + 2\frac{A}{B}\right) + 2\mu, \tag{74}$$

where the ratio A/B is given by (73). By the equilibrium of the rod in the axial direction (Altenbach et al. 2013), the effective Young's modulus can be defined as $E^{\text{eff}} = (S_{zz} + \frac{2}{R_0}\mathbb{S}_{zz})/E_{zz} = \lambda\left(1 + 2\frac{A}{B}\right) + 2\mu + \frac{2}{R_0}[\lambda_0(1 + \frac{A}{B}) + 2\mu_0]$.

This simple example makes clear that the effective or apparent elastic response of nanostructures becomes size-dependent as a result of surface energy effects and that with smaller R_0, the effective elastic modulus may become significantly different than its bulk value. We remark that if surface effects are ignored, that is, $\lambda_0 = \mu_0 = 0$ and the ratio $A/B = -\lambda/(2\lambda + 2\mu)$ in (73), then the Young modulus in (74) becomes $E^{\text{rod}} = \mu(3\lambda + 2\mu)/(\lambda + \mu)$–which is essentially the relation between bulk Young's modulus and the Lamé constants.

4.2   Influence of Surface Effects on the Thermoelastic State of a Ball

We consider the equilibrium state of a spherical ball (radius R_0) in vacuum undergoing thermal expansion – following Murdoch (2005). Only traction boundary conditions need to be considered here; hence, $(21)_1$ can be omitted. Moreover, $(38)_2$ and $(38)_4$ are also omitted since the entire surface of the sphere is considered, that is, $\mathsf{S}_0 = \partial\Omega_0 = \partial\Omega_0^t$.

We now proceed to solve the reduced boundary value problem that consists of $(38)_1$, $(38)_3$, and (29). As before, we assume the ball material to be isotropic and that the ball is in its natural, stress-free state in the reference configuration. Hence, we assume constitutive equations for the bulk and surface as follows:

$$S = \lambda_b \mathrm{tr}(E)I + 2\mu_b E - \alpha \Delta T I, \tag{75}$$

$$\mathbb{S} = \tau_0 \mathbb{I} + (\lambda_s + \tau_0)\mathrm{tr}(E_s)\mathbb{I} + 2(\mu_s - \tau_0)E_s + \tau_0(\nabla_s u) - \alpha_0 \Delta T \mathbb{I}, \tag{76}$$

where S is the first P-K stress field in the reference configuration, λ_b and μ_b are Lamé constants for the ball material, ΔT is the temperature difference, α is the coefficient of thermal expansion, and α_0 is the thermal expansion coefficient of the surface.

In this example we use spherical coordinates with basis (e_r, e_θ, e_ϕ) with the origin at the center of the ball. For the surface with outward unit normal $n = e_r$, the projection tensor is $\mathbb{P} = e_\theta \otimes e_\theta + e_\phi \otimes e_\phi$. Given that the problem is spherically symmetric, the form of the displacement field can be assumed to be

$$u = u_r(r)e_r. \tag{77}$$

Then the displacement gradient, the strain tensor, the surface gradient of the displacement, and the surface strain in (75) and (76) can be written, similar to (67), as

$$\nabla u = E = \frac{\partial u_r}{\partial r}e_r \otimes e_r + \frac{u_r}{r}(e_\theta \otimes e_\theta + e_\phi \otimes e_\phi),$$
$$\nabla_s u = E_s = \frac{u_r}{r}(e_\theta \otimes e_\theta + e_\phi \otimes e_\phi). \tag{78}$$

From (78), (75), and (76), we have

$$S = \lambda_b\left(\frac{\partial u_r}{\partial r} + \frac{2u_r}{r}\right)I + 2\mu_b\left(\frac{\partial u_r}{\partial r}e_r \otimes e_r + \frac{u_r}{r}(e_\theta \otimes e_\theta + e_\phi \otimes e_\phi)\right) - \alpha \Delta T I,$$

$$\mathbb{S} = \tau_0\mathbb{I} + \lambda_s\left(\frac{2u_r}{r}\right)\mathbb{I} + (2\mu_s + \tau_0)\left(\frac{u_r}{r}\right)(e_\theta \otimes e_\theta + e_\phi \otimes e_\phi) - \alpha_0 \Delta T \mathbb{I}. \tag{79}$$

Substituting (79)$_1$ into the equilibrium equation (38)$_1$, together with the divergence of a tensor in spherical coordinates, we obtain

$$r^2\frac{\partial^2 u_r}{\partial r^2} + 2r\frac{\partial u_r}{\partial r} - 2u_r = 0. \tag{80}$$

The general solution of (80) is $u_r(r) = Ar + Br^{-2}$, where A and B are constants to be determined by the boundary conditions. The displacement at the origin must vanish for the field to be bounded, i.e., and consequently $B = 0$ and then

$$u_r(r) = Ar. \tag{81}$$

Thus, the stresses (79) are reduced to

$$S = [A(3\lambda_b + 2\mu_b) - \alpha \Delta T]\mathbf{I}, \quad \mathbb{S} = [\tau_0 + A(2\lambda_s + 2\mu_s + \tau_0) - \alpha_0 \Delta T]\mathbb{I}. \quad (82)$$

We must now deploy the boundary condition in $(38)_3$. Since the ball surface is traction-free, that is, $\mathbf{t}_0 = \mathbf{0}$, $(38)_3$ can be reduced to $\mathbf{S}\mathbf{e}_r = \mathrm{div}_s\mathbb{S}$. By (82), we can further simplify this to

$$S_{rr} = \mathbf{e}_r \cdot \mathbf{S}\mathbf{e}_r = \mathbf{e}_r \cdot \mathrm{div}_s\mathbb{S} \quad \text{at } r = R_0. \quad (83)$$

By using the identity (16), $(82)_2$, $(38)_3$, and (14), we finally obtain

$$S_{rr} = \mathbf{e}_r \cdot \mathrm{div}_s\mathbb{S} = -\mathbb{S} \cdot \nabla_s \mathbf{e}_r = -\mathbb{S} \cdot \left\{ \frac{1}{r}(\mathbf{e}_\theta \otimes \mathbf{e}_\theta + \mathbf{e}_\phi \otimes \mathbf{e}_\phi) \right\} = -\frac{2\mathbb{S}_{\theta\theta}}{R_0}. \quad (84)$$

By (82) and (84), we have

$$3\lambda_b A + 2\mu_b A - \alpha \Delta T = -\frac{2}{R_0}\{\tau_0(1 + A) + 2A(\lambda_s + \mu_s) - \alpha_0 \Delta T\}. \quad (85)$$

By defining the *surface modulus* K_s as

$$K_s = 2(\lambda_s + \mu_s), \quad (86)$$

(85) yields the solution of A in the displacement (81), namely:

$$A = \frac{-2\tau_0/R_0}{(3\lambda_b + 2\mu_b + 2K_s/R_0 + 2\tau_0/R_0)} + \frac{(\alpha + 2\alpha_0/R_0)\Delta T}{(3\lambda_b + 2\mu_b + 2K_s/R_0 + 2\tau_0/R_0)}. \quad (87)$$

This example, as can be noted from (87), nicely shows how a positive surface residual stress τ_0 may hinder the thermal expansion in a size-dependent manner.

4.3   Effect of Residual Stress of Surfaces on Elastic State of Spherical Inclusion

The solution of an embedded inclusion in another material is a canonical problem in classical solid mechanics (known as Eshelby's inclusion problem) and has been applied to situations as diverse as phase transformation to effective properties of composites. Accordingly, in this section, we consider the problem of spherical inclusion with radius R_0 but incorporating surface effects. The solution here is a slight modification of the work by Sharma et al. (2003). We use spherical coordinates with basis $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$ with the origin at the center of the sphere in this problem. We assume that a far field stress is exerted on the matrix as follows:

$$S_{rr}\big|_{r\to\infty} = S^\infty. \quad (88)$$

To make analytical progress, we further stipulate that the inclusion, the matrix, and the interface are isotropic and that the inclusion is in its natural state in the reference configuration. The constitutive equations for the bulk are as follows:

$$S = \lambda_v \text{tr}(E)\mathbf{I} + 2\mu_v E, \tag{89}$$

where S is the first Piola-Kirchhoff stress and λ_v and μ_v are Lamé constants of the material corresponding to either the inclusion(I) or the matrix(M). For the surface we use the constitutive equation defined in (58), i.e.:

$$\mathbb{S} = \tau_0 \mathbb{I} + (\lambda_s + \tau_0)\text{tr}(E_s)\mathbb{I} + 2(\mu_s - \tau_0)E_s + \tau_0(\nabla_s u). \tag{90}$$

Like the previous example, this problem also has spherical symmetry and accordingly the displacement vector is of the form

$$\boldsymbol{u} = u_r(r)\boldsymbol{e}_r. \tag{91}$$

Therefore, the displacement gradient and the strain are similar to (78). Hence, the stress in the bulk and on the surface stress can be expressed as

$$S = \lambda_b \left(\frac{\partial u_r}{\partial r} + \frac{2u_r}{r} \right) \mathbf{I} + 2\mu_b \left(\frac{\partial u_r}{\partial r} \boldsymbol{e}_r \otimes \boldsymbol{e}_r + \frac{u_r}{r}(\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta + \boldsymbol{e}_\phi \otimes \boldsymbol{e}_\phi) \right),$$

$$\mathbb{S} = \tau_0 \mathbb{I} + \lambda_s \left(\frac{2u_r}{r} \right) \mathbb{I} + (2\mu_s + \tau_0)\left(\frac{u_r}{r} \right)(\boldsymbol{e}_\theta \otimes \boldsymbol{e}_\theta + \boldsymbol{e}_\phi \otimes \boldsymbol{e}_\phi). \tag{92}$$

We have used (38) in spherical coordinates. The boundary value problem that consists of $(38)_1$–$(38)_3$ becomes

$$\left. \begin{aligned} r^2 \frac{\partial^2 u_r}{\partial r^2} + 2r\frac{\partial u_r}{\partial r} - 2u_r = 0 \qquad & r \leqslant R_0, \\[2mm] r^2 \frac{\partial^2 u_r}{\partial r^2} + 2r\frac{\partial u_r}{\partial r} - 2u_r = 0 \qquad & r > R_0, \\[2mm] [\![S_{rr}]\!] = \frac{2\mathbb{S}_{\theta\theta}}{r} \qquad & r = R_0, \\[2mm] S_{rr}\big|_{r\to\infty} = S^\infty \qquad & r \to \infty. \end{aligned} \right\} \tag{93}$$

The "$[\![\cdot]\!]$" denotes the jump across the interface. Similar to (80), the general solutions of $(93)_1$ and $(93)_2$ are

$$u_r(r) = \begin{cases} A\,r + B\,r^{-2} & r \leqslant R_0, \\ C\,r + D\,r^{-2} & r > R_0, \end{cases} \tag{94}$$

where A, B, C, and D are constant. Since $u_r(0) = 0$, $B = 0$. Substituting $(94)_2$ into $(92)_1$, the far field $(93)_4$ gives $C = S^\infty/(3K_M)$, where $K_M = \lambda_M + 2\mu_M/3$. Using the continuity of the displacement at the interface, that is, $[\![u_r]\!] = 0$ at $r = R_0$, and

the Young-Laplace equation $(93)_3$, we obtain two algebraic equations of A and D. The routine calculation is not shown here and we just list the final results. Thus, the displacement (94) is obtained as

$$
u(r) = \begin{cases} \alpha r & r \leqslant R_0, \\[2ex] E^\infty r + (\alpha - E^\infty) \dfrac{R_0^3}{r^2} & r > R_0, \end{cases} \tag{95}
$$

where $E^\infty = S^\infty/(3K_M)$, $\alpha := \dfrac{(3K_M + 4\mu_M)E^\infty - 2\tau_0/R_0}{4\mu_M + 3K_I + 2K_s/R_0 + 2\tau_0/R_0}$, $K_I = \lambda_I + 2\mu_I/3$ is the inclusion bulk modulus, and $K_s = 2(\lambda_s + \mu_s)$ is defined as the surface modulus.

In particular, by setting $\lambda_I = 0$ and $\mu_I = 0$, we obtain the much-studied case of a void in a solid. The displacement $(95)_2$ for this special case is

$$
u(r) = E^\infty r + \beta \frac{R_0^3}{r^2} \qquad r > R_0, \tag{96}
$$

where $\beta := \dfrac{(3K_M + 4\mu_M)E^\infty - 2\tau_0/R_0}{4\mu_M + 2K_s/R_0 + 2\tau_0/R_0} - E^\infty$ and $E^\infty = S^\infty/(3K_M)$. Thus, the bulk stress $(92)_1$ in the matrix $(r > R_0)$ is

$$
\boldsymbol{S} = S^\infty \mathbf{I} + 2\mu_M \beta \frac{R_0^3}{r^3} \left(\mathbf{I} - 3\boldsymbol{e}_r \otimes \boldsymbol{e}_r \right). \tag{97}
$$

If there is no far stress field, $S^\infty = E^\infty = 0$, the bulk stress $\boldsymbol{S}$ becomes

$$
\boldsymbol{S}^0 = 2\mu_M \left(\frac{-2\tau_0/R_0}{4\mu_M + 2K_s/R_0 + 2\tau_0/R_0} \right) \frac{R_0^3}{r^3} \left(\mathbf{I} - 3\boldsymbol{e}_r \otimes \boldsymbol{e}_r \right). \tag{98}
$$

Assume a nonzero far stress field, $S^\infty \neq 0$. By (97) and (98), we may then define the stress concentration factor at $r \to R_0$ as

$$
S.C. = \left. \frac{S_{\theta\theta} - S_{\theta\theta}^0}{S^\infty} \right|_{r \to R_0} = 1 + \frac{1}{2} \left(\frac{1 - 2\left(K_s + \tau_0\right)/(3K_M R_0)}{1 + K_s/(2\mu_M R_0) + \tau_0/(2\mu_M R_0)} \right). \tag{99}
$$

In the absence of surface energy effects, i.e., $\tau_0 = K_s = 0$ in (99), we obtain a stress concentration factor of 1.5 which is the well-known classical elasticity result for a spherical void under hydrostatic stress. We have graphically plotted the results in Fig. 5 that illustrates the qualitative behavior of how stress concentration on a void alters due to size and the surface elasticity modulus (Interestingly, the surface elasticity modulus can have negative values as shown via atomistic simulations by Miller and Shenoy 2000.).
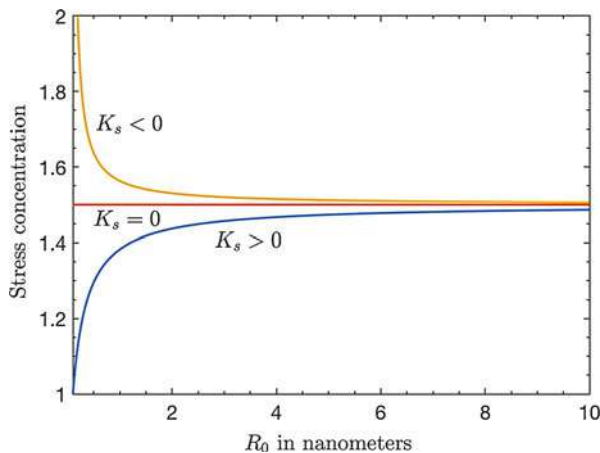
Fig. 5 Stress concentration on a spherical void under hydrostatic stress is plotted with respect to the size of the void for three different surface modulus K_s. For coarser void size, the stress concentration asymptotically approaches the bulk value of $3/2$. For the plot, we have chosen properties of Aluminum as bulk, with ($\lambda_M = 62.29$ GPa, $\mu_M = 36.71$ GPa). The surface properties are the following: *Case* (a) considering positive surface modulus $K_s = 2(\lambda_s + \mu_s)$ by using $\lambda_s = 6.842\,N/m$, $\mu_s = -0.3755\,N/m$, and $\tau_0 = 0$; *Case* (b) without considering surface effects; and $K_s = 0$; *Case* (c) considering negative surface modulus $K_s = 2(\lambda_s + \mu_s)$ by using $\lambda_s = 3.48912\,N/m$, $\mu_s = -6.2178\,N/m$, and $\tau_0 = 0$. (The surface and material properties are taken from Miller and Shenoy 2000; Cammarata et al. 2000)

5   Perspectives on Future Research

Despite extensive work on surface elasticity, some aspects of this field are still somewhat understudied and represent avenues for future exploration. We briefly articulate them below:

- As well-motivated by Steigmann and Ogden (1999, 1997), under certain circumstances, the dependence of surface energy on curvature must be accounted for. This was recently explored by Fried and Todres (2005), who examined the effect of the curvature-dependent surface energy on the wrinkling of thin films, and Chhapadia et al. (2011) who (using both atomistics and a continuum approach) explained certain anomalies in the bending behavior of nanostructures. However, due to the complexity of the Steigmann-Ogden curvature-dependent surface elasticity, relatively few works exist on this topic.
- Intriguing recent experiments by Style and co-workers (Style et al. 2013, 2017) on capillarity and liquid inclusions in soft solids have revealed that the pertinent size effects due to surface effects may be observed at *micron* length scales (in contrast to the nanoscale for hard materials). This represents an important future direction and requires the use of *nonlinear* surface elasticity due to the need to account for the inevitable large deformations in soft matter. Arguably, the

study of capillarity in soft matter will also require the development and use of numerical methods cf. Henann and Bertoldi (2014).

- Finally, the literature on coupling of capillarity with electrical and magnetic fields is quite sparse.

References

Altenbach H, Eremeyev VA, Morozov NF (2013) Mechanical properties of materials considering surface effects. In: IUTAM symposium on surface effects in the mechanics of nanomaterials and heterostructures. Springer, Berlin Heidelberg, pp 105–115

Biria A, Maleki M, Fried E (2013) Continuum theory for the edge of an open lipid bilayer. Adv Appl Mech 21:1–78

Cammarata RC (2009) Generalized thermodynamics of surfaces with applications to small solid systems. Solid State Phys 61:1–75

Cammarata R, Sieradzki K, Spaepen F (2000) Simple model for interface stresses with application to misfit dislocation generation in epitaxial thin films. J Appl Phys 87(3):1227–1234

Chatzigeorgiou G, Meraghni F, Javili A (2017) Generalized interfacial energy and size effects in composites. J Mech Phys Solids 106:257–282

Chhapadia P, Mohammadi P, Sharma P (2011) Curvature-dependent surface energy and implications for nanostructures. J Mech Phys Solids 59(10):2103–2115

Courant R, Hilbert D (1953) Methods of mathematical physics, vol I (First English ed.). Interscience Publishers, Inc., New York

Duan H, Wang Jx, Huang Z, Karihaloo BL (2005) Size-dependent effective elastic constants of solids containing nano-inhomogeneities with interface stress. J Mech Phys Solids 53(7): 1574–1596

Duan H, Wang J, Karihaloo BL (2009) Theory of elasticity at the nanoscale. In: Advances in applied mechanics, vol 42. Elsevier, Amsterdam, pp 1–68

Fischer F, Waitz T, Vollath D, Simha N (2008) On the role of surface energy and surface stress in phase-transforming nanoparticles. Prog Mater Sci 53(3):481–527

Fried E, Todres RE (2005) Mind the gap: the shape of the free surface of a rubber-like material in proximity to a rigid contactor. J Elast 80(1–3):97–151

de Gennes PG, Brochard-Wyart F, Quere D (2004) Capillarity and wetting phenomenon. Springer, New York

Gurtin ME, Murdoch AI (1975a) Addenda to our paper a continuum theory of elastic material surfaces. Arch Ration Mech Anal 59(4):389–390

Gurtin ME, Murdoch AI (1975b) A continuum theory of elastic material surfaces. Arch Ration Mech Anal 57(4):291–323

Gurtin ME, Murdoch AI (1978) Surface stress in solids. Int J Solids Struct 14(6):431–440

Gurtin M, Weissmüller J, Larche F (1998) A general theory of curved deformable interfaces in solids at equilibrium. Philos Mag A 78(5):1093–1109

Gurtin ME, Fried E, Anand L (2010) The mechanics and thermodynamics of continua. Cambridge University Press, Cambridge

Haiss W (2001) Surface stress of clean and adsorbate-covered solids. Rep Prog Phys 64(5):591

Henann DL, Bertoldi K (2014) Modeling of elasto-capillary phenomena. Soft Matter 10(5): 709–717

Huang Z, Sun L (2007) Size-dependent effective properties of a heterogeneous material with interface energy effect: from finite deformation theory to infinitesimal strain analysis. Acta Mech 190(1–4):151–163

Huang Z, Wang Jx (2006) A theory of hyperelasticity of multi-phase media with surface/interface energy effect. Acta Mech 182(3–4):195–210

Huang Z, Wang J (2013) Micromechanics of nanocomposites with interface energy effect. Handbook of micromechanics and nanomechanics. Pan Stanford Publishing, Singapore

Ibach H (1997) The role of surface stress in reconstruction, epitaxial growth and stabilization of mesoscopic structures. Surf Sci Rep 29(5–6):195–263

Javili A, McBride A, Steinmann P (2013) Thermomechanics of solids with lower-dimensional energetics: on the importance of surface, interface, and curve structures at the nanoscale. A unifying review. Appl Mech Rev 65(1):010802

Javili A, Ottosen NS, Ristinmaa M, Mosler J (2018) Aspects of interface elasticity theory. Math Mech Solids 23:1004–1024

Li S, Wang G (2008) Introduction to micromechanics and nanomechanics. World Scientific Publishing Company, Singapore

Liu L, Yu M, Lin H, Foty R (2017) Deformation and relaxation of an incompressible viscoelastic body with surface viscoelasticity. J Mech Phys Solids 98:309–329

Miller RE, Shenoy VB (2000) Size-dependent elastic properties of nanosized structural elements. Nanotechnology 11(3):139

Müller P, Saúl A (2004) Elastic effects on surface physics. Surf Sci Rep 54(5–8):157–258

Murdoch AI (2005) Some fundamental aspects of surface modelling. J Elas 80(1–3):33

Pala RGS, Liu F (2004) Determining the adsorptive and catalytic properties of strained metal surfaces using adsorption-induced stress. J Chem Phys 120(16):7720–7724

Park HS (2008) Strain sensing through the resonant properties of deformed metal nanowires. J Appl Phys 104(1):013516

Park HS, Klein PA, Wagner GJ (2006) A surface Cauchy–Born model for nanoscale materials. Int J Numer Methods Eng 68(10):1072–1095

Sharma P, Ganti S, Bhate N (2003) Effect of surfaces on the size-dependent elastic state of nano-inhomogeneities. Appl Phys Lett 82(4):535–537

Steigmann D, Ogden R (1997) Plane deformations of elastic solids with intrinsic boundary elasticity. Proc R Soc London A 453:853–877

Steigmann D, Ogden R (1999) Elastic surface substrate interactions. Proc R Soc London A 455:437–474

Style RW, Hyland C, Boltyanskiy R, Wettlaufer JS, Dufresne ER (2013) Surface tension and contact with soft elastic solids. Nat Commun 4:2728

Style RW, Jagota A, Hui CY, Dufresne ER (2017) Elastocapillarity: surface tension and the mechanics of soft solids. Ann Rev Conden Matter Phys 8:99–118

Suo Z, Lu W (2000) Forces that drive nanoscale self-assembly on solid surfaces. J Nanopart Res 2(4):333–344

Wang J, Huang Z, Duan H, Yu S, Feng X, Wang G, Zhang W, Wang T (2011) Surface stress effect in mechanics of nanostructured materials. Acta Mechanica Solida Sinica 24(1):52–82

Wang ZQ, Zhao YP, Huang ZP (2010) The effects of surface tension on the elastic properties of nanostructures. Int J Eng Sci 48(2):140–150

Part XII
Glass Science and Technology: Predictive Modeling

Modeling of Glasses: An Overview

80

Mehmet C. Onbaşlı and John C. Mauro

Contents

Abstract

Glass modeling includes unique advantages and challenges with respect to other fields of materials modeling owing to lack of long-range order, strong dependence on temperature and pressure history, statistical nature of glass-forming liquid, and the availability of almost the entire periodic table for constituents in glass. In this chapter, we introduce a range of methods used for glass modeling and overcoming these challenges. We first briefly compare how glass modeling is different from crystalline materials. Next, we briefly outline some of the techniques used for modeling glass and finally present the outstanding challenges in glass modeling and design. As glass modeling merges

M. C. Onbaşlı (✉)
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Electrical and Electronics Engineering, Koç University, Istanbul, Turkey
e-mail: monbasli@ku.edu.tr

J. C. Mauro
Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA, USA
e-mail: jcm426@psu.edu

empirical techniques (i.e., data-driven machine learning, finite element models for mechanical and acoustic properties, composition/property/processing relationships) with fundamental physical methods (i.e., statistical physics, diffusion, first principles quantum mechanical theories, energy landscapes), many orders of magnitude in time- and length scales may be simultaneously modeled across vast composition spaces whose experimental exploration would be prohibitively expensive.

1   Introduction

In current and emerging technologies, functional glasses play a critical role due to their unique final properties such as optical transparency, mechanical strength and elasticity, thermal insulation, chemical durability, liquidus temperature, and refractory compatibility. Traditionally, glass research has been based on trial-and-error experimentation on structure, property, and processing relationships. Compared with other classes of materials such as crystals, glass lacks long-range order and glass structure changes over short- and intermediate-range orders on different length scales, which dramatically alters macroscopic properties. In glass, local bonding and packing of cations and anions are composition dependent and typically differ from those in the compositionally analogous crystals. Since nearly all elements of the periodic table are available for glass formation, one can explore virtually infinitely many glass compositions. As glass structure also evolves as a strong function of its thermal and pressure history, one also needs to trace microstructural evolution in glass as a function of thermal input and pressure changes. As molecular topology is driven by the statistical nature of glass-forming liquid, the intermediate properties of this liquid are critical in understanding and engineering of glass composition and properties. Hence, due to the compositional and structural complexity of glass, a rigorous understanding of composition, property, and processing relationship through modelling has been difficult and glass-modeling studies have been limited.

Recent developments in ab initio models for organic-glass adhesion, constitutive mechanical modeling of metallic glass, data-driven models of glass, and other modeling methods have made major strides in accurately predicting glass properties and in optimizing competing manufacturing-related glass attributes and final glass properties. In this book chapter, we introduce the recently developed glass modeling methods, which enable accurate prediction of final glass properties and manufacturing related attributes. These methods also help bridge modeling across multiple time- and length scales and merge insights from fundamental physical and empirical models.

This chapter serves as an introduction to this part "Glass Science and Technology: Predictive Modeling" of the handbook. Owing to the complexity of glass properties and structure, one may classify glass-modeling methods in terms of their empirical or fundamental physical origins. Depending on the range of length and timescales, the number of constituents involved, the properties of interest, the types

of fundamental physical insights to be captured, and error tolerance for coarse-graining, researchers pick from the methods shown on Fig. 1 or develop new modeling methods.

Figure 1 shows the range of methods that one may use for glass modeling. Each method has its own advantage and disadvantage and depending on the modeling purpose (i.e., kinetic, thermodynamic, or multicomponent glass phase diagrams), different limitations arise. As shown in the classification of Fig. 1, one may use purely empirical modeling techniques such as regression, neural networks, and data mining approaches. In these methods, one starts with an experimental data set for glass composition-property relationships. Then, regression analysis is used to identify linear relationships and correlations between glass composition and major glass properties. When the process and composition parameters interact nonlinearly, neural network models with many layers and neurons might be more useful in capturing the subtle correlations within many-component glass systems. The use cases for regression and machine learning methods such as artificial neural networks are described in ▶ Chap. 81, "Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass," of the glass modeling part (Onbasli et al.).

In this section on glass modeling, there are four chapters, which use different sets of methods. These methods span multiscale finite element modeling, high-
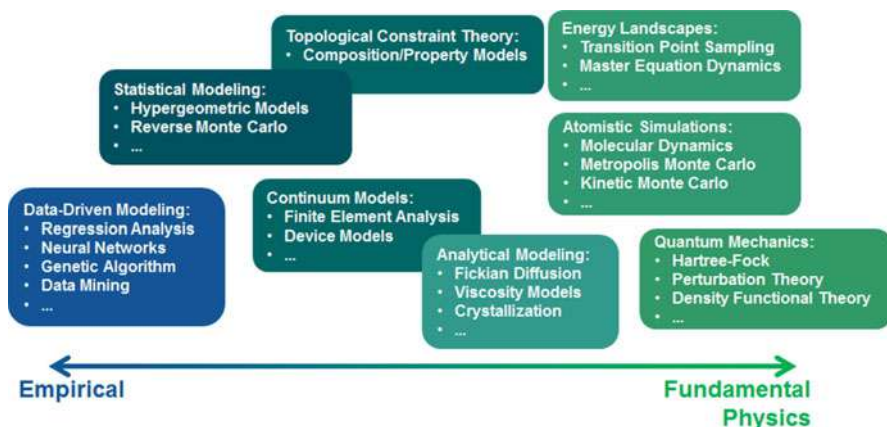
Fig. 1 Glass modeling methods based on their length-scale and the classes of techniques involved. The methods based on first principles calculations such as quantum mechanics, molecular dynamics, and energy landscapes involve shorter length scales while capturing the glass behavior more rigorously. The methods, which involve longer length scales (microscopic) as continuum and thermodynamic models, are introduced in the middle category. These methods cover finite element modeling, diffusion models, topological constraint theory, viscosity, and crystallization models. Finally, as much longer length, time, and composition spaces are involved, statistical and nonlinear machine learning models prove to be more effective in capturing the overall change in glass properties with empirical and data-driven approaches. (Reprinted with permission from Mauro et al. 2016, 2016, American Chemical Society.)

throughput compositional screening, and data science techniques to gain insight into glass composition-property relationships as well as industrially viable glass property predictions. The second chapter of this part, ▶ Chap. 81, "Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass" (Onbasli et al.), describes a purely data-driven approach into composition optimization for obtaining commercially viable Corning Gorilla® Glass liquidus temperature, Young's modulus, and viscosity. The fact that more than 10 elemental and compound constituents in glass composition increase the composition space significantly and exploring this vast parameter space experimentally becomes prohibitively expensive in expenses and in time. As a result, Corning's unique advantage of its long experimental records of glass trial-and-errors has been used to empirically predict and classify glass compositions, which may be promising for the desired ranges of liquidus temperature, Young's moduli, and viscosities.

As mechanical properties of glass are modeled across micrometer to centimeter length scales, one needs to use multiscale formalisms such as finite element methods (FEM) and numerical analysis techniques. The bridge between macroscopic and mesoscopic length scales requires an accurate modeling of the constitutive relationships of mechanical properties of glasses and metallic glass compositions. On Fig. 1, these techniques are classified as continuum models, which do not consider the discrete molecular structure of glass. Avoiding the molecular precision allows for bridging granular mechanical behavior to macroscopic mechanical properties of materials without significantly compromising accuracy. The predictive accuracy of these mechanical FEM models is mostly determined by the accuracy of the constitutive mechanical property models for materials. As a result, we include the third chapter, ▶ Chap. 82, "Constitutive Modeling in Metallic Glasses for Predictions and Designs" (Li), in this glass modeling section to address the multiscale modeling challenge for analyzing and designing mechanical properties of metallic glasses. This chapter demonstrates how constitutive modeling can help engineer new high-strength metallic glasses with superior optimized mechanical properties and the same techniques could be used for mechanical modeling and rational design of other materials.

The next chapter in glass modeling is ▶ Chap. 83, "Fundamentals and Applications of Organic-Glass Adhesion" which focuses on first-principles calculations with rigorous physical insights into the thermodynamics of adhesion of organic molecules on glass (Goyal et al.). In this chapter, density functional theory and molecular dynamics models are used to predict the binding energies of different common organic molecules such as polyimide on silica. Molecular dynamics models help develop predictive power over the binding kinetics up to nanometer length scales. This predictive power originates from the development of accurate force fields such as ReaxFF and Interface FF. Once the accurate force fields are used in the molecular dynamics simulations, one can design and optimize organic/glass interfaces of desired strength (i.e., binding energies), which would be necessary for technological display applications.

The final chapter in glass modeling is ▶ Chap. 84, "Design and Modelling of High-Strength, High-Transmission Auto Glass with High Sound Transmission Loss" which uses laminated composite glass-polymer layers for enhanced sound transmission loss and sound isolation while retaining the mechanical strength, wide elastic range, scratch resistance, and high optical transparency (Onbasli, this volume). For automotive, industrial, and architectural purposes, glass must have high sound transmission loss to achieve good sound isolation. While engineering glass compositions allows for tuning the mechanical properties of glass for high strength and scratch resistance, geometric requirements of applications and acoustic phonon band structure of glass originating from its composition limit the flexibility on tuning the sound transmission loss via intrinsic band structure engineering. As a result, composite multilayers need to be used such that the mass density and effective acoustic impedance can be engineered for maximum reflection over audible sound frequency ranges. In this chapter, this challenge is posed as a multilayer composite materials optimization problem in which optical transmission and mechanical properties must be retained while the effective wideband acoustic impedance can be maximized. For enhancing acoustic impedance, sound wave equation, constitutive relations, and appropriate boundary conditions are defined to model the acoustic reflection, transmission, and absorption spectra of glass. Once the formalism is established and the key modeling steps are streamlined, optimization for automotive and other wide area glass applications becomes feasible.

In the rest of this chapter, the methods used in these studies are described as an overview. The methods covered also find emerging and groundbreaking uses in multiscale modeling for other materials classes. By finite element modeling and accurate constitutive modeling, one can achieve multiscale modeling of materials. Through data-driven approaches, one can systematically screen vast composition spaces and defect states with high-throughput and streamline material property optimization. DFT- and MD-based techniques help bridge material behavior at mesoscopic (i.e., molecular interfaces) and microscopic (i.e., experimentally easily accessible micron-scale properties) scales. This introductory chapter ends with a short overview of new glass modeling problems. A multiscale modeling approach in which the advantages of the methods discussed here could be merged might address these glass modeling problems.

2   Methods in Glass Modeling

In this part, we describe the three categories of methods used in the glass modeling section of the handbook. These categories are:

1. Data-driven and high-throughput compositional screening
2. Multiscale finite element modeling
3. DFT- and MD-based first-principles modeling

2.1   Data-Driven and High-Throughput Compositional Screening

In the data-driven materials modeling approach, the main purpose is to understand and predict the structure-property and processing relationships in glass by using statistical models as well as nonlinear machine learning models (such as artificial neural networks or genetic algorithms). The main advantage in this approach is that if an experimental data set is available to partially describe the structure-property and processing relationships, one may develop predictive models for the key properties such as liquidus temperature, viscosity, and Young's modulus of multicomponent glass systems treated under different conditions. By spanning more than ten main and trace components in glass, one can computationally narrow down on the composition subspaces, which may yield the optimal liquidus temperature, viscosity, and Young's modulus and many other desired properties. A disadvantage of this approach is that an experimental dataset must be available to build descriptive and predictive analytical models and building this dataset might be expensive. The models are coarse-grained in the sense that some of the effects of trace components may not necessarily be captured within the predictive error margin of a data-driven model. Nevertheless, such models help identify trends and effects of composition on final glass properties and manufacturing-related attributes, which could later be studied in finer detail using fundamental physical or finite element models. Below is a brief description of the data-driven methodology used for glass composition development for optimized liquidus temperature, viscosity, and Young's modulus predictions.

From a high-impact commercial research and development standpoint, final glass properties and manufacturing-related attributes need to be optimized simultaneously. Glass composition design can thus be posed as a multiobjective optimization problem with constrains on the stoichiometry and percentage weights of the constituents and the range of each stoichiometry. One objective function for each property can be given as a range ($70\ \text{GPa} \leq E \leq 85\ \text{GPa}$) or lower-bounded range ($AP \geq 800\ °C$). Corning's proprietary glass composition and property database developed over many years span key constituents and linear regression models are used whenever existing experimental data may coincide with the desired properties. For conflicting constraints or nonlinear material behavior, nonlinear optimization and machine learning models need to be used. Below is the description of the steps involved in data-driven materials modeling for glass.

Data consolidation and cleaning: Regardless of the type of data-driven approach used, data quality is the most important aspect that determines the success of the models to be built. Data quality is defined in terms of both data organization and the signal-to-noise ratio at which data were acquired. In typical data-driven projects, data consolidation and cleaning consists of restructuring the data in a format and structure that could be computationally processed. In data cleaning, the gaps within the data structure are eliminated by either interpolation or complete elimination of the data row associated with the missing measurement. Once the dataset is constructed, regression and machine learning models are built for predictions.

In ▶ Chap. 81, "Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass," Corning researchers used the proprietary Corning glass composition and property relations. These data sets are based on Corning's proprietary experimental data from the company's long history of glass composition research. The composition information has been stored as the weight or atomic or molar percentages of each oxide constituting glass. For outputs, the viscosities η, Young's moduli E, glass transition temperatures T_g, densities ρ, as well as the liquidus temperatures T_L are used. For commercial purposes, Corning researchers had to optimize glass compositions, which would help achieve a high enough strength and scratch resistance while allowing for high enough manufacturing yield through low enough viscosities. The purpose of the models therefore is to find the composition that yields the optimal mechanical, chemical, or rheological properties of glass.

After data consolidation and cleaning comes a step called *feature detection*. This step is a group of computational algorithms used for input data sets without labeled response variables (i.e., properties). These algorithms help detect anomalous data entries. These algorithms help group different composition types and predict based on their measured mechanical properties. As part of this step, we describe cluster analysis, regression, genetic algorithms, and neural networks.

Cluster analysis: For building a reliable model, one needs to break down the data into subspaces of glass compositions. Since manual identification of systematic glass compositional variations is not feasible and is error-prone, a classification algorithm called cluster analysis is developed. First, the proximity and similarity between each data point in the data are calculated. Then, a criterion function is defined for effective clustering based on the proximity and similarity metrics. Dissimilar data items are grouped under different clusters. As a result, trends within subsections of glass composition space can be analyzed. Since glass composition data span more than 15 orders of magnitude; their values cannot be directly used in models as numerical computational errors may become larger than the weights of the properties with smaller values. In order not to cause one column to dominate distances based on their absolute values, each value is normalized to index values before calculating the distances. A conversion formula is defined in the beginning for each property or column for converting these normalized values back to their original scales. After normalization, the distances between each data point can be found.

Principal component analysis (PCA): PCA is a dimension reduction technique that allows omitting variables, which do not have significant effect on the desired glass property (▶ Chap. 81, "Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass," Onbasli et al.). This method uses an orthogonal mathematical transformation on data to convert a mapping of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. The first principal component has the largest possible variance, and every next component has progressively lower variances. The output of PCA is a set of vectors transformed out of the original data. These vectors form an

uncorrelated orthogonal basis set. As variances depend on the absolute mean values of each variable, PCA output vectors strongly depend on the relative scaling of the initial variables. Therefore, normalizing each variable is essential before running PCA calculations. After running the PCA algorithm on data, a significance threshold is determined for the eigenvectors. This threshold step helps determine the most significant variables in a glass composition project and focus on cluster analysis using those fewer and more important variables (Kullback and Leibler 1951). PCA reduces computational costs by reducing the number of variables involved while allowing for a small enough prediction error due to the omitted eigenvectors. PCA is one of the steps that enable higher throughput screening across large composition spaces.

Regression analysis: PCA may fail to capture nonlinear correlations among variables and regression analysis might be necessary to distinguish if there are hidden complex dependencies. In regression analysis, the conditional expectation $E(X|H)$ is estimated when the independent variables are given. In other words, either a generalized linear model is constructed by using a multivariable linear superposition of independent variables or a least squares fit is solved numerically among variables to identify if there is a nonlinear correlation.

Multivariable linear regression: This is a generalized linearization method that assumes one or more independent variables:

$$Y_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n + \varepsilon \tag{1}$$

where each X is an independent variable and ϵ i is an error term which is minimized until the optimal set of regression coefficients α_0 to α_n are found. These coefficients also indicate the relative significance of each independent variable.

Least squares regression model: This fitting model assumes a sophisticated nonlinear correlation between variables. No closed-form solutions are assumed here and one may pick a Taylor expansion or a different basis set to fit the dependent variable by minimizing the sum of the squared residuals S.

Genetic algorithms: Genetic algorithms (GA) are methods for solving constrained optimization problems based on an evolutionary natural selection process. An iterative convergence into the optimal solution is essential in any optimization problem. In a classical numerical solution algorithm (i.e., Newton's method solving the zeros of equations), one generates a single point at each iteration. The sequence of points progressively approaches the solution. In GA, instead of iterating only on one point, a population of points are generated at each iteration and the best point in population approaches the solution.

A key difference between classical algorithms and GA is that in GA, each successive population is generated using random number generators and some combinations from the best solutions of the previous iteration instead of deterministic computation. GA are useful for accelerated global optimization problems where there are multiple objectives with competing requirements. GA can also be used for maximization, minimization, constrained optimization, multidimensional fitting, classification, and prediction problems. GA are therefore beneficial for composition development.

Below is a description of a simplified GA for solving a generic multi-objective optimization problem: The GA starts by generating a random population of glass compositions. The GA then evaluates the glass composition population based on their predicted properties and the fitness values. The GA then scores each composition of the current population by computing the fitness value for each and normalizing their scores for comparisons. The GA picks a few candidates from the population (parents) based on their fitness. Some of the individuals in the current population that have lower fitness are chosen as elite. These elite individuals are passed to the next population (children). Children are produced either by making random changes to a single parent (mutation) or by combining the vector entries of a pair of parents (crossover). Then, the algorithm replaces the current population with the children to form the next generation. The GA then creates a sequence of new populations and successive generations. At each step, the GA uses the compositions in the current generation to create the next population. The algorithm stops when the output meets the stopping criteria. The stopping criteria may include a maximum number of generations, time limit for iterations, fitness limit for population (when the fitness function for the best point in the population is less than or equal to the fitness limit), and marginal change limits per generation (i.e., when the average change over each generation is less than a functional tolerance limit). The algorithms mentioned can be implemented in MATLAB, R, or Python.

Neural Networks: Neural networks (NN) are a set of machine learning algorithms that can be used to predict nonlinear behavior of different systems. A neuron is a nonlinear and parameterized function of input variables. Therefore, a neural network is the composition (weighted sum) of nonlinear functions of two or more neurons. The nonlinearity of neural networks helps model nonlinear behaviors that may not otherwise be captured in regression models or in linear techniques such as PCA. Despite the biological flavor of the term neural network, neural networks in machine learning are pure mathematical constructs. For glass modeling, this handbook covers feedforward networks, which do not include feedback. In an optimized NN model, a sequence of neurons may also yield physical insight about a given physical parameter and help develop physics-based models later. Although correlations may not necessarily imply causality, consistent trends in NN models may guide fundamental physical models on composition spaces not studied previously.

A NN model consists of a graph structure where each edge indicates the weight of the given variable entering the node (neuron). Each node indicates the nonlinear activation function that captures the nonlinear glass behavior. The outputs of each activation function are entered into the next layer of activation functions with different weights. After such successive nonlinear active calculations are done, the output neuron, which is the node at the output, calculates the final classification or prediction result. The rest of the layers between the inputs and the output neuron are hidden layers of neurons. In NN optimization, the number of hidden layers, the number of neurons in each layer, the activation function of each layer (tanh, sigmoid, inverse tangent), and the learning rate (continuous number between 0.0001 and 0.1) are optimized by entering the data set as training data. After the prediction outputs

are obtained, these outputs are compared with the final glass properties and error rate is calculated as a percentage difference. The data are fed into the NN model to optimize the model parameters until the error rate converges to a minimum.

2.2   Multiscale Finite Element Modeling

Finite element modeling is class of modeling techniques that can be applied to multiple different scales and physics from mesoscopic to microscopic and classical mechanical models to acoustic sound transmission. Finite element modeling is a geometric divide-and-conquer technique for modeling the local material behavior for given physics (i.e. Maxwell's equations, mechanical elasticity/plasticity or flow relations, acoustic wave equation relations, semiclassical electronic properties etc.), constitutive relations which capture material properties and the appropriate boundary conditions for known behaviors of interfaces (i.e. rigid or absorbing boundaries). In glass modeling section of this handbook, finite element methods have been used to describe the mechanical strength and ductility of metallic glasses and to describe sound transmission loss of optically clear, high-strength automotive glasses. In both cases, finite element modeling describes the geometries in mesh structures and the adjacent meshes affect each other's response variables such as deformation or displacement or sound transmission. In this part, first, finite element modeling (FEM) used in describing the mechanical behavior metallic glasses is introduced. Next, FEM used in acoustic sound transmission of glass is presented.

2.2.1   FEM for Designing High Strength and Highly Elastic Metallic Glasses

Metallic glasses include metallic elements and have amorphous structure. These materials have extraordinary high strength and high elastic limits at room temperature. Metallic glasses are brittle and suffer from inhomogeneous deformation characteristics of shear banding, flow serration and cavitation. The net-shape forming ability is typically observed for most glasses when heated above glass transition temperature, but metallic glasses deform in a homogeneous manner. Understanding both the homogeneous and inhomogeneous deformation in metallic glasses is important for designing metallic glasses with improved ductility while preserving their high strength and elasticity. As a result, finite element modeling accurate constitutive models for metallic glasses are necessary to understand the deformation behavior of metallic glasses and develop their ductility.

In the glass modeling section of this handbook, ▶ Chap. 82, "Constitutive Modeling in Metallic Glasses for Predictions and Designs" (Li, this volume) has been included to address the multiscale modeling challenge for analyzing and designing mechanical properties of metallic glasses. Constitutive modeling captures the homogeneous and inhomogeneous deformation behavior and helps identify brittle failure conditions as a function of loading, shear banding and flow serration and cavitation.

The finite element models in this chapter use a free volume model-based constitutive relationship to describe homogeneous and inhomogeneous plastic flow. Since the atoms in an amorphous metallic glass are arranged in a disordered and non-closely packed manner, the excess space of the structure is described as free volume per atom. This volume is quantified as the difference between the average atomic volume in an amorphous structure and that in an ideal ordered structure. In the free volume model, macroscopic plastic flow is assumed to be induced by a number of individual atomic jumps. When no external force is applied, these jumps are driven by thermal activation energy. A dynamic equilibrium is assumed such that the number of atoms jumping forward balances the number of those jumping backward. Therefore, no net macroscopic flow occurs. When a shear stress is assumed due to an external force, a net flux of atoms along one direction are going to be observed. This net flux describes the irreversible macroscopic deformation. When shear deformation is modeled in a simplified 1D case, the plastic shear strain rate is the fraction of the thermally activated atomic jumps within the free volume where the success rates of the jumps are determined by temperature and the activation energy barrier (ΔG^m):

$$\frac{\partial \gamma_p}{\partial t} = 2f \, \exp\left(-\frac{\alpha v^*}{v_f}\right) \exp\left(-\frac{\Delta G^m}{k_B T}\right) \sinh\left(\frac{\tau \Omega}{2k_B T}\right), \qquad (2)$$

where f is the frequency of the atomic vibrations, α is a geometric factor, v^* is the hard-sphere volume of an atom, v_f is the average free volume per atom, is the activation energy, Ω is the atomic volume, k_B is the Boltzmann constant, and T is the temperature in Kelvin.

In this standard viscoplastic model description shown with Eq. (2), the strain softening effect found in metallic glasses cannot be fully accounted for. To account for strain softening, the free volume, v_f, is treated as an order parameter whose creation and annihilation govern the structural disordering and reordering processes in metallic glasses. The free volume can be created by a sufficiently high stress squeezing an atom into a neighboring vacancy of smaller volume; it can be annihilated by diffusional atomic jumps. The net rate of free volume increase is:

$$\frac{\partial v_f}{\partial t} = v^* f \exp\left(-\frac{\alpha v^*}{v_f}\right) \exp\left(-\frac{\Delta G^m}{k_B T}\right) \left\{ \frac{2\alpha k_B T}{v_f C_{\text{eff}}} \left[\cosh\left(\frac{\tau \Omega}{2k_B T}\right) - 1\right] - \frac{1}{n_D} \right\},$$
$$(3)$$

where n_D is the number of atomic jumps needed to annihilate a free volume equal to v^* and is usually below 10, and the effective modulus is $C_{\text{eff}} = E/3(1 - v)$ with E and v being the Young's modulus and Poisson's ratio, respectively. Equations (2) and (3) describe the inelastic portion of the continuum mechanics of deformation mechanism in metallic glasses. In conjunction with an elasticity formula given below, a complete one-dimensional isothermal constitutive law for metallic glasses is formed

$$\frac{\partial \gamma_e}{\partial t} = \left(\frac{\dot{\tau}}{G}\right)_T \tag{4}$$

where G is the shear modulus. In addition to this isothermal model, thermal transport equation is also used to take plastic energy dissipation and heat conduction into account. As a result, the yield criteria for metallic glasses under multiaxial stress states could be modeled. Further details on these models are presented by W. Li in ▶ Chap. 82, "Constitutive Modeling in Metallic Glasses for Predictions and Designs" of this handbook. These models can be implemented in ABAQUS software for numerical modeling of multiaxial shear strain effects with respect to J_2 plasticity, which assumes that the hydrostatic stress does not cause plastic flow or any changes in free volume.

In addition to free volume model, constitutive models could also be built phenomenologically to capture the macroscopic mechanical response of metallic glasses. Anand and Su (2005) developed a finite-deformation constitutive theory that captures some of the essential inelastic deformation features in metallic glasses. These features include the hydrostatic-stress dependence of yielding, plastic dilatation, and plastic strain localization. This model was first developed for room temperature and then extended to high homologous temperatures within $0.7T_g \lesssim T \lesssim 0.9T_g$ for simulating homogeneous plastic flow (Anand and Su 2007) and $0.9T_g \lesssim T \lesssim T_x$ (Henann and Anand 2008) for the transition from a viscoelastic-plastic solid-like response to a Newtonian fluid-like response for isothermal condition.

Plastic deformation does work on atoms during mass transport and dissipates heat. As a result, temperature changes during plastic deformation need to be incorporated into the models for more accurate description of the thermomechanical changes. Yang et al. (2006) consider the temperature rise in metallic glasses resulting from the conversion from plastic work to heat but treats the system as adiabatic. Thamburaja and Ekambaram (2007) applied fundamental thermodynamic laws and the principle of microforce balance to develop a thermomechanically coupled finite-deformation model for predicting within supercooled liquid regime. This model was further developed to include a damage criterion to simulate impact and damage processes (Ekambaram et al. 2011). These constitutive models have been implemented in commercial finite-element program ABAQUS by writing a user-defined material script for the explicit (VUMAT) or the implicit procedure (UMAT). The model by Anand and Su at room temperature (Anand and Su 2005) has been used extensively. Further details on the models are presented in ▶ Chap. 82, "Constitutive Modeling in Metallic Glasses for Predictions and Designs," by Li, this volume.

2.2.2 Finite Element Modeling for Sound Transmission Loss Calculations

In the second half of the FEM part, acoustic transmission modeling of glass is introduced. Sound transmission loss of laminated composite glass layers is

modeled with using linear acoustic wave equations and constitutive relations. After defining the geometry of the structure, the material properties of each layer are defined and the boundary conditions are established. Next, a commercial software package solves for the temporal and spatial evolution of the sound-induced pressure inside glass and external medium. Finally, the pressure and associated power distribution is calculated and normalized with respect to the incoming acoustic wave energy density to find sound transmission loss as a function of sound excitation frequency.

In linear acoustics, pressure changes are assumed to be small with respect to the mean pressure:

$$p = p_0 + p', p' \ll p_0 \tag{5}$$

Momentum and mass are conserved as described by Eqs. (6) and (7):

$$\frac{\partial \overrightarrow{v}_a}{\partial t} = -\frac{1}{\rho_0} \nabla p_a + \frac{4\mu}{3\rho_0} \nabla \left(\nabla \cdot \overrightarrow{v}_a \right) \tag{6}$$

$$\frac{1}{\rho_0 c_0{}^2} \frac{\partial p_a}{\partial t} = -\nabla \cdot \overrightarrow{v}_a + \frac{Q}{\rho_0} \tag{7}$$

where ρ_0 is the mass density of the medium (kg·m^{-3}), p_a is the acoustic pressure (N·m^{-2}), v_a is the sound velocity (m·s^{-2}), and the Q/ρ_0 term (s^{-2}) is that captures mass loss due to sound wave-medium interaction. Using Eqs. (6) and (7), the equation captures the spatial and time dependence of linear acoustic pressure profile is derived:

$$\nabla \cdot \left(\frac{1}{\rho_0} \nabla p_a \right) - \frac{1}{\rho_0 c_0{}^2} \frac{\partial^2 p_a}{\partial t^2} + \nabla \cdot \left[\frac{1}{\rho_0} \left(\frac{4\mu}{3} + \mu_B \right) \nabla \left(\frac{1}{\rho_0 c_0{}^2} \frac{\partial p_a}{\partial t} \right) \right]$$
$$= -\frac{\partial}{\partial t} \left(\frac{Q}{\rho_0} \right) + \nabla \cdot \left[\frac{1}{\rho_0} \left(\frac{4\mu}{3} + \mu_B \right) \nabla \left(\frac{Q}{\rho_0} \right) \right] \tag{8}$$

This equation is used for solving for the pressure distribution across finite elements. The solution types to the acoustic transmission problem include:

- *Mode solutions*, in which eigenmodes of the system are found.
- *Harmonic solutions*, in which sound pressure level and power distribution at each junction is found. This helps understand the sound transmission loss spectrum of a laminated windshield glass. ▶ Chapter 84, "Design and Modelling of High-Strength, High-Transmission Auto Glass with High Sound Transmission Loss" (Onbasli, this volume), we focus on harmonic solutions.

- *Transient solutions*, in which time evolution of the spatial dependence of sound pressure or acoustic energy density profile is solved.

2.2.3 Material Properties

The material properties that govern sound isolation simulations include mass density, sound speed in the material, and dynamic viscosity. Dynamic viscosity is defined as $\eta = F/(A \times SR)$, where F is force, A is area, and SR is the shear rate. Dynamic viscosity is a term in units of Pa·s and it introduces dissipative effects and sound absorption losses, which is maximized for high sound transmission loss applications.

Sound wave propagation in air is modeled as a nonuniform ideal gas and differently than in glass. The spatial and temporal evolution of sound waves in air is described using Eqs. (9), (10), and (11):

$$\nabla \cdot \left(\frac{1}{\rho_0\left(\overrightarrow{r}\right)} \nabla p \right) - \frac{1}{\rho_0\left(\overrightarrow{r}\right) c^2\left(\overrightarrow{r}\right)} \frac{\partial^2 p}{\partial t^2} = -\frac{\partial}{\partial t}\left(\frac{Q}{\rho_0\left(\overrightarrow{r}\right)} \right) \tag{9}$$

$$c^2\left(\overrightarrow{r}\right) = \gamma R T\left(\overrightarrow{r}\right) \tag{10}$$

$$P_{state}\left(\overrightarrow{r}\right) = \rho\left(\overrightarrow{r}\right) R T\left(\overrightarrow{r}\right) \tag{11}$$

The solutions to the acoustic wave equation strongly depend on the boundary conditions and how they are defined (i.e., Dirichlet, Neumann, or mixed). The boundary conditions in acoustics are defined as follows:

- *Perfectly matched layer boundary (PML)*: absorbs incoming pressure waves
- *Sound-hard surface*: A surface on which particle normal velocity is zero (or Neumann boundary with $\widehat{n} \cdot \nabla p = 0$)
- *Sound-soft surface*: A surface on which sound pressure is constrained (also known as Dirichlet boundary, where $p = p_0$ fixed pressure)
- *Mixed boundary*: The surface whose frequency-dependent wave transmission properties are modeled using a mixed impedance mismatch model, where acoustic impedance is defined as $Z = p/(vS)$, with p being pressure, v sound velocity, and S being the surface area.
- *Wave absorbing boundary conditions*: The domain or open medium is terminated with infinite, homogeneous, inviscid fluid (gas) and thus, outgoing propagating waves keep propagating outwards and do not reflect back.

Among these boundary conditions, radiation and absorbing fluid element boundaries are used for all types of analyses (modal, harmonic, and transient) and perfectly matched layers are used for harmonic analysis. In FEM, mesh sizes are typically chosen as $\lambda/10$–$\lambda/5$ where $\lambda = c_0/f$.

2.2.4   Definition of Sources

Depending on the application, the sound wave source is defined in different forms including incident pressure plane wave $P_{inc} = P_{inc,0}e^{j\varphi}e^{-j\vec{k}_i(\vec{r}-\vec{r_s})}$ and particle velocity $v_{n,f} = \frac{1}{\omega\rho_0}\hat{n} \cdot \vec{k}_i p_{inc}$. Here, the mass density and the sound speed in the given environment must be known to solve for sound transmission loss.

After the simulations are completed, sound pressure level and sound transmission loss are calculated:

$$SPL \; (dB) = 10\log_{10}\left(\frac{\sum_{i=1}^{N}p_i^2}{p_{ref}^2}\right) \tag{12}$$

Sound pressure level at the threshold of human hearing is 0 dB, which has the same pressure as the reference pressure, $p_{ref} = 2 \times 10^{-5}$ Pa. Here, p_i terms are pressures at the sampled frequencies. To calculate sound transmission loss, time-averaged sound power through an area S is calculated:

$$P_W = \frac{1}{2}\iint\limits_{S} p\vec{v}^* \cdot \hat{n}ds \tag{13}$$

$$\text{Sound transmission loss (TL) (dB)} = 10\log_{10}\left(\frac{P_{in}}{P_t}\right) \tag{14}$$

$$\text{Sound reflection loss (RL) (dB)} = 10\log_{10}\left(\frac{P_{in}}{P_{ref}}\right) \tag{15}$$

Sound transmission here is described in terms of reflected and transmitted fractions of incoming acoustic power. The transmitted acoustic power fraction is determined by each layer or mesh's acoustic impedance Z, which is defined as $Z = p/(vS)$, here, p is pressure, v is sound velocity, and S is the surface area. The reflections occur due to impedance mismatch between two dissimilar layers.

2.3   DFT- and MD-Based First-Principles Modeling

First-principles modeling provides an atomic scale description of glass structure, its surface binding properties, as well as the effects of each atomic dopant on the local electronic properties. Since density functional theory (DFT) is an approximate solution to the Hamiltonian describing the glass system, the electron density solutions are not strictly exact solutions. Nevertheless, DFT solutions provide accurate enough predictions for the electronic properties including band structures and gaps as well as the equilibrium structure. While DFT models can only describe the structure and electronic properties within a few nanometers at most, molecular dynamics (MD) models could describe the physical movements of atoms and

molecules within a few tens of nanometers. In these models, the positions of the molecules are first initialized according to the configuration of interest (i.e., an interface) and then force distributions are then calculated based on the potential profile originating from the electrostatic or other force field-profiles of molecules. Based on the force fields, the positions of the molecules are updated within a step size of Δt by accounting for the acceleration that molecules acquire. Such iterations are repeated successively until the structure converges to an equilibrium distribution of molecules. By developing more accurate force fields, one may describe the behavior of glass molecules and their interactions with organics, fluids, or other classes of materials as a function of time and different constituents.

In glass modeling, one application that has immediate technological application is organic/glass adhesion. In this application, the organic, such as polyimide, needs to bind on glass with sufficient strength (i.e., large enough binding energy), but the organic layer must later be easy to peel away without damaging the glass surface. As a result, the binding energy must also be low enough for practical peeling forces. ▶ Chapter 83, "Fundamentals of Organic-Glass Adhesion" describes this challenge and presents DFT and MD-based numerical solutions to organic/glass adhesion.

The models for characterizing adhesion are developed at two different scales: electronic (DFT) and atomic levels (MD). While DFT-based approaches are rigorous and more accurate, they are computationally more expensive than MD-based solutions to the adhesion problem. While MD-based solutions are less accurate than DFT-based electronic structure solutions, longer time- and length scales are accessible using MD. By employing a multiscale approach, the effect of polymer chemistry, polymer rigidity, structure of inorganic substrates, and inorganic surface attributes such as the role of hydroxylation and surface roughness can be studied. The key aspect that allows the multiscale approach is the potential energy profile between the organic and glass interface. The Interface potential consists of Lennard Jones and Columbic interactions as nonbonded interactions, harmonic bonds to mimic covalent bonds, and harmonic angles to model the structure of polymers more accurately. One major example is the PCFF-interface potential, which is an extension of the well-established PCFF potential:

$$
\begin{aligned}
E_{pot} = {} & \sum_{ij, \text{nonbonded } 1,2 \text{ and } 1,3 \text{ excluded}} \varepsilon_{ij} \left[2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^9 - 3 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
& + \frac{1}{4\pi\varepsilon_0} \sum_{ij, \text{nonbonded } 1,2 \text{ and } 1,3 \text{ excluded}} \frac{q_i q_j}{r_{ij}} + \sum_{ij, \text{bonded}} k_{r,ij} \left(r_{ij} - r_{0,ij} \right)^2 \\
& + \sum_{ijk, \text{bonded}} k_{\theta,ijk} \left(\theta_{ijk} - \theta_{0,ijk} \right)^2
\end{aligned}
$$

$$(16)$$

MD-based modeling focuses on physisorbed systems consisting only of non-bonded interactions between the organic and glass interface. Hydrogen bonding,

surface roughness, electrostatics, and the flexibility of organic molecules affect adhesion significantly. This result was also found to be consistent across length scales for polymers ranging from molecular fragment scale to full polymer film thickness. Based on such MD models with more accurate force field descriptions, one may capture both the dynamics, kinetics, and the chemical interaction between different molecules and glass. Such models have significant applications in engineering and optimizing the functional performance of organic/inorganic multilayers as they are becoming ubiquitous in smartphones and other embedded electronic devices.

3   Open Challenges in Glass Modeling and Design

Owing to the range of emerging technological applications of glass ranging from smart windows, smartphones, ultra-strong and ductile (metallic) glasses, and biocompatible glass surfaces and containers, new glass chemistries and processing capabilities must be developed. In this quest, multiscale modeling as well as fundamental physical models are going to continue to deliver many insights into the intrinsic mechanical, structural and rheological properties of glass and how glass surfaces interact with organics of different molecular structures and polarities, metals and other kinds of materials. From a glass composition engineering standpoint, developing streamlined machine learning models and experimental datasets for training these models for different chemistries are necessary for more accurate predictions. Below, we present and briefly discuss some of the open questions in glass modeling.

Data-driven models for new glass compositions with optimized mechanical, optical, electronic, acoustic, surface, and rheological properties: While the chapter presented in this section focuses on the rheological and mechanical properties, many other key properties could be improved by developing new glass chemistries through a data-driven optimization approach. These properties include surface conductivity, microwave transmission, impedance matching, as well their simultaneous optimization with optical transmission and mechanical properties.

Merging fundamental physical insights with data-driven approaches: Data-driven approaches can span large composition spaces owing to dimension reduction techniques such as PCA and ICA without compromising accuracy significantly. Since this approach is highly coarse-grained, the effects of defect chemistry as well as thermal history may not necessarily be captured properly. Therefore, insights on defect and constituents from fundamental physical models such as molecular dynamics and DFT need to be incorporated into genetic algorithms and NN models. There are multiple ways to incorporate fundamental insights into the models. The first way could be training the models with a narrower range of chemistries and their corresponding properties such that the effects of small concentrations of dopant oxides could be captured more effectively. The second way could be training the NN models with different kernel functions that actually resemble the physical behavior such as thermal activation, rectified linear, or other mechanisms. The third way

could be directly applying GA and NN models to fundamental physical models for high-throughput prediction of glass chemistries and their properties of interest. Instead of training with experimental data sets, which are expensive to acquire, the results of ab initio models could be used for training NN models and predicting the properties of vast composition spaces. Iterative NN and DFT/MD model building helps identify the physical mechanisms and new chemistries that may lead to optimized glass properties. Optimized chemistries predicted from NN models could be studied in greater detail using DFT and MD to elucidate the new physics yielding superior glass properties.

Nontraditional glass chemistries: Glass industry specialized on borosilicate glasses, but there are many other new chemistries with potentially higher mechanical, thermal, optical, and functional performance. Bulk metallic and chalcogenide glasses have attracted significant research interest due to their unique functionalities. Metallic glasses have brittle-to-ductile transition, which yields increased fracture toughness. Chalcogenide glasses have unique optical properties for infrared bands. Since the length- and timescales relevant for these glasses can span more than nine orders of magnitude, molecular dynamics or other approaches are not feasible for metallic and chalcogenide glasses. Data-driven approaches could help identify the compositions that can optimize multiple functionalities.

Finite element modeling with improved constitutive relationships: Constitutive models help capture the novel physics and the micromechanical behavior that enables high strength and high ductility in emerging metallic glasses. The constitutive models, once implemented in a finite element program as user subroutines, can be used for guiding material, product, or process designs. This capability is one of the most remarkable advantages in the continuum mechanics modeling that is unmatched by experiments and atomistic simulations. Finite element modeling could be used for designing metallic glasses for high ductility, strength, and toughness at room temperature and optimizing the net-shape forming processes of metallic glasses at high homologous temperatures.

Modeling adhesion of physisorption of nanometer thick polymers on glass across larger length- and timescales: As inorganic/organic interfaces are becoming ubiquitous in the glass world, the bonding chemistry at the interface needs to be understood in atomic scale detail using DFT-based calculations. Due to the computational cost of DFT, the length scale of models cannot exceed a few atoms. As a result, DFT-based models cannot provide microscopic description of inorganic/organic interfaces. As a result, microscopic factors such as surface roughness, steric effects which are nonbonding interactions that influence the molecule shapes and reactivity cannot be described. Although force fields such as ReaxFF and INTERFACE are used to describe the chemistry, the parameterization is still a demanding task. As a result, new force fields which are computationally inexpensive and which can capture bond chemistry with electronic structure accuracy are needed for a wide variety of functional organics on glass and other inorganic materials.

References

Anand L, Su C (2005) A theory for amorphous viscoplastic materials undergoing finite deformations, with application to metallic glasses. J Mech Phys Solids 53:1362–1396

Anand L, Su C (2007) A constitutive theory for metallic glasses at high homologous temperatures. Acta Mater 55:3735–3747

Henann D, Anand L (2008) A constitutive theory for the mechanical response of amorphous metals at high temperatures spanning the glass transition temperature: application to microscale thermoplastic forming. Acta Mater 56:3290–3305

Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22:79–86

Mauro JC, Tandia A, Vargheese D, Mauro YZ, Smedskjaer MM (2016) Accelerating the design of functional glasses through modeling. Chem Mater 28:4267–4277

Thamburaja P, Ekambaram R (2007) Coupled thermo-mechanical modelling of bulk-metallic glasses: theory, finite-element simulations and experimental verification. J Mech Phys Solids 55:1236–1273

Yang Q, Mota A, Ortiz M (2006) A finite-deformation constitutive model of bulk metallic glass plasticity. Comput Mech 37:194–204

Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass

81

Mehmet C. Onbaşlı, Adama Tandia, and John C. Mauro

Contents

Abstract

For complex glass compositions with more than eight constituent compounds, experimental measurements of the entire composition space are prohibitively expensive and time-consuming. In addition, for systems with such complexity,

M. C. Onbaşlı (✉)
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Electrical and Electronics Engineering, Koç University, Istanbul, Turkey
e-mail: monbasli@ku.edu.tr

A. Tandia
Science and Technology Division, Corning Incorporated, Corning, NY, USA
e-mail: TandiaA@corning.com

J. C. Mauro
Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA, USA
e-mail: jcm426@psu.edu

there is no physically predictive model. There are requirements imposed on the end properties of glass and manufacturability requirements such as appropriate liquidus temperature and sufficiently low viscosity at a given temperature. These competing requirements necessitate the development of data-driven machine learning models of glass composition and properties. These models enable accurate and systematic prediction of glass properties such as Young's moduli and liquidus temperature. Research companies with long track records of exploratory experimental research are in unique position to develop data-driven models by compiling and using their earlier internal experimental results. In this chapter, we present how Corning used this unique advantage for developing neural network and genetic algorithmic models for predicting compositions that would yield a desired liquidus temperature and Young's modulus.

1 Introduction

Glass is a transparent and non-crystalline inorganic material used in virtually all aspects of human life, and it is a key enabler for our civilization. Glass can be manufactured in sheets, wires, fibers, or arbitrary forms. The growing need for highly functional, manufacturable, and inexpensive glasses has prompted glass researchers to engineer glass compositions using almost the entire periodic table. As a result, state-of-the-art commercial glass may contain more than eight different oxides and substituents with desired weight fractions. In glass composition and property engineering, each of these constituents significantly influences the structural, rheological, thermal, and mechanical properties. These properties are critical for display applications. For different glass applications, the desired glass properties are different, and glass composition must be engineered accordingly. In addition, rheological and thermomechanical properties of glass affect the fusion draw rates in production and ultimately determine glass production throughput and yield. As a result, for low-cost and high-volume manufacturing of glass, rheological (viscosity) and thermomechanical properties (liquidus temperature, Young's modulus) must be engineered simultaneously through glass composition. Sometimes, the manufacturability requirements and end property requirements may pose competing constraints on glass composition. As a result, engineering glass composition becomes a complex multi-objective optimization problem for which a physics-based predictive solution is not currently available.

Traditional glass research have been using experimental trial-and-error approach for improving glass properties through composition engineering and measuring the properties of each new composition. Corning Glass Research also have been investigating the mechanical, optical, rheological, and other properties of glass samples as a function of glass composition over more than a century. As a result, Corning accumulated a vast internal glass composition-processing-property database. Since a fully physics-based solution is not currently available for designing new glass compositions and since experimentally spanning and testing the entire composition space is prohibitively expensive and time-consuming

for consumer electronics and other major display glass applications, Corning started developing data-driven machine learning models to accelerate glass and ceramic materials engineering. As data-driven models require a reliable set of glass composition, processing and property data, Corning is in a unique position to draw from its century-old glass research data (Sullivan and Taylor 1919) and build machine learning (ML) models. Such models enabled Corning to accelerate glass research significantly. A detailed description of Corning's glass research achievements has been presented in Tandia et al. (2019) and Mauro et al. (2016).

In this chapter, we present how data-driven models can be developed for accurately predicting display glass properties for given compositions. In particular, we present neural network-based composition models for predicting glass liquidus temperature and genetic algorithm-based glass models for predicting Young's moduli for given compositions. The liquidus temperature is a key property which must be known for the given composition to prevent the formation of crystalline phases during fusion draw and high-volume production of display glass. Young's modulus is a key stiffness metric that must be engineered for display applications. Using both linear and nonlinear data-driven models, we demonstrate how material properties can be accurately predicted and optimized.

Corning developed extensive expertise on functional glass compositions and properties that span tough, strong, and ultralow CTE (coefficient of thermal expansion) borosilicates with different dopants and tunable glass transition temperatures to different fusion draw rates for controlling production throughput and yield. Publicly available inorganic materials databases and high-throughput screening of materials (Curtarolo et al. 2013; Pilania et al. 2013; Potyrailo et al. 2011; Xiang et al. 1995; Mueller et al. 2015) mainly focus on crystalline materials studies. Therefore, a combination of glass manufacturing parameters and composition-property relationship data must be used from Corning's internal database for reliable modeling. The data analysis and prediction methods as well as composition and process information presented herein are intended to describe some of Corning's research and to guide the readers in doing similar data-driven studies on different materials databases. Corning's internal database, which was used for this study, includes composition, property, and process parameters. These properties include experimental characterization results on the structural (topology and defect structure, stoichiometry, microstructure), electrical, thermomechanical (glass transition temperature, viscosity, fictive temperature), and mechanical properties (elastic modulus, stress-strain profile) of different glass compositions at different temperatures.

The chapter continues with Sect. 2, which is an overview of methods necessary for building data-driven models. The methods section is a tutorial on initial data modeling (unsupervised learning, cluster analysis, and regression), genetic algorithms (GA), and neural networks (NN). Next, in Sect. 3, the data-driven models are presented. In the first part of Sect. 3, neural network-based composition models for predicting glass liquidus temperature are presented and discussed. In the second part of Sect. 3, genetic algorithm-based glass models are presented for predicting Young's moduli for given glass compositions. Finally, in Sect. 4, new glass and other materials discovery opportunities with data-driven approaches are discussed.

The next section describes the materials modeling methods used for predicting new glass compositions.

2   Methods

In this section, the methods used for predicting glass properties from glass compositions are presented in four major steps (Dreyfus 2004): data consolidation and cleaning, initial data modeling, genetic algorithms, and neural networks. These statistical learning methods provide the basis for developing data-driven models for predicting properties for a given glass composition.

Data consolidation and cleaning: The predictive accuracy and robustness of any data-driven model depends on the accuracy and quality of the input data. In order to build the models in this study, proprietary Corning glass composition and property relations have been used. The predictive models for compositional and mechanical properties of glass use training, validation, and test data sets. These data sets are Corning's proprietary experimental data drawn from decades of glass composition research. Weight or atomic or molar percentages of each oxide constituting glass were used for training data set. For outputs, the viscosities η, Young's moduli E, glass transition temperatures T_g, densities ρ, as well as the liquidus temperatures T_L are stored. At Corning Glass Research, the composition experiments were done at different dates and by different staff. For each sample, an extensive sample preparation record (date, time, technician name, furnace name, process monitor parameters) is saved. On a later date, characterization reports (one for each of mechanical, electrical, chemical, rheological properties) have been recorded. These records were accumulated over multiple years, and many reports were entered manually into digital records. This step has been one of the most time-consuming aspects of the model as an automated processing for thousands of such reports was impossible. The reports had also gaps in entries and a data cleanup procedure needed to be developed. The data cleanup is an iterative process that is updated as more data were entered into the system.

Corning's earlier glass composition development experiments focused on refining composition toward improving the glass properties for a specific product at the time. Some products required specifications related to mechanical properties such as Young's modulus. The experiments then focused on Young's modulus and relevant mechanical properties. If a glass sample with a given composition fails to be within Young's modulus threshold, no further experiments were done on that sample and the other mechanical, chemical, or rheological properties of glass were not recorded. These missing entries must be handled either by using statistical methods such as interpolation of Young's modulus values between glasses with similar compositions or simply by ignoring the entire sample's data row. If the model requires numerous rows of missing entries (i.e., predicting viscosity without density or temperature or other relevant data), then the samples with missing rows are needed to be ignored. If the model results can have strong dependence on the preprocessing procedures for filling in the gaps, then the model is likely not going to be accurate. As a result,

the model accuracy was determined during these data curation steps long before the actual model parameters or architectures were determined. In each of these data consolidation and cleanup steps, the purpose is to keep reliable and accurate data and minimize useless data while preserving similar naming conventions.

2.1   Initial Data Modeling

Unsupervised learning is a group of initial algorithms used for input data sets without labeled response variables (i.e., properties). These algorithms allow a "divide-and-conquer" strategy for distinguishing misbehaving or outstanding data entries (if any). In addition, unsupervised learning methods help group different composition types and classify and predict based on their generalized behavior. For instance, this could be a systematic dopant dependence study of coefficient of thermal expansion (CTE) in alumino-borosilicate glasses within a more controlled and lower-dimensional composition space. Below, we describe cluster analysis, principle component analysis, and regression methods. These methods could be used to identify subgroups in the data sets to assess generalization capabilities of models.

Cluster analysis: The glass composition data could consist of more than ten dimensions where some parameters such as dopant fractions were systematically varied. Since manual identification of systematic variations in the input space is not feasible, a classification algorithm based on unsupervised learning has been used. This method is also known as feature selection. In unsupervised learning, one first calculates the proximity and similarity between each data point in the data and then defines a criterion function to maximize the quality of clustering based on the proximity and similarity metrics. Finally a clustering function is computed such that its values satisfy the criterion function for all data points when mapped to each data point in the data. These steps were followed to cluster experimental data:

1. Assemble composition-property database with composition entries, CTE, Young's moduli, and viscosities.
2. Normalize CTE, Young's moduli, and viscosity scales.
3. Describe a proximity criterion function (i.e., function whose output depends on the distance between data points) (Ullman et al. 2016).
4. The function gives numerical category values based on separation thresholds.
5. Customize the clustering function thresholds based on the desired properties and iterate until classification converges.

Here, we define a cluster as a collection of data point with "similar" features (i.e., composition, property, chemistry, etc.), and dissimilar data items are grouped under different clusters. Since the values in the glass composition data span more than 15 orders of magnitude, their absolute values cannot be directly used. For instance, there are significant scale differences among parameters such as CTE (10^{-7} to 10^{-6} $(°C)^{-1}$) and weight percentages (10^{-5}–10^{-1}) and Young's moduli ($\sim 10^{11}$ GPa) and

viscosities (10–10^5 Pa·s). In order not to cause one column to dominate distances based on their absolute values, each value is normalized to index values before calculating the distances. A conversion formula is defined in the beginning for each property or column for converting these normalized values back to their original scales. After normalization, the distances between each data point can be found either using Euclidean distance, which is translation-invariant, or using Manhattan distance or their generalized version which is Minkowski distance as discussed in more detail by Tandia et al. (2019) and Mauro et al. (2016).

After distances are calculated, similarity indices can also be calculated to specify if compositions are nearly identical. Then, a criterion function needs to be defined to set the rules for distinguishing different clusters. For instance, this function can impose multiple specific conditions, such as eliminating hazardous dopants like lead for a given application or requiring lead for an x-ray shield glass application ($x_{Pb} > 10^{-6}$), and impose distance threshold conditions based on overall calculated separations or material property differences (high versus low viscosity samples, etc.). The definitions of the clustering criteria are the points where materials engineering and specifications for the given application may also be built into the models. Since models can later be trained with subsets of clusters, one may filter for data in model training and testing stage as well. After defining cluster criteria, a clustering function is calculated such that all data entries are grouped to satisfy the criteria. This clustering function can be implemented with MATLAB, R, or Python. The clustering algorithms can be grouped in three types: hierarchical, partitional, and Bayesian. In the hierarchical approach (i.e., inductive method), each data row can start as a separate cluster, and based on differences and distance criteria, clusters can be merged. In the divisive hierarchical approach (i.e., deductive method), all data rows belong to one cluster and can then differentiate based on distance criteria and similarity metrics. The details of these algorithms are presented by Tandia et al. (2019) and Mauro et al. (2016).

While implementing clustering algorithms, one must make sure that the clusters are not sensitive to random seeds. Cluster analysis should be repeated sufficiently many times to prevent different random seeds from altering the way the data are clustered. In addition, outlier data points from this clustering calculation can be separately analyzed by glass scientists. These points could be removed from analysis to prevent any unintentional bias in cluster centroids. Visualization methods may help in reporting clusters and identifying outliers.

Principal component analysis (PCA): PCA is a dimension reduction technique that allows omitting variables, which do not have significant effect on the desired glass property (Jolliffe 2002; Ringner 2008). This method uses an orthogonal mathematical transformation on data to convert a mapping of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. The first principal component has the largest possible variance, and every next component has progressively lower variances. The output of PCA is a set of vectors transformed out of the original data. These vectors form an

uncorrelated orthogonal basis set. As variances depend on the absolute mean values of each variable, PCA output vectors strongly depend on the relative scaling of the initial variables. Therefore, normalizing each variable is essential before running PCA calculations.

One can view PCA as fitting a multidimensional ellipsoid to the data. The axes of the ellipsoid are called principal components. Whenever the data have weak dependence on an independent variable, the variance with that variable is small, and the corresponding axis of the ellipse is small. As a result, ignoring that variable of **ellipsoid** is not going to cause a significant loss in information. Instead, this simplification reduces the number of variables and reduces the eventual model complexity. The detailed steps for implementing PCA are presented by Tandia et al. (2019).

After PCA, one can define a variance threshold for an eigenvector to be accepted as significant and pick only the eigenvectors with variances above the threshold. This threshold step helps researchers determine the most significant variables in a glass composition project and focus on cluster analysis using those fewer and more important variables (Kullback and Leibler 1951; Kullback 1959). While PCA reduces computational costs, consider the following limitations of PCA in initial data processing:

1. **Initial data scaling has strong impact on PCA vector variances:** The relative ranges and absolute values of each variable can distort the PCA ellipsoid and alter the covariance matrix. The principal components must be normalized properly to avoid such distortion.
2. **PCA relies on linear variances:** PCA cannot capture information gain from variables which have complex nonlinear correlations. As a result, PCA may add spurious variables. Nonlinear correlations must also be considered.
3. **PCA may fail at blind signal separation cases:** PCA assumes that when a variable has a large variance, the variable is going to have low covariance hence high importance. This assumption helps eliminate the noise and pick up the major variables. In specific problems called blind signal separation (or blind source separation) however, multiple signal sources can have almost equally significant effects and lead to a mixed signal response. In such multiple source effects that may lead to competing driving forces, one must run an independent component analysis (ICA) in addition to PCA.
4. **Mean and covariance do not describe some distributions:** Mean and covariance values generally describe Gaussian distributions, but there are many common statistical distributions in which mean and covariance do not describe the distribution functions. Other ways of testing significance of variables such as information gain analysis (also known as Kullback-Leibler divergence or Gram-Schmidt) must also be used.

Regression analysis: Since PCA may fail to capture nonlinear correlations among variables, regression analysis is necessary to distinguish if there are hidden

complex dependencies. In regression analysis, the conditional expectation (E(X|H) of variable X given H) is estimated when the independent variables are given. In other words, either a generalized linear model is constructed by using a multivariable linear superposition of independent variables or a least squares fit is solved numerically among variables to identify if there is a nonlinear correlation. Most common regression methods are presented below:

Multivariable linear regression: This is a generalized linearization method that assumes one or more independent variables:

$$Y_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \cdots + \alpha_n X_{in} + \epsilon_i,$$

where X_{i1} to X_{in} are independent variables and ϵ_i is an error term which is minimized until the optimal set of regression coefficients α_0 to α_n are found. These coefficients also indicate the relative significance of each independent variable.

Least squares regression model: This fitting model assumes a sophisticated nonlinear correlation between variables. As no closed-form solutions are assumed, one may pick a Taylor expansion or a different basis set to fit the dependent variable behavior by minimizing the sum of the squared residuals S:

$$S = \sum_{i=1}^{n} (y_i - f(x_i, \beta))^2$$

Domain expert interpretation of statistically identified causalities: Although correlations can be identified using regression analysis, identifying causality requires an extra layer of physical interpretation by glass scientists. Linear and nonlinear regression models can be implemented in MATLAB, R, or Python.

Genetic algorithms: Genetic algorithms (GA) are methods for solving constrained optimization problems based on an evolutionary natural selection process. In a classical numerical solution algorithm (i.e., Newton's method solving the zeros of equations), one generates a single point at each iteration. The sequence of points progressively approaches the solution. In GA, instead of iterating only on one point, a population of points are generated at each iteration and the best point in population approaches the solution.

A fundamental difference between classical algorithms and GA is that in GA, the next population is generated using random number generators and some combinations from the best solutions of the previous iteration instead of deterministic computation. GA are useful for accelerated global optimization problems where there are multiple objectives with competing requirements. GA can also be used for maximization, minimization, constrained optimization, multidimensional fitting, classification, and prediction problems. In this chapter, glass models are optimized with respect to multiple competing objectives. As a result, GA are particularly beneficial for composition development. Below is a description of a simplified GA for solving a generic multi-objective optimization problem:

1. The GA generates a random initial population of glass compositions.
2. The GA then evaluates the glass composition population based on their predicted properties and the fitness values:
   (a) Scores each composition of the current population by computing its fitness value and normalizing these scores for comparisons.
   (b) Picks a few candidates from the population (parents) based on their fitness.
   (c) Some of the individuals in the current population that have lower fitness are chosen as elite. These elite individuals are passed to the next population (children).
   (d) Children are produced either by making random changes to a single parent (mutation) or by combining the vector entries of a pair of parents (crossover).
   (e) Replaces the current population with the children to form the next generation.
3. The GA then creates a sequence of new populations. At each step, the GA uses the compositions in the current generation to create the next population.
4. The algorithm stops when the output meets the stopping criteria. The stopping criteria may include a maximum number of generations, time limit for iterations, fitness limit for population (when the fitness function for the best point in the population is less than or equal to the fitness limit), and marginal change limits per generation (i.e., when the average change over each generation is less than a functional tolerance limit). GA can be implemented in MATLAB, R, and Python.

Neural Networks: Neural networks (NN) are a set of machine learning algorithms (supervised or unsupervised) that can be used to predict nonlinear behavior of different systems. A neuron is a nonlinear and parameterized function of input variables. Therefore, a neural network is the composition of nonlinear functions of two or more neurons. The nonlinearity of neural networks can help model nonlinear behaviors that may not otherwise be captured properly in regression models or other linear techniques such as PCA. Despite the biological flavor of the term neural network, neural networks in machine learning are pure mathematical constructs that consist of either feedforward or feedback networks (recurrent). For glass modeling, we focus on feedforward networks.

A feedforward neural network (static network) is a nonlinear function of its inputs and their corresponding weighting parameters. So a feedforward network consists of functions of its neurons which are nonlinear functions of inputs.

Figure 1 shows a representation of a feedforward neural network in which information flows only in forward direction. In this graph, the vertices are the neurons, while the edges are the connections. Different from feedback networks, feedforward NN have a noncyclic graph topology. The neurons at the outputs are called output neurons, and the rest of the layers between the inputs and the output neurons are hidden networks. In an optimized NN model, a sequence of neurons may also yield physical insight about a given physical parameter and help develop physics-based models later.

Feedforward networks are static and their outputs are constant when their inputs are constant. A mathematical representation of the single-layer static network in Fig. 1 that uses nonlinear sigmoid activation functions is:

Fig. 1 Feedforward neural network, which consists of multiple n inputs, N_c hidden neurons and N_o output neurons

N inputs   Hidden   Output
                  neurons   neurons
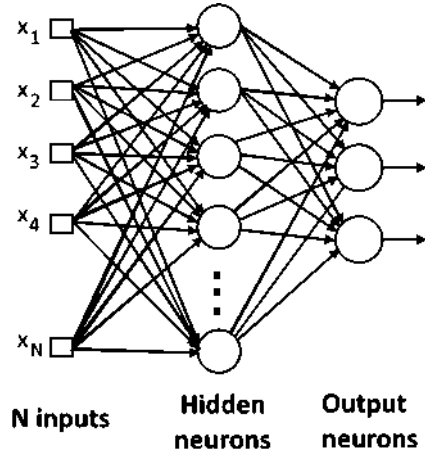
$$g\left(\mathbf{x}, \mathbf{w}\right) = w_{N_c+1,0} + \sum_{i=1}^{N_c} \left[w_{N_c+1,i} \cdot \tanh\left(\sum_{j=1}^{n} \left(w_{ij}x_j + w_{i0} \right) \right) \right]$$

where $\mathbf{x}$ is the input vector (size: n elements) and $\mathbf{w}$ is a vector of size $(n + 1)N_c + (N_c + 1)$. The hidden neurons are numbered from 1 to N_c, and the output neuron is indexed as $N_c + 1$. The parameter w_{ij} is assigned to a connection from neuron (or input edge) j to neuron i. The output function g(x,w) is a linear function of the parameters in the last connection layer (between the final hidden layer to the output neuron), and the connections between the inputs and the hidden layers are nonlinear functions of the inputs. In g($\mathbf{x}$, $\mathbf{w}$), the activation function is a hyperbolic tangent although other activation functions such as step, linear, sigmoid, rectified linear, and exponential functions can be used.

3   Data-Driven Development for Glass Composition Design

The design of glass compositions for a targeted application is a long process due to many product specifications and the manufacturing technology. Glass specifications vary among different applications. Many customer requirements for display glasses include strain point (SP), annealing point (AP), softening point (SfP), coefficient of thermal expansion (CTE), density, Young's modulus (E), dimensional stability, and fracture toughness (FT) for strengthened glasses. Corning Incorporated use the fusion draw technology to deliver surface pristine glass products. This technology imposes manufacturability requirements for a given glass composition. Some of the most important ones are the liquidus temperature, the viscosity at the liquidus temperature, and the temperatures at 20 Pa·s and 3500 Pa·s.

Glass composition design can be posed as a multi-objective optimization problem with constrains on the sum of the variables (different oxides that make up the glass

composition) and the range of each variable (oxide). One objective function for each attribute can be given as a range (70 GPa $\leq$ E $\leq$ 85 GPa) or lower-bounded range (AP $\geq$ 800 °C). Corning's proprietary glass composition and property database developed over many years span key constituents, and linear regression models are used whenever existing experimental data may coincide with the desired properties. For conflicting constraints or nonlinear material behavior, nonlinear optimization and machine learning models must be used.

Liquidus temperature (T_L) and viscosity are two important properties for glass design. T_L of glass is defined as the temperature at which the first crystalline phase precipitates from a glass melt of a given composition when it is cooled very slowly. The accurate knowledge of the T_L is critical for both glass property prediction and manufacturing purposes. During glass manufacturing using an isopipe, as used by Corning, knowing the T_L is important when the melt is flowing on the isopipe surface. The temperature of the melt can be maintained above the T_L of the melt to avoid its crystallization on the isopipe. If crystalline phases nucleate, the isopipe, which is a multimillion-dollar equipment, can be damaged. Depending on the desired sheet thickness, the melt must have a viscosity at the liquidus higher than a required threshold. This last requirement necessitates the accurate knowledge of the viscosity-temperature curve for the glass melt. Both T_L and viscosity-temperature curve are critical melt properties tightly related to the manufacture of surface pristine glass sheets when drawn from an isopipe. For display applications or for ion-exchange glasses for damage-resistant layer, the knowledge of the viscosity-temperature is also important for the final glass attributes or process parameters such as annealing temperature, softening temperature, or the viscosity at a given temperature. Despite the necessity for the knowledge of these parameters for industrial glasses in display applications, there is no known accurate or generalizable physics-based models for glass melt T_L or melt viscosity. Therefore, using Corning's large glass composition database and robust machine learning techniques, accurate predictive models for both liquidus and viscosity have been developed. In this chapter, we present the model development for T_L and Young's moduli. The viscosity model is presented in Springer Handbook of Glass as a separate chapter (Tandia et al. 2019).

One of the approaches to finding glass compositions that meet both material specifications and manufacturing requirements is to use a fine grid search on the composition space and compute each property at the grid. Developing accurate and robust ML models and running on this grid is the challenge addressed in this section. Most ML methods consist of many parameters. For neural networks, one has to optimize the number of layers, the number of neurons in each layer, and the functional form to use for the activation function for the most accurate predictions. Depending on the type of optimizer to be used, additional parameters such as learning rate and momentum need to be added in the list of parameters. For genetic algorithm, key parameters to optimize for during the model design are population size, different probabilities for crossover, and mutation, along with the fraction of new random solutions to be generated at each iteration. Optimizing the model parameters through an exhaustive search for the best architecture for either

the NN or GA is prohibitively time-consuming due to the parameter space size and due to continuous parameters such as learning rate. Although one can use Monte Carlo methods, i.e., randomly select values/states for the model parameters and run the decided number of simulations, one cannot guarantee that a global optimum can be achieved by random parameter selection. In this study, we use Bayesian optimization (BO), which is a method of finding the maximum/minimum of cost functions (Shahriari et al. 2016). BO is based on the Bayesian method to set up a priori over the model objective function and combine it with evidence to find the posterior function. BO is applicable to cases where we do not have a closed-form formula for the objective function, but where we can collect observations of the objective function.

In the following subsections, we describe the composition-dependent models for liquidus temperature and Young's modulus.

3.1   Neural Network-Based Composition Models for Predicting Glass Liquidus Temperature

The liquidus data used in this study were sampled from Corning's internal liquidus data, which represents a historical collection with clusters of different sizes. First, a neural network model for liquidus temperature was build using the whole data set. The initial liquidus model is built using the flowchart in Fig. 2. In model development, one needs to split data into training, validation, and test sets. Each subset must have a similar statistical distribution as the whole set. Methods such as Kullback-Liebler divergence can be used in testing if two distributions are different (i.e., divergence is 0 if statistical distributions are similar). The design of optimal neural networks architecture is going to be based on the incremental increase of

Fig. 2 Flowchart describing the steps for the liquidus model development
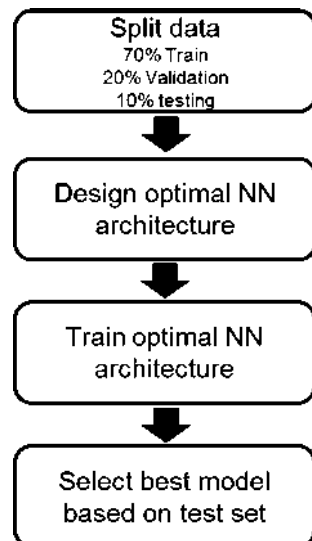
model complexity through more neurons in any given layer or an increase of the numbers of hidden layers. For each NN architecture with given parameters, a series of training steps are conducted. Next, the root mean square error (RMSE) and R^2 on the test set and validation set are recorded. R-squared (R^2), also known as the coefficient of determination, is a statistical measure of how close the experimental and predicted data are to each other. The optimal architecture is the one with the smallest number of parameters (least complexity) and for which the RMSE and R^2 are about the same for all three sets (training, validation, and testing). One may also systematically study the effect of the choice of the activation function for each layer in the NN as part of the optimization process, except the output neuron, which is kept as a linear activation function.

Once an optimal NN architecture is identified, the model is then trained for a few hundred times. The model is trained to increase the likelihood of finding a set of parameters that give an output set of parameters that minimize the objective function. For this training, we found that the optimal architecture has two layers (first, ten neurons; second, eight neurons), and both layers have a Gaussian activation function. On Fig. 3, a close look in the plot of the predicted liquidus for the test set suggests that there are points with high accuracy and others with low
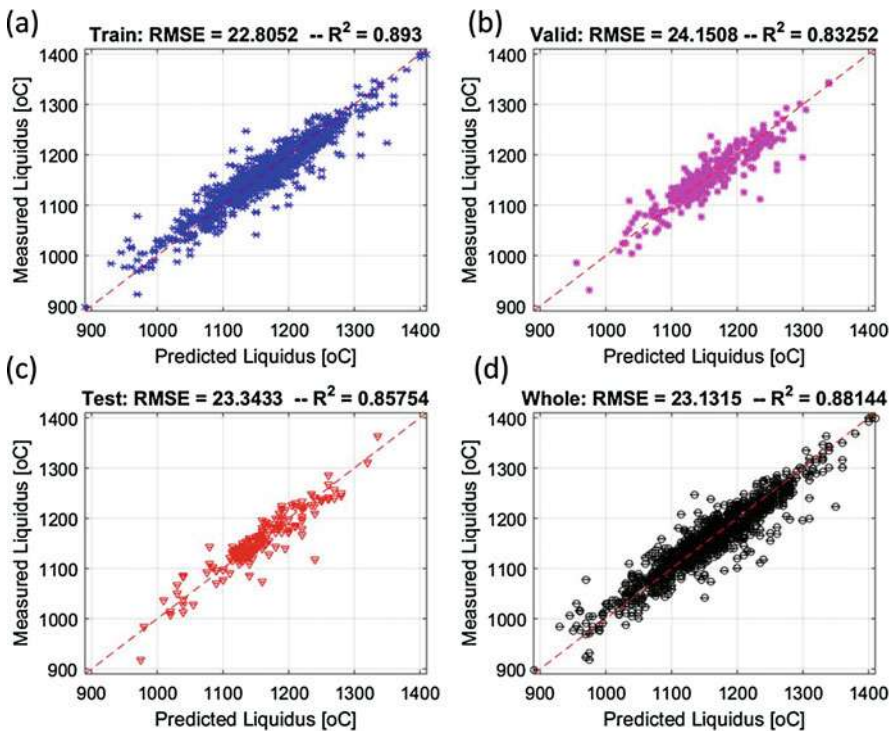
Fig. 3 Best liquidus model based on the whole data set. RMSE on all sets are about the same but higher than the experimental values ($\sim$10 °C). (**a**) represents the training data, (**b**) validation data, (**c**) the test set or blind prediction, and (**d**) the whole data set

accuracy. Given that the original data is historical data and were not based on a design of experiments, one may expect that the original data might be a clustered data set.

We ran a k-means cluster analysis to better understand how the data are organized. Since we do not know a priori the number of clusters to consider, we will assume the number of clusters between 2 and 15. The choice of the optimal number of clusters to consider will depend on how well the chosen clusters separate from each other. This is determined by the similarity between elements of the same cluster. In 1986, Peter J. Rousseeuw defined a metric called silhouette, which is a robust way of interpretation and validation of regularity within clusters of data. The technique provides a concise graphical representation of how well each object lies within its cluster or how similar an object is to other members in its own cluster compared to other clusters. The silhouette ranges from −1 to 1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is suitable. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters. The silhouette can be calculated with any distance metric. Average silhouette in each cluster for all 14 clusters we consider is plotted in Fig. 4. It appears that the data is best split in nine clusters. Further analysis of the different clusters shows the cluster's population.
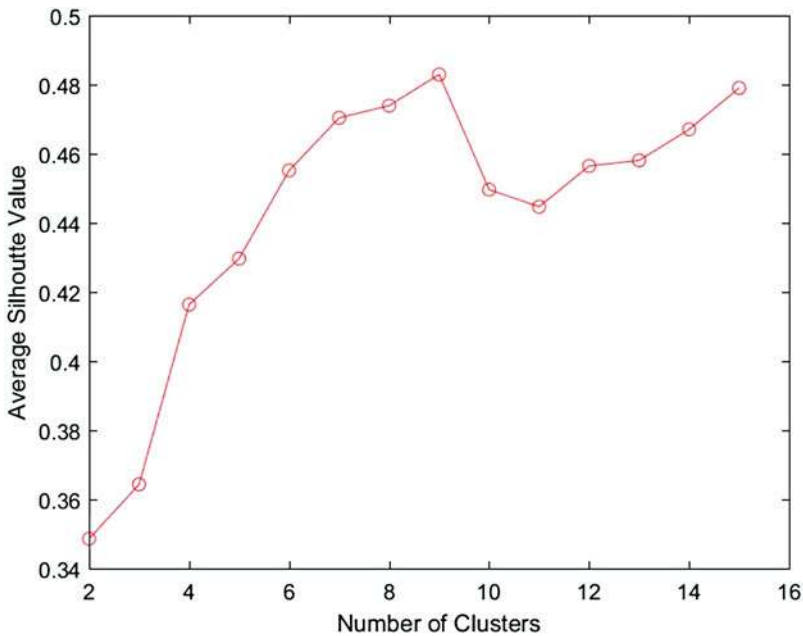
Fig. 4 Average silhouette value for all 14 clusters considered in the k-means cluster analysis

Fig. 5 (**a**) Clusters population for liquidus temperature data set. (**b**) Average distance to centroid for all 14 clusters we considered in the k-means cluster analysis
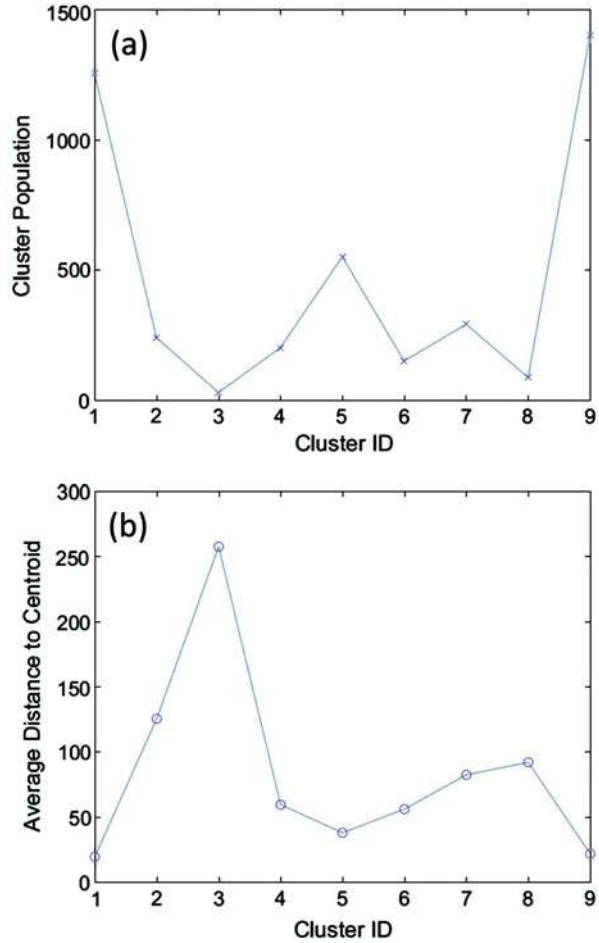
For the liquidus data, the clusters have different population sizes. In Fig. 5a, two clusters stand out with populations higher than 1100 compositions. The third largest cluster has about 500 compositions. The six remaining cases have populations lower than 300 compositions. Other than the cluster population, the average distance of a cluster population to the cluster centroid also characterizes the cluster. The smaller the average distance the more compact the cluster is. We observe in Fig. 5b that the two most populated clusters (1, 9) have the smallest average distances to their respective centroids. Being most populated and most compact, clusters 1 and 9 are suitable to be used for building localized liquidus models. Using the same procedure for splitting the data into training, validation, and test sets, we conducted a best architecture search for data in cluster 1. We found the same architecture as when we used the whole set although this is not necessarily the case for all cases.
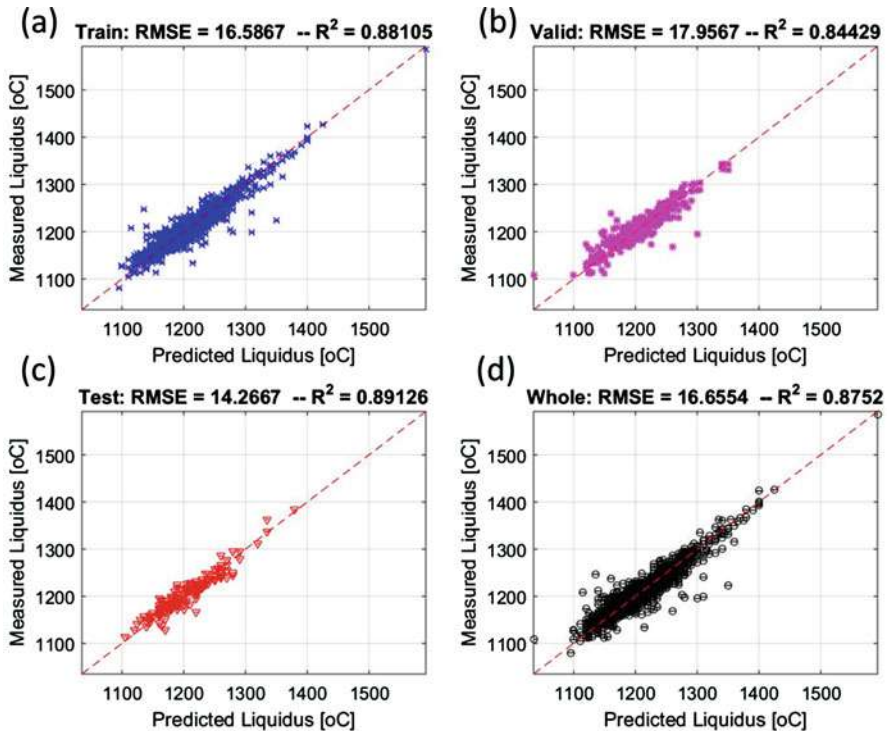
Fig. 6 Best liquidus model based on the data in the largest cluster. RMSE on all sets are about the same, within few degrees of difference. They are higher than the experimental values ($\sim$10 °C) but much better than the RMSE we got when we used the whole data set to build a global model. (**a**), (**b**), (**c**), and (**d**) represent the training data, the validation data, the test set or blind prediction, and the whole data set, respectively

We see from Fig. 6 a significant improvement of the predicted liquidus temperature for the test sets. Over all the RMSE of the test set decreased significantly with the use of the largest cluster (14 °C), compared to (23 °C) when we used the whole data set. The liquidus model we built from the second largest cluster also has a lower RMSE on the test set (15 °C). To make liquidus temperature predictions from both models will require that the composition of interest to be a member of cluster 1 or cluster 9. This is very likely to be the case since these clusters are the largest ones in the set. When using this approach for designing glass compositions with low liquidus temperature, one has to start by selecting compositions that are only in cluster 1 or cluster 9 and then use the corresponding cluster based model to make the prediction.

We now consider a case in which we have many different models from N_c different clusters. One can consider an ensemble prediction which consists of building a model for each cluster, make a prediction from each cluster based model, and take the ensemble average computed as the mean predicted value. This mean

predicted value is the average from the cluster-based predictions over the distance from the composition of interest to the centroid of each cluster:

$$T_{Liq} = \frac{\sum\limits_{i=1}^{Nc} \frac{T_i}{d_i}}{\sum\limits_{i=1}^{Nc} \frac{1}{d_i}}$$

N_c is the total number of clusters for which we have built a liquidus model from, T_i is the liquidus predicted by the model built from cluster i, while d_i is the Euclidian distance from the point of interest to the centroid of the cluster i. The coordinates of the composition can be in mole percent (mol %), weight percent (w %), or atomic percent (at %) depending on the coordinates that was used to do the cluster analysis.

Based on the above model development description, optimizing the NN model architecture is tedious. We have searched for the best architecture by starting with a simple one-layer model with a small number of neurons, and we increase its size one or two neurons at a time for a fixed activation function. We increased the number of neurons depending on the ratio between the RMSE of the training and validation sets, which we want to keep close to one. At some point, we can insert a second hidden layer to resolve the nonlinearity of the liquidus temperature. Let us consider for a moment the number of variables we can play with in the design of the neural network architecture, without being explicit:

- Number of hidden layers (1, 2) and number of neurons in each layer (6 to 20)
- Activation function of each layer (tanh, sigmoid, inverse tangent)
- Learning rate (continuous number between 0.0001 to 0.1)

  Simulating all possible architectures is a prohibitively expensive computational task. Although one could randomly select a few architecture cases to run from the whole parameter composition space, there is no guarantee of having the optimal architecture in the selected cases. Ideally, one would like to find the optimal neural network architecture without having to run thousands of simulations. Bayesian optimization (BO) offers that possibility. In this step, by using BO, we rebuild a liquidus model by splitting the data into two sets: a training and a validation set. BO enables an accelerated optimization of the architecture with its modeling parameters such as the number of layers, neurons in each layer, the learning rate, and the activation function for each hidden layer.

3.2   Genetic Algorithm-Based Glass Models for Predicting Compositions for Desired Ranges of Young's Moduli

Young's modulus is a critical glass property for many different applications. When alkali-free glasses are used in the thin-film transistor manufacturing process, they are sought to have high Young's modulus to help reduce stress-induced distortions.

RMSE=11.9°C R²=96.4%
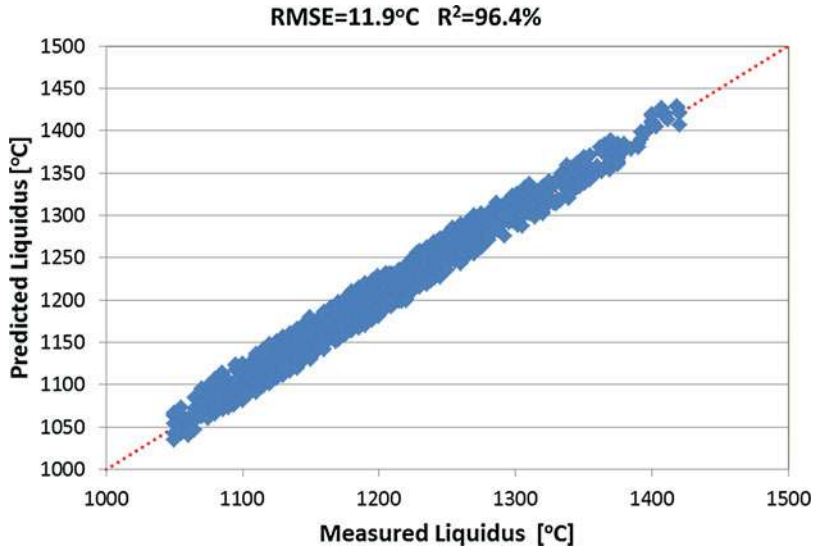
Fig. 7 Best liquidus model based on the validation set. Use of the Bayesian optimization framework helped improve the RMSE and the R^2 on the validation set

In Corning's design of glass compositions with native damage resistance, composi-tions with low Young's modulus are of critical interest. Glass compressive stress due to ion exchange is given by formula in which Young's modulus is a key contributor:

$$CS = \frac{BE\Delta C}{1 - \alpha}$$

B is the network dilation coefficient, E Young's modulus, ΔC an atomic concen-tration variation, and α the glass Poisson ratio. If everything is kept constant, the higher the Young's modulus, the higher the compressive stress CS.

The development of an accurate and robust model for Young's modulus is very important for glass composition design for different applications. The model's accuracy and robustness is important. Young's modulus was not measured for each glass that had been melted. This lack of systematic measurement led to scarcity of Young's modulus data. As a result, the extrapolation of any Young's modulus model can be a major challenge. Neural networks typically do not extrapolate very well when the input deviates too far from the training data. One cannot extrapolate over composition space as no composition-dependent physics-based model exists yet.

With proper parameter choices and optimization however, one can build genetic algorithmic models with much better extrapolation capability than neural networks. Genetic algorithms (GA) converge slowly, and many more iterations are required when building models. The model's accuracy and robustness depends on the size of the population under consideration and the number of trials. To build GA models,
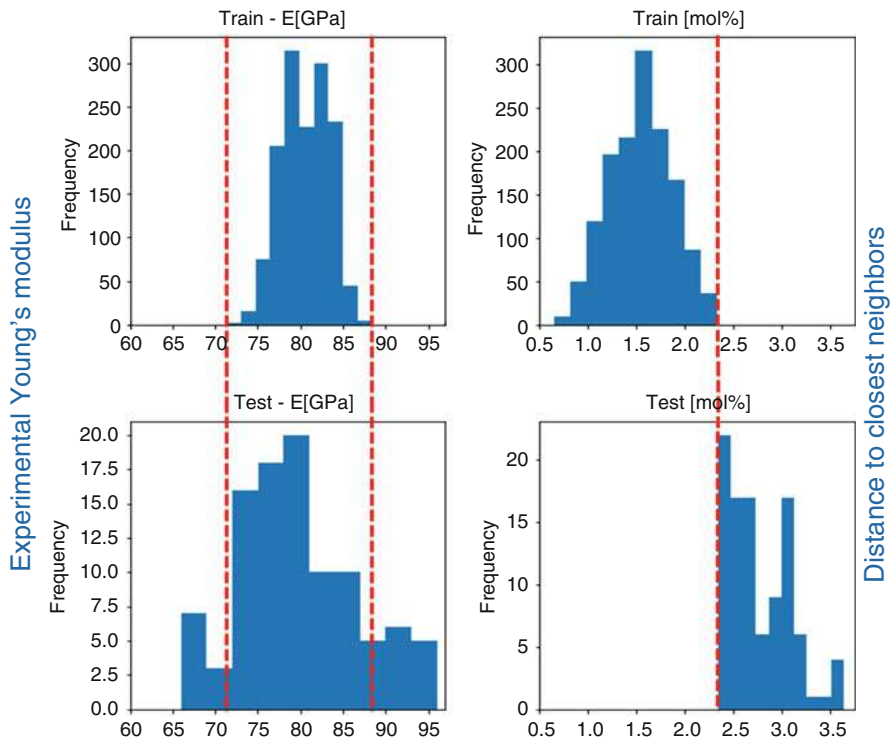
Fig. 8 Modeling of Young's modulus with genetic algorithm. Horizontal axis on the right-hand side is for the distance to the closest neighbor in the whole set. Compositions we selected for the test set are the ones very far from their closest neighbors meaning the ones at the boundary. A look at their corresponding Young's modulus shows that they have higher Young's modulus

we used an in-house GA modeling tool similar to Python. Once we constrained the tool to simple models by penalizing complexity and large number of parameters of long and complex solutions, we find more accurate Young's modulus models, which extrapolate over the training data set. One extreme situation to build such models where to extract data that are at the boundary of the whole data set and use it as a test set for validation and selection of the best model (Figs. 7, 8, 9, and 10).

Overall, we see that GA makes very accurate extrapolated predictions. If we use the same data splitting with a neural network model, the results are not as accurate, as one can see in Fig. 11.

4   Conclusions and New Glass Modeling Opportunities

In this chapter, we presented an overview of glass material composition and property modeling based on GA and NN. These models were used to develop glass compositions that achieve low-cost and high-volume manufacturing and the desired
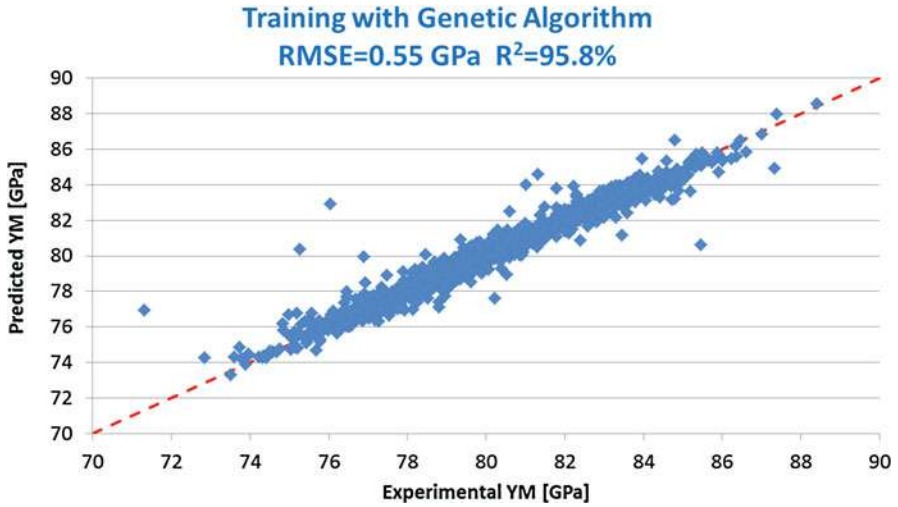
Fig. 9 GA makes good predictions for the compositions in the training set
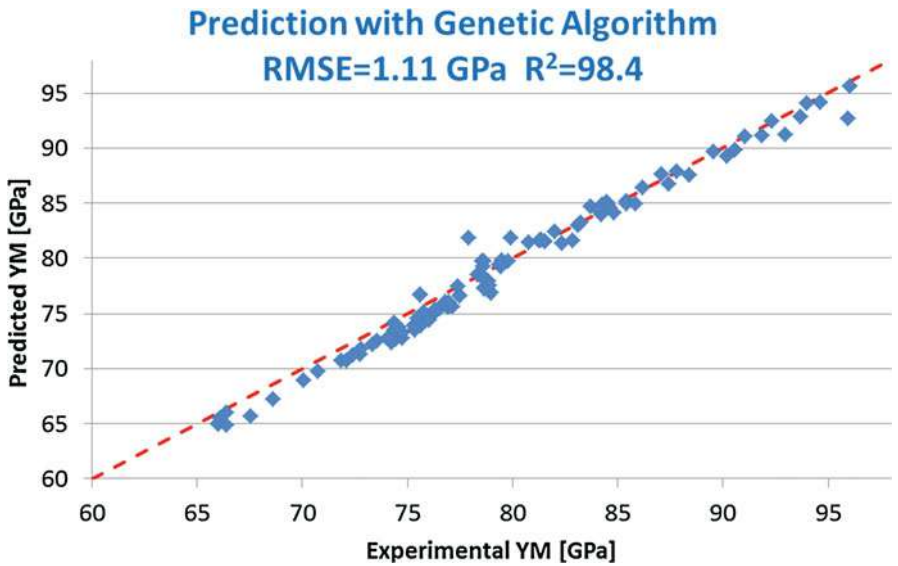
Fig. 10 GA makes good predictions for the compositions in an extrapolated region

Young's modulus. By the end of 2015, Corning's damage-resistant glass products have been in over 4.5 billion devices including smartphones, tablets, smartwatches, automotive glass, and in interior architecture. Billions of display devices are expected to use Corning Gorilla® Glass as display substrates (Materials Innovation
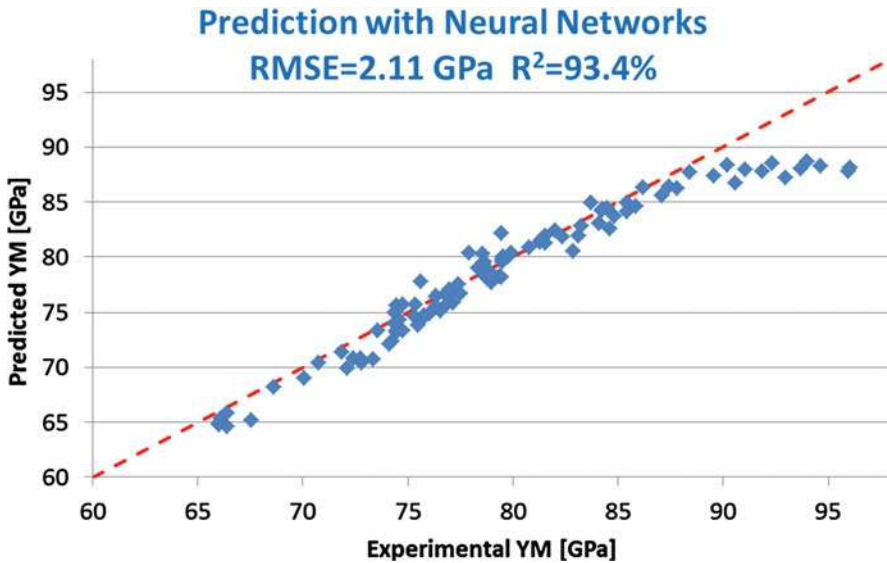
Fig. 11 NN makes poor predictions for the compositions in an extrapolated region

Case Study 2016). Data-driven models for CS, as well as integrated computational materials engineering (ICME) approach that incorporated both materials design for performance and manufacturing-related models, are key ways to consider in design of glass compositions for damage resistance applications.

Using data-driven GA and NN models, one can predict and engineer glass liquidus temperature and Young's modulus as discussed in Sect. 3. Data-driven glass modeling and prediction allows researchers to accelerate the materials invention process by computationally screening tens of thousands of different chemical compositions and their corresponding predicted or measured properties.

For a display application with a desired set of glass property specifications, a database of materials properties must be constructed by using Corning's internal experimental database of glass compositions and structural, optical, mechanical, and rheological properties. Next, data simplification methods such as PCA, ICA, data cleaning, cluster, and regression analysis techniques should be used to identify the properties that have the most significant effects on the desired end properties. Finally, GA and NN models should be used to predict the properties of a given stoichiometry over a target chemical composition space to identify (new) compositions with optimal mechanical and rheological properties. Rheological properties such as viscosity and liquidus temperature determine the high-volume and low-cost manufacturability of glass. Young's modulus, chemical composition, and other mechanical and structural properties determine how well the glass composition meets the desired specifications. Eliminating a subgroup in a composition space (i.e., due to topology restrictions, known behavior of dopants, etc.) can help simplify the NN and GA models' parametric sweep over the phase diagram.

In terms of fundamental glass modeling, we believe that data-driven predictive methods for chemical compositions and their corresponding physical properties offer the following opportunities (Mauro 2014):

New glass compositions with minimal crack propagation: Damage-resistant glass compositions are engineered to prevent crack propagation and eliminate failure modes induced by external impact. Chemical strengthening on the surface of glass reduces stress concentration and the likelihood of crack propagation. New failure modes are emerging with new use cases. Different form factors such as automobile windshield, smartwatch, tablet, or smartphone display glasses cause differences in stress distribution and dopant concentration profiles. Enhancing glass strength both through the bulk glass composition and through the surface dopant profile is needed for display applications. Data-driven predictive modeling can help identify new surface dopants, bulk chemical compositions, and the effect of any other post-processing recipes and their parameters that can enhance the ductility and fracture toughness of glass.

Borate and phosphate glass chemistry: Major oxide glasses such as borate and phosphates offer interesting scientific challenges as well as useful technological applications. Different effects can be observed when incorporating these network-forming oxides into a silicate glass matrix. Modeling approaches similar to this chapter can help identify structure-property relations previously not known for these less explored glass chemistries.

Nontraditional glass chemistries: Bulk metallic and chalcogenide glasses have attracted significant interest due to their unique functionalities. Metallic glasses have brittle-to-ductile transition, which yields increased fracture toughness. Chalcogenide glasses have unique optical properties for infrared bands. These materials are fundamentally different from silicate, phosphate, or borate glasses in terms of their chemistries and optical, thermal, and mechanical properties. Since the length and timescales relevant for the functionalities of these materials can span more than nine orders of magnitude, molecular dynamics or other approaches are not feasible modeling approaches for engineering or understanding the properties of these materials. Data-driven approaches can help identify the compositions that can optimize multiple functionalities with respect to a range of elemental constituents whose behaviors are partially understood.

High pressure glass formation: Corning uses fusion drawing of glass melt at high temperature and reasonably lower viscosities for large volume manufacturing of state-of-the-art glasses (Mauro et al. 2009; Vogel 1921). A fundamental and practical open research area is high-pressure glass formation, which occurs for volcanic and magmatic systems. Today's glass production is done at ambient pressure. As high-pressure glass fabrication experiments can be expensive, data-driven modeling of volcanic or magmatic glasses may help develop pressure-based glass fabrication methods. These methods may enhance new combinations of glass properties, which may not necessarily form by thermal history.

Polyamorphism: In polyamorphism, two glasses of identical chemistries may display different short-range structural ordering. The short-range order differences significantly alter the observed properties. Since the glass chemistries which may

exhibit polyamorphism are not well understood, one may construct a data set of polyamorph glasses and their corresponding chemistries. Then running a data-driven model can help identify the most important parameters and defect chemistries and conditions that may alter short-range order for identical chemistries. As a result, one may predict mechanical and rheological properties based on data-driven polyamorph glass models.

References

Curtarolo S, Hart GLW, Nardelli MB, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12:191–201

Dreyfus G (2004) Neural networks, methodology and applications, 2nd edn. Springer, Berlin

Jolliffe IT (2002) Principal component analysis. Springer, New York

Kullback S (1959) Information theory and statistics. Wiley, New York

Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22:79–86

Materials Innovation Case Study: Corning's Gorilla glass 3 for consumer electronics. https://www.nist.gov/system/files/documents/2018/06/26/materials_innovation_case_study_gorilla_glass_3_020816.pdf

Mauro JC (2014) Grand challenges in glass science. Front Mater 1:20

Mauro JC, Yue Y, Ellison A, Gupta PK, Allan DC (2009) Viscosity of glass forming liquids. PNAS 160:19780–19784

Mauro JC, Tandia A, Vargheese KD, Mauro YZ, Smedskjaer MM (2016) Accelerating the design of functional glasses through modeling. Chem Mater 28:4267–4277

Mueller T, Kusne AG, Ramprasad R (2015) Chapter 4: Machine learning in materials science: recent progress and emerging applications In: Parrill AL, Lipkowitz KB (eds) Reviews in computational chemistry. 1st edn. Rev. Comput. Chem. 2016, vol 29. Wiley, Hoboken, pp 186–273

Pilania G, Wang C, Jiang X, Rajasekaran S, Ramprasad R (2013) Accelerating materials property predictions using machine learning. Sci Rep 3:2810

Potyrailo R, Rajan K, Stoewe K, Takeuchi I, Chrisholm B, Lam H (2011) Combinatorial and high-throughput screening of materials libraries: review of state of the art. ACS Comb Sci 13:579–633

Ringner M (2008) What is principal component analysis? Nat Biotechnol 26:303–304

Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N (2016) Taking the human out of the loop: a review of Bayesian optimization. Proc IEEE 104:148–175

Sullivan EC, Taylor WC (1919) Glass compositions containing silica with 40% to 90% silica, by weight containing boron containing aluminium. US Patent 1,304,623

Tandia A, Onbasli MC, Mauro JC (2019) Machine learning for glass. In: Musgraves JD, Hu J, Calvez L (eds) Springer Handbook of Glass. Springer, Cham, pp 1157–1192

Ullman S, Poggio T, Harari D, Zysman D, Seibert D. Massachusetts Institute of Technology 9.54: computational aspects of biological learning, Fall 2014 course notes. Retrieved 17 Sept 2016 from http://www.mit.edu/~9.54/fall14/slides/Class13.pdf

Vogel H (1921) Das Temperaturabhängigkeitsgesetz der Viskosität von Flüssigkeiten. Phys Z 22:645–646

Xiang XD, Sun X, Briceno G, Lou Y, Wang KA, Chang H, Wallace-Freedman WG, Chen SW, Schultz PG (1995) A combinatorial approach to materials discovery. Science 268:1738–1740

Constitutive Modeling in Metallic Glasses for Predictions and Designs

82

Weidong Li

Contents

Abstract

Metallic glasses, made of metallic elements yet possessing an amorphous structure, have attracted considerable interest due to their unique mechanical properties and excellent net-shape forming ability. In terms of mechanical properties, they are extraordinary in elastic limit and strength but inferior in ductility at room temperature, where the inhomogeneous deformation characteristics of shear banding, flow serration, and sometimes cavitation prevail. The net-shape forming ability surfaces when the temperature rises above T_g (i.e., the glass transition temperature), where metallic glasses deform in a homogeneous manner. Understanding both the homogeneous and inhomogeneous deformation in metallic glasses is pivotal for promoting their potential applications. In this

W. Li (✉)
Department of Materials Science and Engineering, The University of Tennessee, Knoxville, TN, USA
e-mail: wli20@utk.edu

course, the continuum mechanics constitutive modeling has played a critical role. The constitutive modeling can be used for either predicting the mechanical responses of metallic glasses under various loading conditions or guiding material, product, and process designs. In this chapter, a number of important constitutive models for metallic glasses are reviewed; their applications in predicting both the homogeneous and inhomogeneous plastic flows are presented; and finally, the attempts of using the constitutive modeling to assist the design of ductility and net-shape forming processes are discussed.

1   Introduction to Metallic Glasses

Depending on the composition and cooling rate, metallic melts could solidify in either a crystalline or an amorphous form, which are two completely distinct solid states. The former arranges atoms in an ordered lattice structure whereas the latter in a disordered configuration. While the use of crystalline metals has accompanied the human history and civilization, the first metallic glass (or termed amorphous alloy) was synthesized in 1960, in the thin-ribbon Au-Si system cooled at a very high rate of $\sim$ 105–106 K·s^{-1} (Klement et al. 1960). This field was not fast-moving until the late 1980s and early 1990s, when metallic glasses in a bulk form were successfully produced in multicomponent La-, Mg-, Zr-, Pd-, Fe-, Cu-, and Ti-based alloy systems at relatively low cooling rates of 1–100 K·s^{-1} (Inoue et al. 1989, 1990, 1991; Peker and Johnson 1993; Telford 2004). After these discoveries, metallic glasses have been extensively studied and persistently pursued for applications due to their unparalleled functional and structural properties originating from their amorphous microstructure.

While a limited number of metallic glass systems were commercialized by virtue of their exceptional functional properties (e.g., Fe-based metallic glasses for low-loss power distribution transformers due to their low magnetization loss), large space in structural applications is still expected and actively explored. This high expectation originates from two facets of the merits found in metallic glasses. The first merit is their excellent room-temperature mechanical properties, including the superior yield strength ($\approx$ 2.0 GPa), high elastic limit ($\approx$ 2%), and potentially high fracture toughness (1–100 MPa·m$^{1/2}$) (Anand and Su 2005; Schuh et al. 2007), which are obvious when comparing with many other materials in Fig. 1. The second merit is their thermoplastic forming ability in the supercooled liquid region ($T_g < T < T_x$, where T_g and T_x are the glass transition and crystallization temperature, respectively), which enables metallic glasses to be easily processed into complicated shapes as plastics and oxide glasses. All these advantageous properties, however, are compromised by their lack of macroscopic ductility and brittle fracture at temperatures well below T_g, because of shear localization. For instance, under unconstrained uniaxial tension, a metallic glass often undergoes catastrophic failure along a predominant shear band, generating near-zero plastic strain. This drawback, again, is attributed to the amorphous structure of metallic glasses in which any strain-hardening and crack-blunting mechanisms are missing.
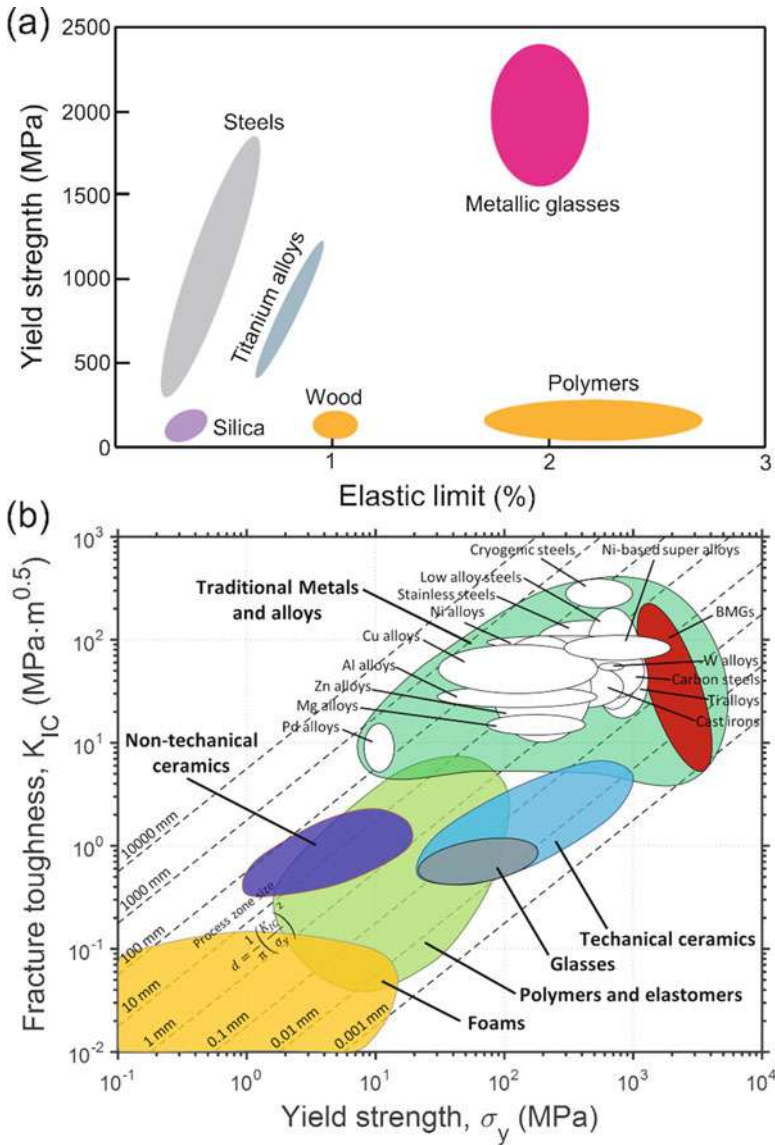
Fig. 1 (**a**) The comparison of the yield strength and elastic limit of metallic glasses with other materials including crystalline metals (Telford 2004). (**b**) Ashby plot of the fracture toughness versus yield strength of metallic glasses along with traditional metals/alloys and other material classes, with the dashed lines roughly representing the process zone size, $\frac{1}{\pi}\left(\frac{K_{IC}}{\sigma_f}\right)^2$, at the crack tip (Ashby 2005; Li et al. 2016c)

Given that brittleness is the biggest barrier preventing metallic glasses from extensively commercialized applications, designing improved ductility in these materials has been a constant pursuit of numerous studies in the past few decades. Along this line of endeavor, some extrinsic means have been proposed and proven to effectively enhance the ductility of metallic glasses through inducing abundant shear bands. Examples of such extrinsic methods include coating metallic glasses with ductile films (Jia et al. 2014), fabricating metallic glass composites (Jia et al. 2015), creating built-in compressive residual stress on surface layers (Wang et al. 2011), and imposing geometric constraints (An et al. 2012; Liu et al. 2012). Despite many different attempts, these extrinsic approaches do not solve the intrinsic brittleness problem in metallic glasses. Therefore, a multitude of efforts are focused on fundamentally understanding the micro-mechanisms of the inelastic deformation in metallic glasses (Argon 1979; Langer 2006; Li et al. 2013, 2015; Spaepen 1977). Along this line of work, the structural heterogeneity has been widely recognized as an intrinsic microstructural element that governs the ductility or brittleness of metallic glasses. The structural heterogeneity originates from the spatial fluctuation in the atomic packing density of metallic glasses and exhibits varied mechanical responses at different locations, usually at the micro- or sub-microscale (Li et al. 2013, 2015). A high density of structural heterogeneities helps retain good ductility of metallic glasses; on the contrary, reducing the amount of structural heterogeneities (e.g., by annealing) leads to significant embrittlement (Li et al. 2013, 2015).

2 Constitutive Models for Metallic Glasses

Constitutive models developed at the continuum level are useful in predicting the macroscopic mechanical behaviors of metallic glasses under various conditions and revealing the deformation phenomena that are difficult to observe in experiments. The constitutive models developed for metallic glasses include those motivated by the underlying deformation micro-mechanisms and those purely based on phenomenological considerations. Below we briefly summarize these two types of constitutive models for metallic glasses.

2.1 Physics-Based Free-Volume Model

Atoms in an amorphous structure are arranged in a disordered and non-closely packed manner. This naturally leaves some extra space for interstitial atoms. The excess space averaged on each atom is termed as the free volume per atom, which can be quantified as the difference between the average atomic volume in a real amorphous structure and that in the ideally ordered structure (Gao 2006). In an amorphous solid, the free volume is statistically distributed over all atoms. The free volume is not just a fictitious concept but actually a physical observable which has been explicitly measured by different experiments (Flores et al. 2011; Slipenyuk and

Eckert 2004; Yavari et al. 2005). The concepts and facts outlined above underlie the free-volume constitutive model. In the free-volume model, it is assumed that the macroscopic plastic flow in a metallic glass is induced by a number of individual atomic jumps. In the absence of external forces, atomic jumps are caused by thermal fluctuations, which obey an Arrhenius dependence on the activation energy, and the number of atoms jumping forward is balanced by that of those jumping backward. As a result, no net macroscopic flow is produced in the metallic glass. On the other hand, when the energy landscape is biased by applying an external stress, e.g., shear stress, there will be more jumps in one direction than the other. Consequently, there will be a net flux of atoms in one direction, the metallic glass exhibiting irreversible deformation. Under a simple one-dimensional shear problem, the plastic shear strain rate is thought of as the fraction of potential jump sites times the net number of atomic jumps on each of such sites per second (Spaepen 1977). With the free-volume theory (Cohen and Turnbull 1959; Spaepen 1977), the fraction of potential jump sites under homogeneous deformation is calculated as $exp\left(-\frac{\alpha v^*}{v_f}\right)$. Under the applied shear stress, τ, the net number of atomic jumps on each potential site per second is $f\left[exp\left(-\frac{\Delta G^m - \tau\Omega/2}{k_B T}\right) - exp\left(-\frac{\Delta G^m + \tau\Omega/2}{k_B T}\right)\right]$, where the first and second terms represent the forward and backward fluxes of atomic jumps, respectively (Spaepen 1977). As a result, the plastic flow equation is given as (Spaepen 1977)

$$\frac{\partial \gamma_p}{\partial t} = 2f \, exp\left(-\frac{\alpha v^*}{v_f}\right) exp\left(-\frac{\Delta G^m}{k_B T}\right) sinh\left(\frac{\tau\Omega}{2k_B T}\right), \quad (1)$$

where γ_p is the plastic shear strain, t is the time, f is the frequency of the atomic vibration, α is a geometric factor on the order of 1, v^* is the hard-sphere volume of an atom, v_f is the average free volume per atom, ΔG^m is the activation energy, Ω is the atomic volume, k_B is the Boltzmann constant, and T is the absolute temperature.

Equation (1) is a standard viscoplasticity model that alone cannot account for the strain-softening effect found in metallic glasses. To account for strain softening, the free volume, v_f, is treated as an order parameter whose creation and annihilation govern the structural disordering and reordering processes in metallic glasses. Specifically, the free volume can be created by a sufficiently high stress squeezing an atom into a neighboring vacancy of smaller volume; it can be annihilated by diffusional atomic jumps. Considering the competition of these two processes, the net rate of free volume increase is written as

$$\frac{\partial v_f}{\partial t} = v^* f \, exp\left(-\frac{\alpha v^*}{v_f}\right) exp\left(-\frac{\Delta G^m}{k_B T}\right)\left\{\frac{2\alpha k_B T}{v_f C_{eff}}\left[cosh\left(\frac{\tau\Omega}{2k_B T}\right) - 1\right] - \frac{1}{n_D}\right\}, \quad (2)$$

where n_D is the number of atomic jumps needed to annihilate a free volume equal to v^* and is usually taken to be 3 (Gao 2006; Huang et al. 2002; Steif et al. 1982), and

the effective modulus is $C_{eff} = E/3(1 - v)$ with E and v being the Young's modulus and Poisson's ratio, respectively.

Equations (1) and (2) constitute the inelastic portion of the continuum mechanics description of the deformation mechanism in metallic glasses. At constant T, a complete one-dimensional isothermal constitutive law for metallic glasses can be formed by combining Eqs. (1) and (2) with Hooke's law for elasticity

$$\frac{\partial \gamma_e}{\partial t} = \left(\frac{\dot{\tau}}{G} \right)_T,$$ (3)

where γ_e is the elastic shear strain and G is the shear modulus.

The isothermal model can be further extended to be a thermomechanically coupled constitutive relation by incorporating the following thermal transport equation encompassing the heat conduction (first term) and plastic energy dissipation (second term) (Bergman et al. 2011; Gao 2006)

$$\dot{T} = \frac{1}{\rho C_P} \left(k \frac{\partial^2 T}{\partial x^2} + \alpha_{TQ} \tau \dot{\gamma}_p \right),$$ (4)

where T is the temperature, k is the thermal conductivity, α_{TQ} is the Taylor-Quinney coefficient representing the fraction of the plastic work converted into heat, ρ is the material density, C_P is the specific heat, and ρC_P denotes the volumetric heat capacity of a material (Bergman et al. 2011).

The one-dimensional partial differential equations, i.e., Eqs. (1) - (3) that form the isothermal model as well as Eqs. (1) - (4) that form the thermomechanically coupled model, can be solved numerically. Solving the isothermal model gives the evolution of the stress, τ, and free volume, v_f, with the shear strain, γ (usually from 0 to 1), or time, t (in seconds), in metallic glasses undergoing homogeneous deformation. Solving the thermomechanically coupled model can, in addition, give the evolution of the temperature, T. The one-dimensional isothermal and thermomechanical models are only applicable to the case of simple shear. To extend their applicability to arbitrary multiaxial stress states, models need to be generalized with appropriate yield criteria. Simply put, the shear stress, τ, in the one-dimensional constitutive equations needs to be replaced with an equivalent stress under a given yield criterion in the process of generalization, to form the constitutive equations for arbitrary stress states. One of such generalizations is to use the J_2 plasticity on the basis of von Mises yield criterion, which assumes that the hydrostatic stress does not cause plastic flow and changes the internal state variable – free volume (Gao 2006). If considering the contribution from the hydrostatic stress, a generalization based on the Drucker-Prager criterion was made by Zhao and Li (2011a). Finally, a generalized constitutive model, based on any yield criterion, can be implemented into a finite element program (e.g., ABAQUS) through developing a user-material subroutine to maximize its usage (Gao 2006; Zhao and Li 2011a).

2.2   Phenomenological Elastic-Viscoplastic Models

In addition to the physics-based free-volume model, many researchers have attempted to develop constitutive models that can phenomenologically capture the macroscopic mechanical responses of metallic glasses. Anand and Su (2005) developed a finite-deformation, Mohr-Coulomb-type constitutive theory that captured some of the essential inelastic deformation features in metallic glasses, which include the hydrostatic stress dependence of yielding, plastic dilatation, and plastic strain localization. This model was initially developed for room-temperature applications but later extended to high homologous temperatures in the range of $0.7T_g \lesssim T \lesssim 0.9T_g$ for simulating the homogeneous plastic flow (Anand and Su 2007) and $0.9T_g \lesssim T \lesssim T_x$ (Henann and Anand 2008) for the transition from a viscoelastic-plastic solid-like response to a Newtonian fluid-like response. Nonetheless, the applications of these models were limited to the isothermal condition. The finite-deformation constitutive model developed by Yang et al. (2006) considers the temperature rise in metallic glasses resulting from the conversion from plastic work to heat but simply treats the thermal problem as adiabatic. Later, Thamburaja and Ekambaram (2007) applied fundamental thermodynamic laws and the principle of micro-force balance to develop a thermomechanically coupled finite-deformation model, targeting for predictions in the supercooled liquid region. This model had also been augmented to include a damage criterion to simulate damage processes (Ekambaram et al. 2011).

All these constitutive models have been implemented in the commercial finite element program, ABAQUS, by writing a user-material subroutine either for the explicit procedure (VUMAT) or implicit procedure (UMAT) (Manual 2013). The former is easy for implementation, but the computation is conditionally stable, whereas the latter is unconditionally stable but more difficult to implement. Given that the model by Anand and Su at room temperature (Anand and Su 2005) has been used rather extensively, below we briefly outline the key governing equations for this model. For the details of the other models, one is referred to the relevant references (Anand and Su 2007; Ekambaram et al. 2011; Henann and Anand 2008; Thamburaja and Ekambaram 2007; Yang et al. 2006).

In the finite-deformation framework, a multiplicative decomposition of the deformation gradient tensor into elastic and plastic parts is used to map the initial, stress-free reference configuration to the current, deformed configuration (Bower 2009):

$$\mathbf{F} = \mathbf{F}^e\mathbf{F}^p, \tag{5}$$

where $\mathbf{F}^p$ is the plastic deformation and rotation mapping the initial reference configuration to an intermediate, unstressed configuration, whereas $\mathbf{F}^e$ is the elastic deformation and rotation mapping the intermediate configuration to the current configuration. The free volume is made to be associated with the plastic volumetric strain $\eta = \ln J^p$, where $J^p = \det \mathbf{F}^p$. The plastic flow in metallic glasses is assumed

to result from shearing along with dilatation relative to potential slip systems. As no preferred directions other than the principal direction of stress exist in amorphous isotropic materials, six potential slip systems are defined relative to the principal directions of Kirchhoff stress. Each slip system is specified by a slip direction, $\mathbf{s}^{(a)}$, and a slip plane normal, $\mathbf{m}^{(a)}$, where $a \in \{1, 2, \ldots, 6\}$ represents the ath slip system. Based on the Mohr-Coulomb criterion, the six potential slip systems are defined as

$$
\begin{aligned}
\mathbf{s}^{(1)} &= \cos\theta\widehat{\mathbf{e}}_1 + \sin\theta\widehat{\mathbf{e}}_3, & \mathbf{m}^{(1)} &= \sin\theta\widehat{\mathbf{e}}_1 - \cos\theta\widehat{\mathbf{e}}_3, \\
\mathbf{s}^{(2)} &= \cos\theta\widehat{\mathbf{e}}_1 - \sin\theta\widehat{\mathbf{e}}_3, & \mathbf{m}^{(2)} &= \sin\theta\widehat{\mathbf{e}}_1 + \cos\theta\widehat{\mathbf{e}}_3, \\
\mathbf{s}^{(3)} &= \cos\theta\widehat{\mathbf{e}}_1 + \sin\theta\widehat{\mathbf{e}}_2, & \mathbf{m}^{(3)} &= \sin\theta\widehat{\mathbf{e}}_1 - \cos\theta\widehat{\mathbf{e}}_2, \\
\mathbf{s}^{(4)} &= \cos\theta\widehat{\mathbf{e}}_1 - \sin\theta\widehat{\mathbf{e}}_2, & \mathbf{m}^{(4)} &= \sin\theta\widehat{\mathbf{e}}_1 + \cos\theta\widehat{\mathbf{e}}_2, \\
\mathbf{s}^{(5)} &= \cos\theta\widehat{\mathbf{e}}_2 + \sin\theta\widehat{\mathbf{e}}_3, & \mathbf{m}^{(5)} &= \sin\theta\widehat{\mathbf{e}}_2 - \cos\theta\widehat{\mathbf{e}}_3, \\
\mathbf{s}^{(6)} &= \cos\theta\widehat{\mathbf{e}}_2 - \sin\theta\widehat{\mathbf{e}}_3, & \mathbf{m}^{(6)} &= \sin\theta\widehat{\mathbf{e}}_2 + \cos\theta\widehat{\mathbf{e}}_3,
\end{aligned}
\tag{6}
$$

where $\widehat{\mathbf{e}}_1$–$\widehat{\mathbf{e}}_3$ represent three orthogonal unit vector, $\theta = \frac{\pi}{4} + \frac{\varphi}{2}$, $\varphi = \tan^{-1}\mu$ is the angle of internal friction, and μ is the friction coefficient.

The plastic volumetric strain (or free volume) evolves according to

$$
\dot{\eta} = \beta \sum_{a=1}^{6} \dot{\gamma}^{(a)},
\tag{7}
$$

where β is the dilatancy function and $\dot{\gamma}^{(a)}$ is the plastic shearing rate on the ath slip system, which are defined, respectively, as

$$
\dot{\gamma}^{(a)} = \dot{\gamma}_o \left\{ \frac{\tau^{(a)}}{c + \mu\sigma^{(a)}} \right\}^{1/m},
\tag{8}
$$

$$
\beta = g_o \left(1 - \frac{\eta}{\eta_{cv}} \right)^p,
\tag{9}
$$

where c is the internal variable named cohesion and its initial value, c_o, is related to the uniaxial tensile yield stress through $\sigma_o = 2c_o/(\cos\varphi + \mu(1 + \sin\varphi))$. $\tau^{(a)}$ and $\sigma^{(a)}$ are the resolved shear stress and compressive normal stress for the ath slip system. $\dot{\gamma}_o$ is the reference shearing rate, and m is the rate sensitivity constant. η_{cv} is the saturation value of the plastic volumetric strain. g_o and p are the positive-valued material parameters.

Equations (7) - (9) represent the evolution of the free volume toward an equilibrium value, η_{cv}, after metallic glasses yield. In addition, strain softening induced by the increase in the free volume is phenomenologically modeled by a decaying cohesion, c. Specifically, c has an initial value $c_0 = c_{cv} + b$ and decays according to

$$c = c_{cv} + b\left(1 - \frac{\eta}{\eta_{cv}}\right)^q, \tag{10}$$

where c_{cv}, b, and q are the positive-valued material constants. It is seen from Eq. (10) that after yielding the cohesion c will drop from c_0 gradually to c_{cv}, reflecting the strain softening in metallic glasses. Equations (7) - (10) essentially constitute the core of the Anand-Su model (Anand and Su 2005). During the practical finite-element implementations, the initial cohesion, c_0, is usually perturbed by a statistical distribution over the domain of interest to trigger random and simultaneous nucleation of shear localization at multiple sites (Anand and Su 2005).

In the model, physically observable material parameters (e.g., μ and c_0) can be evaluated by conducting experiments, e.g., compression and instrumented indentation tests; some phenomenological material parameters (e.g., c_{cv}, b, and q) need to be estimated by curve-fitting procedures, i.e., carrying out numerical simulations and physical measurements under identical conditions and then adjusting the values of the material parameters in the simulations to match certain curves from experiments; finally, for other material parameters that are difficult to be estimated from either way aforementioned (e.g., η_{cv}, g_o, and p), reasonable assumptions have to be made for their values. The example procedures for obtaining the values of all material parameters appearing in the model can be found in Anand and Su (2005) and Su and Anand (2006).

3   Prediction of the Homogeneous Plastic Flow

The homogeneous deformation in metallic glasses occurs at low-stress levels and high temperatures ($> 0.7\, T_g$) and is close to the Newtonian viscous state. In uniaxial tension, a specimen thins down uniformly accompanied with extensive and stable plastic deformation until failure takes place by necking. Such a deformation model is very similar to many superplastic materials. Deformation in this regime usually exhibits notable strain-rate sensitivity (Spaepen 1977), in sharp contrast to the strain-rate insensitivity in the inhomogeneous flow which will be discussed shortly. A set of representative uniaxial true stress-strain curves at different strain rates, measured for homogeneously deformed $Pd_{40}Ni_{40}P_{20}$ metallic glass ($T_g = 577$ K) at $T = 564$ K, are given as solid curves in Fig. 2 (Anand and Su 2007; de Hey et al. 1998; Ichitsubo et al. 2004). Different than the strain hardening in crystalline metals and alloys, the absence of dislocation-mediated plasticity in metallic glasses leads to strain softening in the plastic flow. As shown in Fig. 2, following the occurrence of stress overshoot right after yielding, the flow stress drops until a steady state is reached at a strain of ~ 0.1. As the strain rate increases, the entire stress-strain curve is elevated, the stress drop at the overshoot becomes more precipitous, and the steady state is reached more slowly. The prediction of stress-strain responses at a constant T with a constitutive model involves numerically solving the coupled plastic flow and state-variable evolution equations, i.e., Eqs. (1) - (3) for the free-volume model and

Fig. 2 True stress-strain curves of the pre-annealed $Pd_{40}Ni_{40}P_{20}$ metallic glasses measured experimentally at 564 K and four different strain rates (solid lines) (de Hey et al. 1998) along with the simulated data from the isothermal Anand-Su model (dashed lines) (Anand and Su 2007)
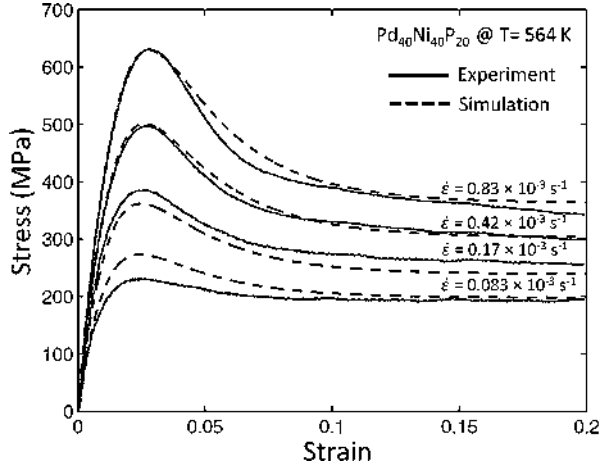
Eqs. (5) - (10) for the Anand-Su model. After estimating the values of the material parameters by curve-fitting to experimental data the isothermal Anand-Su model (Anand and Su 2007) predicts the stress-strain curves as dashed lines in Fig. 2. All characteristics manifested on the experimental stress-strain responses are essentially captured by the model.

The isothermal free-volume model (Eqs. 1) - (3) can also capture the strain-softening characteristic in the uniaxial stress-strain curves of metallic glasses (Gao 2006; Steif et al. 1982). The strain softening in the free-volume model is regulated by the evolution of the order parameter – free volume. According to Eq. (2), the free-volume evolution is a stress-driven process. The increase in the applied stress gives rise to a net accumulation of free volume, as shown by the initial sharp rise in Fig. 3a. Using the viscosity definition $\eta = {}^{\tau}\!/\!_{\dot{\gamma}_p}$ and Eq. (1), the viscosity of a metallic glass is calculated as (Gao et al. 2007)

$$\eta = \frac{\tau}{2 \sinh \left(\tau \Omega / 2 k_B T\right)} f^{-1} \exp \left(\frac{\alpha v^*}{v_f}\right) \exp \left(\frac{\Delta G^m}{k_B T}\right). \tag{11}$$

From Eq. (11), it is seen that the increase in the free volume, v_f, causes a reduction in the viscosity and thus weakens the material. The material weakening caused by the structural disordering is observed as a stress drop on a stress-strain curve (e.g., Fig. 2). The decrease in the flow stress, in turn, slows down the increase of the free volume, as seen from Eq. (2). At some point, the rate of free-volume annihilation by atomic diffusion (second term in Eq. (2)) catches up that of stress-driven free-volume creation (first term in Eq. (2)). An equilibrium state of free volume is therefore established, and the flow stress becomes steady, as denoted by the dashed line in Figs. 3a and 2, respectively.
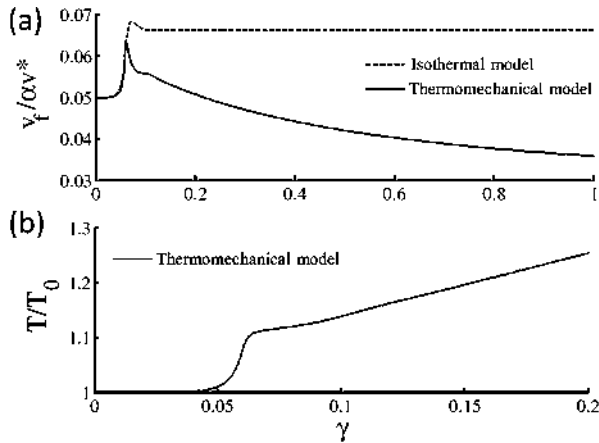
Fig. 3 The normalized (**a**) free volume, $v_f/\alpha v^*$, and (**b**) temperature, T/T_0, as a function of the applied shear strain, γ, pertaining to the homogeneous deformation in a metallic glass. The dashed line is calculated from the one-dimensional isothermal free-volume model, while the solid lines are from the thermomechanically coupled free-volume model. The reference temperature $T_0 = 300$ K and the initial free volume $v_f/\alpha v^* = 0.05$ (Gao et al. 2007)

The homogeneous deformation predicted from a thermomechanically coupled free-volume model Eqs. (1) - (4) shares some similarities with that from the isothermal free-volume model. But, there are also major differences in the predictions of these two models. First, since the thermomechanical model considers the heat generation induced by the plastic work dissipation, the temperature does not remain constant as in the isothermal model but increases gradually as the plastic work accumulates. The temperature rise is slow at the beginning but starts becoming rapid once passing over the stress overshoot, as seen from Fig. 3b in which $\gamma = 0.05$ is approximately where the peak stress happens. According to the free-volume evolution equation (i.e., Eq. (2)), the increase in temperature facilitates the atomic diffusion and therefore the annihilation of free volume. This prediction is shown as the solid curve in Fig. 3a. Unlike the isothermal case, the free volume keeps decreasing after the peak value as shear strain γ increases. No equilibrium is established for the free volume under the thermomechanical scenario.

4   Prediction of the Inhomogeneous Plastic Flow

4.1   Shear Banding

The inhomogeneous deformation in metallic glasses takes place at relatively high stresses (roughly $-2 < \log \frac{\tau}{G} < 0$, where τ and G are the shear stress and shear modulus, respectively) and low temperatures (<0.7 T_g) (Spaepen 1977; Suryanarayana and Inoue 2017). One of the most striking features is that in this

deformation mode, the plastic strain is highly localized into thin bands termed as shear bands. Under unconfined loading conditions such as uniaxial tension, an initiated shear band can propagate promptly throughout the entire specimen, leading to brittle failure. For loading conditions that impose geometric constraints on metallic glass samples (e.g., uniaxial compression (Liu et al. 2012) or indentation (An et al. 2012; Jia et al. 2014), the propagation of shear bands tends to be blocked or redirected by the geometric constraints, consequently resulting in the formation a high density of shear bands. With more shear bands available for bearing the plastic strain, a significant amount of plasticity is possible to achieve (An et al. 2012; Jia et al. 2014; Liu et al. 2012). Therefore, understanding the shear-banding behavior is essential to reveal the deformation mechanism of metallic glasses under various conditions. Diversified experimental techniques have been utilized for fundamentally understanding the nature of shear bands, and a comprehensive summary can be found in ref. (Greer et al. 2013). In complement to experiments, modeling and simulation at the continuum level, particularly with the use of metallic-glass constitutive models, are useful in predicting the shear band-related phenomena.

The shear band inclination angle under uniaxial tension and compression has been an active research topic for years. In crystalline metallic materials, as the yielding is controlled merely by the shear stresses, failure usually occurs along the plane at which the resolved shear stress is maximum, i.e., 45° from the loading axis for both tension and compression (Dieter and Bacon 1986). Nonetheless, the shear band inclination angle and the failure plane in metallic glasses are experimentally found to deviate from 45° (Gao et al. 2011). The angle is greater than 45° in tension and smaller than 45° in compression. Besides, the deviation from 45° for tension and compression is not necessarily symmetric for the same material. To predict the shear band inclination angle in the inhomogeneous deformation mode, the one-dimensional constitutive models of metallic glasses have been generalized for multiaxial stress states based on different failure criteria, including the von Mises (Gao 2006), Mohr-Coulomb (Anand and Su 2005), and Drucker-Prager (Zhao and Li 2011a). The von Mises generalization assumes zero contribution from the hydrostatic stress to yielding, and thus predicts a shear band angle of 45° for both tension and compression, as given in Fig. 4a and d. This assumption is oversimplified and fails to predict the deviation of the shear band angle. The amendment was made by using either the Mohr-Coulomb or Drucker-Prager criterion, both of which are primarily used in soil and rock mechanics considering the contribution of the hydrostatic stress to failure. The Drucker-Prager criterion is simply an extension of the von Mises criterion by adding a term of hydrostatic stress (i.e., Eq. (12) without the I_1 term is the von Mises criterion),

$$Y = aI_1 + \sqrt{J_2} - k = 0, \tag{12}$$

where Y is the yield function; I_1 and J_2 are the first invariant of the stress tensor and the second deviatoric stress invariant, respectively; and a and k are the constants.
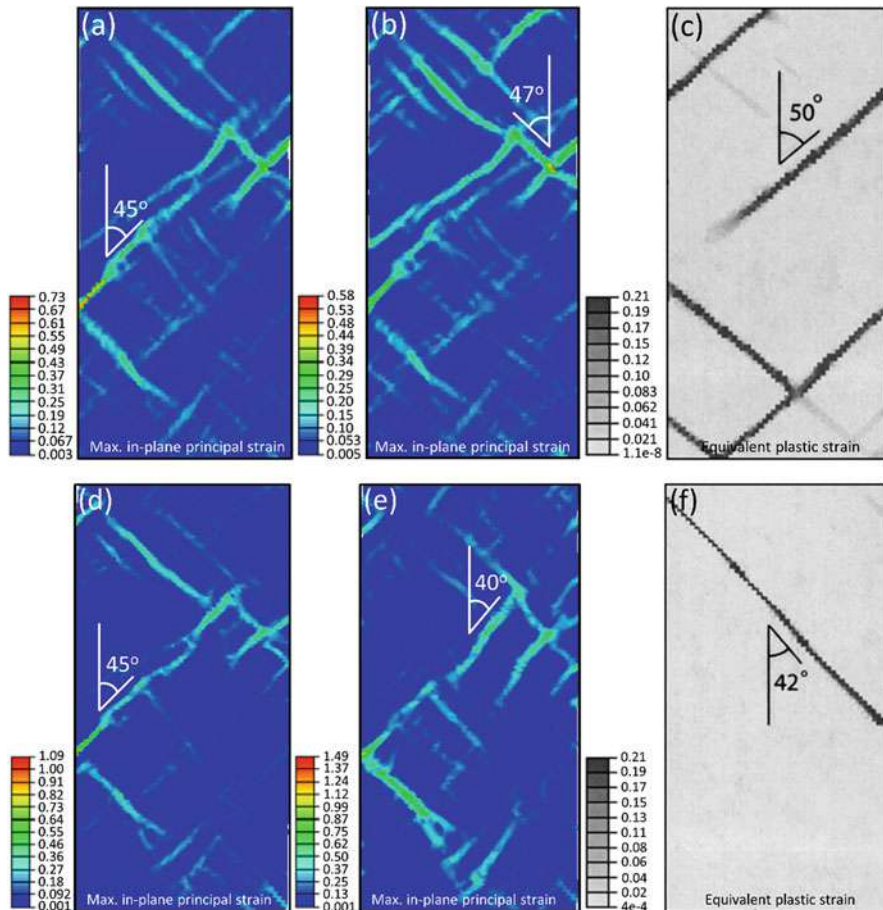
Fig. 4 Shear bands formed under (**a–c**) uniaxial plane-strain tension and (**d–f**) compression. The predictions are, respectively, from the (**a, d**) free-volume model based on the von Mises failure criterion (Zhao and Li 2012), (**b, e**) free-volume model based on the Drucker-Prager criterion (Zhao and Li 2012), and (**c, f**) Anand-Su model based on the Mohr-Coulomb criterion (Anand and Su 2005)

By considering the contribution from the hydrostatic stress, i.e., the first term in Eq. (12), the deviated shear band angles can be accurately predicted. Figures 4b and e give the predicted angles being 47° under tension and 40° under compression. The asymmetry in the angle deviation for tension and compression is captured as well. Additionally, the predicted angles are likely to be further fine-tuned to align more closely with the experimentally observed ones for a given metallic glass by changing the value of a in Eq. (12). Changing a effectively regulates the weight of the hydrostatic stress contribution. The Mohr-Coulomb criterion, whose yield surface is a non-smooth version of that in the Drucker-Prager criterion, can also account for the

dependence of yielding on the hydrostatic stress. But differently, the Mohr-Coulomb model needs a predetermination of shear planes along which inhomogeneous plastic flow could occur, i.e., "slip systems." With the slip systems properly defined, the Mohr-Coulomb model is also capable to predict the asymmetrically deviated shear band angles in tension and compression, as given in Fig. 4c and f (Anand and Su 2005).

The constitutive models can also be used for predicting the shear band activities under more complicated stress states. One of the most widely used applications is to predict shear band patterns under indentations. Figure 5a gives the shear band pattern experimentally observed on the cross-section plane of the $Zr_{56.06}Cu_{23.39}Ni_{15.54}Ta_{2.27}Al_{1.62}Ti_{1.11}$ metallic glass using the bonded-interface
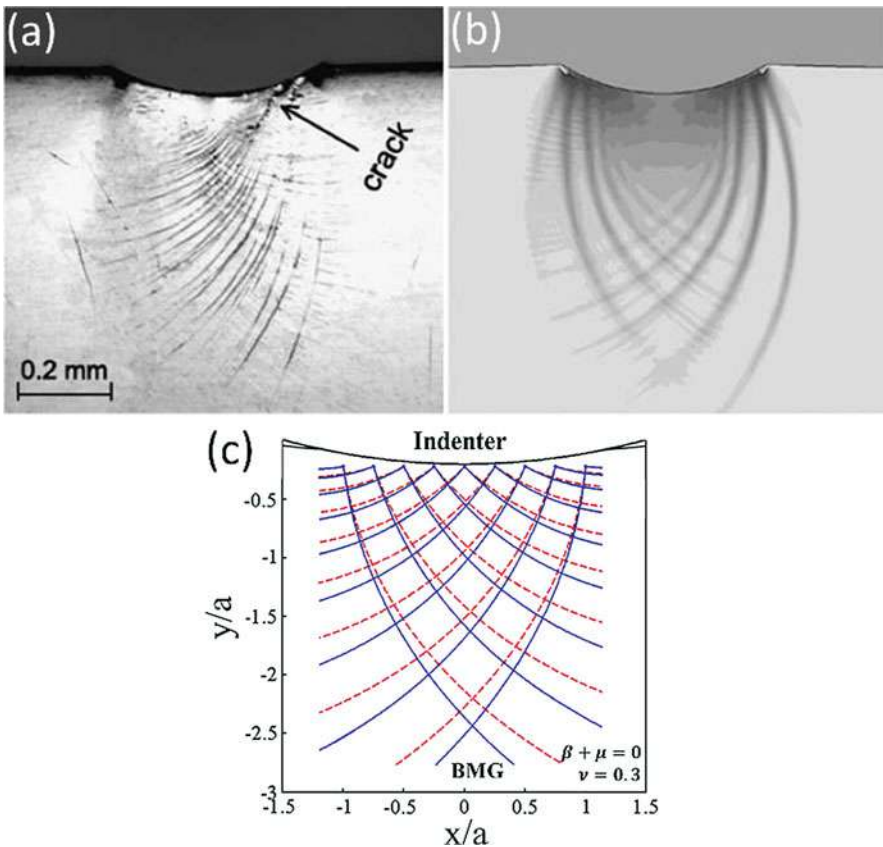
Fig. 5 Shear band patterns from the (**a**) experimental observation on the bonded interface of a Zr-based metallic glass specimen subjected to plane-strain indentation (Su and Anand 2006), (**b**) numerical simulation with the Anand-Su model (Su and Anand 2006), and (**c**) prediction by the Rudnicki-Rice instability theory, with the solid blue curves representing the shear-band directions and the dashed red curves the principal shear-stress directions (Jia et al. 2014)

technique (Su and Anand 2006). The metallic glass was subjected to the plane-strain indentation. As a result, radial shear bands emanating from the contacting interface and intersecting shear bands were observed. The prediction with the Anand-Su model (Su and Anand 2006) in Fig. 5b replicates the shear band pattern closely. Moreover, the initial shear band pattern is also likely to be predicted by a bifurcation theory, considering the onset of strain localization is a result of material instability from the continuum mechanics point of view (Rudnicki and Rice 1975). The exact formulation of the critical condition for strain localization was derived by Rudnicki and Rice (1975). Simply put, the shear band is parallel to the second principal stress component in the principal stress space and is calculated as

$$\theta_0 = \pm \tan^{-1} \sqrt{\frac{\xi - N_{min}}{N_{max} - \xi}}, \tag{13}$$

where $\xi = \frac{1}{3}(1 + \nu)(\mu + \beta) - N(1 - \nu)$, ν is Poisson's ratio, μ is the coefficient of the internal friction (for generally pressure-sensitive materials like metallic glasses), β is the dilatancy factor (describes the flow normality, associated flow if $\mu = \beta$),

$$N_{max} = \sqrt{3}\frac{\sigma'_{\mathrm{I}}}{\sigma_{mises}}, \quad N_{mid} = \sqrt{3}\frac{\sigma'_{\mathrm{II}}}{\sigma_{mises}}, \quad N_{min} = \sqrt{3}\frac{\sigma'_{\mathrm{III}}}{\sigma_{mises}}, \tag{14}$$

and σ'_{I}, σ'_{II}, and σ'_{III} are the principal deviatoric stress components. To apply the Rudnicki-Rice instability theory, the elastic field resulting from a loading condition needs firstly to be obtained either from analytical approaches or finite element simulations. When showing all the shear-band directions on a regular grid, one can construct "streamlines" that are tangent to the shear-band directions at any material point. The constructed streamlines depict the shear band pattern under the given loading condition. The shear band pattern predicted with this approach for a Rockwell indentation is shown in Fig. 5c, which resembles those in Fig. 5a and b (Jia et al. 2014). This procedure has found numerous successful applications in reasonably predicting the shear band patterns under a number of different indentation problems, such as Berkovich indentation and film/substrate systems (An et al. 2012; Gao et al. 2011; Jia et al. 2014).

In addition to predicting the shear band pattern, the constitutive modeling is also useful in revealing other shear band-related quantities that are difficult to be precisely measured by experiments. For instance, the shear band velocity can be traced from the initiation to the final stage of propagation. With the free-volume model, it is found that the propagation of shear bands actually manifests a nonuniform speed, that is, a rather slow stage from the initiation followed by a rapid propagation when they reach the sample surfaces (Zhao et al. 2011). The average speed was estimated to be 100 m/s, with a maximum value of 1400 m/s. The shear band speed assessed with the modeling approach is more thorough than most of the experimental techniques like high-speed camera, SEM, and strain gage (Chen et al. 2009; Song et al. 2010; Yang et al. 2004) because most experiments

are restricted by limited temporal and spatial resolutions and unlikely to detect the initiation and propagation of embryonic shear bands. If a thermomechanically coupled model is applied, local temperature rise inside a propagating shear band can be modeled without significant computational cost as well. Such a modeling method overcomes the measurement uncertainties found in many experiments, as most measurements have to be conducted on the surfaces of samples, e.g., using the infrared imaging and fusible coating methods (Lewandowski and Greer 2005; Yang et al. 2005). The maximum local temperature rise in shear bands measured with the modeling approach was suggested to be a few tens of degrees (Zhao and Li 2011b).

4.2   Flow Serration

Various degrees of plasticity may occur when metallic glasses are loaded in a confined manner, such as uniaxial compression and nanoindentation. In such cases, serrated plastic flows are often observed, manifesting as stress drops on compressive stress-strain curves (Dubach et al. 2009) or displacement bursts ("pop-ins") on nanoindentation load-depth curves (Song and Nieh 2009). The serrated flow in metallic glasses is strain-rate dependent and is believed to be associated with the activation and propagation of individual shear bands (Dubach et al. 2009). The free-volume model, Anand-Su model, and many others render smooth stress-strain or load-depth curves, and only very largely spaced single stress drops or displacement bursts can be captured following the formation of catastrophic shear bands. To capture the flow serration, special treatments are required in the process of developing a constitutive model. One of such models was built up on the rugged free energy landscape, in which the deformation within shear bands is through a continuous transformation from one low energy state to another (Ruan et al. 2011). Since the energy in metallic glasses varies quasiperiodically from peaks to valleys in a rugged free energy landscape, the plastic strain also experiences a quasiperiodic variation as the deformation runs between peaks and valleys. As a result, the serrated flow in metallic glasses is also captured. Conducting uniaxial plane-strain compression tests with this model successfully catches the flow serration on the stress-strain curves, as shown in Fig. 6 (Ruan et al. 2011). The serrated flow appears on all three compressive stress-strain curves conducted at different strain rates even though each exhibits different global plasticity. Alongside shown in Fig. 6 are the tensile stress-strain curves obtained from the identical strain rates, in which brittle failure happens and the serrated flow is absent. These predictions are qualitatively consistent with experimental observations (Dubach et al. 2009).

Another phenomenological model that can also capture the serrated flow in metallic glasses is a rheological model that considers the periodic appearance of flow serration as a result of the interplay of strain softening and strain hardening (Yousfi et al. 2013). As in many other models, strain softening is assumed to originate from the decrease of viscosity following the shear band formation. Strain hardening, on the other hand, is caused by the arrest of shear band propagation by nanocrystalline particles precipitated within the shear zone. Alternating strain

Fig. 6 The stress-strain curves of a metallic glass under plane-strain tension and compression tests at three different strain rates, predicted by the constitutive model based on the rugged free energy landscape (Ruan et al. 2011). The compressive stress-strain curves exhibit different degrees of plasticity, and the serrated flows are reasonably reproduced, whereas brittle failure with zero plasticity features the tensile curves
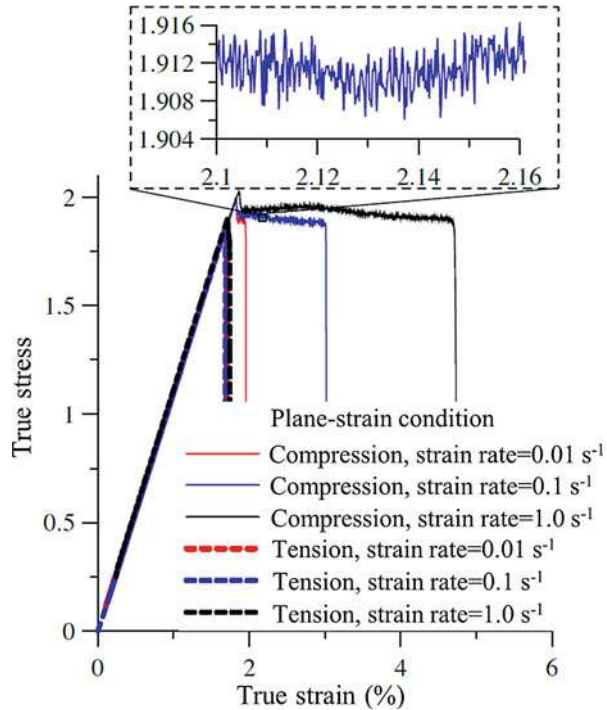
softening and hardening occurs periodically and results in serrated flow. The model, after numerical implementation and determining the material parameters by a data fitting procedure, has already reproduced the serrated flow observed in experiments. Nevertheless, simply assuming that the strain hardening is a result of nanocrystallization lacks generality, as nanocrystallization does not necessarily happen for all metallic glasses exhibiting flow serration. A more generalized serration model should take into account the correlation between the serrated flow and shear-band activation events, and the dependence of the serrated flow on applied strain rates should also be reasonably captured (Schuh and Nieh 2003).

4.3   Cavitation

Some metallic glasses fracture in a sort of ductile manner with profuse shear bands accommodating the large-scale plasticity near the crack or notch tips. On the contrary, many others (e.g., Mg-based metallic glasses) fail in a perfectly brittle manner along a single fast-propagating shear band, just as silicate glasses (Li et al. 2013, 2015; Xi et al. 2006). Observations of the fracture surfaces also reveal salient distinctions: the fracture surfaces of somewhat ductile metallic glasses exhibit vein patterns, whereas the very brittle ones are characteristics of misty-mirror morphologies (Li et al. 2013, 2015). Further inspection of the misty-mirror

morphologies with the high-resolution scanning electron microscopy (SEM) reveals periodic stripes at the nanoscale, which are usually termed as nano-corrugations (Wang et al. 2007, 2008; Xi et al. 2006). It is inferred that these nano-corrugations are formed as a result of the nucleation and coalescence of cavities at the nanoscale, because the atomic force microscopy (AFM) observations found that a peak-to-peak matching relation existed for the nano-corrugations from two complementary fracture surfaces of a Mg-based metallic glass (Wang et al. 2007). The inference was confirmed later by molecular dynamic (MD) simulations of model glasses (Guan et al. 2013; Murali et al. 2011). Additionally, MD simulations suggest that brittle metallic glasses fluctuate spatially in their atomic density and cavities tend to nucleate in the regions of reduced density (Murali et al. 2011). Since the reduced density implies weakened strength, such a finding is linked with another direction of active research, namely, structural heterogeneities in metallic glasses. With the nanoindentation technique, it has been proven that there exists strength fluctuations in metallic glasses. Based on strength fluctuations in metallic glasses, a structure model consisting of a strong matrix and distributed weak zones was suggested (Li et al. 2013, 2015).

Based on these findings, Singh et al. attempted to use finite element simulations to gain more insights on the cavitation in metallic glasses (Singh et al. 2013, 2014, 2016). In their early work, the finite element model of a heterogeneous solid containing a distribution of weak zones was used for representing brittle metallic glasses with structural heterogeneities (Singh et al. 2013, 2014). Both the background material and weak zones were treated as elastic-plastic materials following the J_2 flow theory. The yield strength of the weak zones is reduced in comparison to that of the background material, and after yielding further deformation is bounded by the power law hardening rule. To trigger cavitation, a tiny void was placed at the center of each individual weak zone. By subjecting the heterogeneous solid to equibiaxial stretching under the plane-strain condition, it was demonstrated that the presence of the weak zones permitted cavities to nucleate at substantially lower hydrostatic stresses in comparison with a homogeneous solid. Besides, the critical hydrostatic stress for the onset of cavitation in a heterogeneous solid was dependent on the local properties of the weak zones. These early studies did not explicitly simulate the cavitation process yet gave the cause of nano-cavitation in metallic glasses from the continuum mechanics standpoint, that is, the existing heterogeneities (weak zones) can reduce the cavitation stress and cavitation-induced failure would occur once the properties and distribution of these heterogeneities reach some critical values.

However, a preferable way to simulate the cavitation-induced fracture process in brittle metallic glasses is to assign a more realistic constitutive law to the heterogeneous body. One of such options is to use the Anand-Su model (Anand and Su 2005) based on the Mohr-Coulomb criterion. In the later work by Singh et al. (2016), the heterogeneous metallic glass, by symmetry, was simulated as a classical semicircular domain of radius R_o with three equally spaced weak zones of radius $r_w = 0.13 L_o$ located on the right-half symmetry line (WZ-I, WZ-II, and WZ-III from center to the perimeter). The distance between two consecutive weak zones and from the domain center to WZ-I is L_o. The entire domain was assigned

with the constitutive behavior defined in the Anand-Su model, but a reduced yield strength, σ_{ow}, was used for the weak zones relative to the matrix material strength, σ_o. A small-scale yielding (SSY) fracture problem was simulated by inserting a semicircular notch of length of R_o and diameter of $b_o = 3.3 \times 10^{-3} R_o$ along the left-half symmetry line ahead of the first weak zone and creating a small circular void of radius $r_o = 0.013 L_o$ at the center of each individual weak zone to facilitate cavitation. With such a method, two fracture mechanisms, shear banding-induced fracture and cavitation-induced fracture, are incorporated, and their competition and interplay are therefore likely to be captured. The shear banding process is captured by the constitutive model, and the cavitation is presumed to commence when the void area starts increasing rapidly following the peak of the macroscopic hydrostatic stress, which is defined as

$$\sum_{hw}(t) = \frac{1}{V_w(t)} \int_{V_w(t)} \sigma_h(\mathbf{x}, t) \, dV, \tag{15}$$

where $\sigma_h(\mathbf{x}, t)$ is the current Cauchy (microscopic) hydrostatic stress and $V_w(t)$ is the current volume of a weak zone. The simulated fracture process in such a heterogeneous brittle metallic glass with $\frac{\sigma_{ow}}{\sigma_o} = 0.1$ is depicted in Fig. 7 by the evolution of the principal logarithmic plastic strain $\log \lambda_1^p$ at different values of load parameter $K/\left(\sigma_0 \sqrt{L_0}\right)$. For clarity, only a region containing the notch and weak zones is truncated for display, and the contours are generated to the full domain through symmetry. The cavitation of WZ-I commences first, while the plastic deformation at the notch tip is still negligibly small (Fig. 7a). Progressively, cavitation in WZ-II and WZ-III starts as well, and all three cavities expand rapidly (Fig. 7b and c). In the meantime, yielding occurs at the notch root and shear bands start emanating (Fig. 7b and c). As shear bands continue to evolve and cavities expand, at some point two primary curved shear bands emitting from the notch root exclusively grow, crossing each other and ultimately linking with the growing cavities (Fig. 7d). As loading progresses, these two shear bands keep growing in a sinusoidal manner and connect all three cavities, as shown in Fig. 7e and f. This process is known as void coalescence, and the two wavy shear bands are where failure would happen once the plastic strain inside exceeds a critical value (note that the actual fracture process through crack initiation and propagation was not simulated), as marked by the black dashed lines in Fig. 7d–f. With this simulation, a number of important implications on the brittle fracture of metallic glasses can be observed. First, cavitation prevails in a heterogeneous metallic glass when the strength of weak zones is significantly lower than that of the matrix material, because the critical cavitation stress in such a scenario is substantially below the yield strength of the matrix material (e.g., 0.2–0.25 σ_o for $\frac{\sigma_{ow}}{\sigma_o} = 0.1$). Second, cavities grow unstably due to the absence of strain hardening. Third, the occurrence of void coalescence is by the failure of wavy shear bands that run through neighboring cavities, unlike ductile crystalline metals in which cavities coalesce through necking. This special cavity coalescence behavior explains the formation
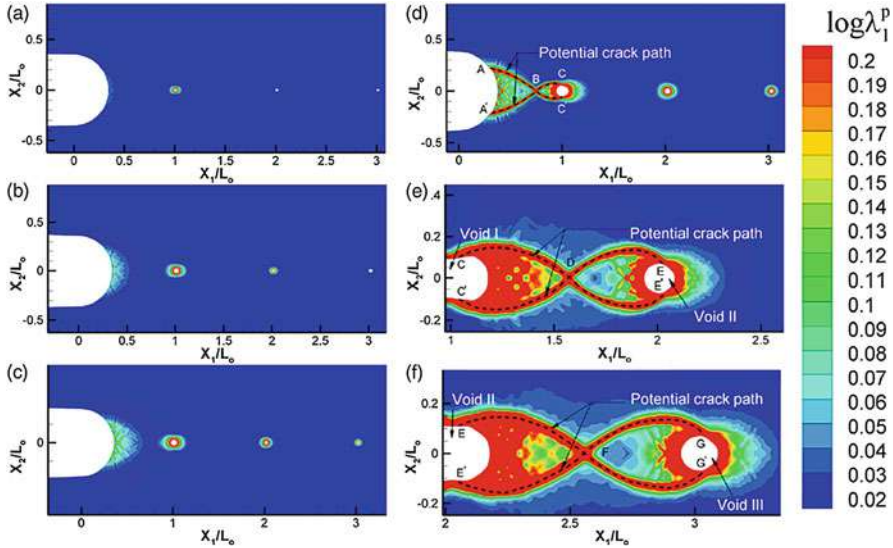
Fig. 7 The simulated course of cavitation ahead of a notch tip in a metallic glass, as conveyed by the evolution of the maximum principal logarithmic plastic strain $\log \lambda_1^p$ at $K / \left(\sigma_0 \sqrt{L_0} \right)$ of (**a**) 0.67, (**b**) 0.96, (**c**) 1.21, (**d**) 1.45, (**e**) 2.09, and (**f**) 2.37 for $\frac{\sigma_{ow}}{\sigma_o} = 0.1$ (Singh et al. 2016). The metallic glass (yield strength, σ_o) is simulated with the Anand-Su constitutive model (Anand and Su 2005), and three circular weak zones with the reduced yield strength σ_{ow} (radii $r_w = 0.13L_o$ with L_o being the distance between two consecutive zones) are placed ahead of the notch to mimic the density fluctuations in brittle metallic glasses. Each weak zone has a small circular void (radii $r_o = 0.013L_o$) in the center for triggering cavitation. The black dashed lines is indicative of potential crack propagation paths

of nano-corrugations on the fracture surfaces of brittle metallic glasses. Fourth, the failure is brittle because of the unstable cavity growth and rapid void coalescence through confined shear bands and also because non-diffuse formation of shear bands involves minimal plastic dissipation.

4.4 Fracture

Many mechanistic aspects of shear localization-induced fracture in "ductile" metallic glasses and cavitation-induced failure in brittle ones have been covered in the previous subsections. A topic that deserves special treatment in this section is the transition in failure mechanism, particularly when edge notches are present in a metallic glass sample. In a brittle solid (e.g., ceramics and oxide glasses), the introduction of notches tends to bring down strength. This phenomenon is called positive notch effect. However, in metallic glasses different types of notch effects are observed in experiments, including the positive notch effect, notch insensitivity, and

inverse notch effect in which the notched sample has higher fracture strength than the one without a notch (Pan et al. 2015). A number of researchers have attempted to use continuum mechanics modeling to understand the origin of these notch effects, particularly the inverse notch effect. A modified Anand-Su model that accounts for the strongly nonlinear and eventually softening, dilatational volumetric elastic response of metallic glasses was used to simulate the notch-tip fracture initiation in a representative metallic glass under small-scale yielding (SSY) and plane-strain mode-I loading (Henann and Anand 2009). The simulations suggest that the fracture toughness of the notched metallic glass sample increases linearly with the square root of the notch-tip radius and decreases with the increasing ratio of the shear modulus to the bulk modulus. Systematic analysis of the shear band patterns near the notches of different sizes with the free-volume model and the Rudnicki-Rice instability theory by Li et al. (2016a) indicates that increasing the notch-tip radius tends to promote the formation of a higher density of shear bands, corroborating the conclusion by Henann and Anand (2009). In addition, when the notches are shallow (e.g., ≤2 mm), only four branches of radial shear bands far apart from each other emit from the notch roots; as the notches gradually become deep (e.g., ≥5 mm), a new category of semicircular shear bands connecting the two notches emerges, and radial shear bands start intersecting with each other (Li et al. 2016a). This implies that the fracture is transiting from a brittle manner to a ductile one as the notches become deep. Since cavitation processes are missing from the free-volume model, a transition in the failure mechanism cannot be captured explicitly. As a result, all considerations are based on the operation of shear bands. Nevertheless, molecular dynamic (MD) simulations did indicate a transition in the failure mechanism from shear banding at the notch tip to cavitation and void coalescence as notches become deep (Pan et al. 2015).

Other factors that the constitutive modeling has found to have crucial effects on the fracture toughness of notched metallic glasses include the internal friction of the material and mode mixity (Li et al. 2016b; Tandaiya et al. 2007, 2009). Higher internal friction tends to suppress the brittle fracture by microcracking and promotes crack-tip plasticity through the formation of multiple shear bands, thereby toughening metallic glasses. Furthermore, it was conjectured that the large plastic strains resulting from high internal friction may give rise to nucleation and coalescence of microvoids (Tandaiya et al. 2007). As for the mode mixity, the increase in mode-II component can significantly enlarge the plastic zone size and elevate the plastic strain near the notches, stimulating ductile fracture (Tandaiya et al. 2009). Finally, it should be noted that even though many insights have been gained on the fracture mechanism of notched metallic glasses by the constitutive modeling, all previous works are limited to analyzing the deformation/stress fields or shear banding activities ahead of notch tips. A comprehensive constitutive model that unifies the cavitation failure mechanism and shear-banding failure mechanism is still lacking for painting a more sensible mechanistic picture for the failure mechanism transition.

5   Design Applications

In addition to the predictive capability, the constitutive models, once implemented in a finite element program as user subroutines, can be used for guiding material, product, or process designs (Dunne and Petrinic 2005). This capability is one of the most remarkable advantages in the continuum mechanics modeling that is unmatched by experiments and atomistic simulations. Finite element modeling has been used for design purposes in multiple areas in metallic glasses. Broadly, these applications can be classified into two categories: (1) designing metallic glasses for high ductility, strength, and toughness at room temperature and (2) optimizing the net-shape forming processes of metallic glasses at high homologous temperatures. In both types of applications, the constitutive laws have been used.

The first category of design applications in metallic glasses driven by finite element modeling is pertinent to material or product designs. Given that the brittleness is the biggest hurdle limiting large-scale practical applications of metallic glasses, finite element modeling under multiaxial stress scenarios is useful for designing routes for improving their ductility, strength, and toughness. Designs can be at the macroscopic scale (e.g., sample size in mm or above) by manipulating the propagation of shear bands in metallic glasses. Existing proven ways that can effectively improve ductility include creating geometric gradients in a metallic glass sample (Liu et al. 2012), coating a bulk metallic glass with a thin film (An et al. 2012; Jia et al. 2014), making metallic glass matrix composites (Jia et al. 2015), and creating porous structures (Sarac et al. 2014). Fracture toughness and strength of metallic glasses can be improved by changing a wide notch into a narrow and deep one in a metallic glass sample to induce many intersecting shear bands (Li et al. 2016a, b). It is also possible to optimize the design for a specific application consideration. For instance, simulations suggest that when the yield strength is a more important consideration than ductility, the periodic alternating pore configuration design in a metallic-glass porous structure should be the designer's first choice over others (Sarac et al. 2014). On the other hand, if the ductility is the primary concern, the stochastic pore distribution comprised of two different pore sizes is more useful (Sarac et al. 2014). Furthermore, innovative design concepts might be explored to perceive novel composite structures. Examples of such explorations include designing metallic-glass composites filled with discrete particles or with network structures (Jiang et al. 2017).

Designs can be also at the microscopic scale (e.g., atomic level) by controlling the initiation site of shear bands and the way they interact with each other. This is achieved by manipulating the distribution of free volume. It has already been proven by experiments that structural heterogeneities exist in metallic glasses at the micro- or sub-microscale and their amount and distribution govern the intrinsic fracture toughness and ductility of metallic glasses (Li et al. 2013, 2015). But the question about at what amount and distribution a metallic glass can gain most in its fracture toughness is still elusive. While this question is not trivial for experimentalists to answer, it is much more tangible by utilizing the low-cost and efficient optimization capability in the finite element modeling. As the free volume is a microscale quantity

Fig. 8 Top row: the shapes of a miniature gear component achieved by the superplastic forming of the $La_{61.4}Al_{15.9}Ni_{11.35}Cu_{11.35}$ metallic glass in experiments. Bottom row: the shapes obtained from the simulation with the use of a metallic-glass constitutive model at high homologous temperatures (Thamburaja and Ekambaram 2007). (**a**) Initial shape, (**b–d**) intermediate shapes, and (**d**) final shape. The forming process is conducted at an ambient temperature of 432 K with the use of a pre-made die (Ekambaram et al. 2010)

implemented in many constitutive models, manipulating the distribution of the free volume is analogous to manipulating that of structural heterogeneities. For this reason, designing the distribution of the free volume with finite element simulations can give promising results. For instance, it was found that certain controlled free-volume gradients could help boost the strength and toughness of metallic glasses (Wang et al. 2017a). Also, a bimodal distribution of the free volume is more effective than a Gaussian-like distribution in toughening and hardening metallic glasses (Wang et al. 2016). These modeling results provide important insights on designing metallic-glass-matrix composites with enhanced ductility and toughness. Specifically, in order to obtain the best performance in toughness, the type and distribution of the two phases in an experimental metallic-glass composite system should be adjusted in such a way that the free-volume distribution matches the optimal scenarios predicted from simulations (Wang et al. 2016, 2017a, b).

The second category of design applications in metallic glasses driven by finite element modeling involves forming process designs. Within the supercooled liquid region, metallic glasses have very low viscosity and excellent formability. As metallic glasses near supercooled region can flow very easily and exhibit superplastic-like deformation characteristics, these materials are much more advantageous than any other metallic materials in the net-shape forming of complex components. Superplastic forming processes like blow molding and injection molding commonly used in the formation of plastic components are all likely to be utilized for metallic glasses in this temperature regime. As in many other metal forming processes,

finite element simulations can find unparalleled values in optimizing the superplastic forming process of metallic glass components. With the metallic-glass constitutive model at high homologous temperatures implemented in finite element software (Thamburaja and Ekambaram 2007), attempts of this kind have been made in the process of designing metallic glass miniature gears (Ekambaram et al. 2010, 2011). The top row in Fig. 8 progressively shows the component shapes formed at different experimental forming stages at an ambient temperature of 432 K (Ekambaram et al. 2010). The bottom row is the prediction at the corresponding stages by the finite element model, whose material parameters have been determined by fitting the constitutive model to a set of stress-strain data obtained from uniaxial compression tests conducted at various strain rates and temperatures. It is seen that the model is capable of precisely replicating the entire forming process achieved in experiments. If a damage criterion is included, the simulation is also able to predict potential failure spots during the forming process (Ekambaram et al. 2011).

References

Abaqus manual (Version 6.13-2) (2013) Dassault Systémes Simulia Corp, Providence

An ZN, Li WD, Liu FX, Liaw PK, Gao YF (2012) Interface constraints on shear band patterns in bonded metallic glass films under microindentation. Metall Mater Trans A 43:2729–2741

Anand L, Su C (2005) A theory for amorphous viscoplastic materials undergoing finite deformations, with application to metallic glasses. J Mech Phys Solids 53:1362–1396

Anand L, Su C (2007) A constitutive theory for metallic glasses at high homologous temperatures. Acta Mater 55:3735–3747

Argon AS (1979) Plastic deformation in metallic glasses. Acta Metall 27:47–58

Ashby MF (2005) Materials selection in mechanical design, 3rd edn. Butterworth-Heinemann, Burlington

Bergman TL, Lavine AS, Incropera FP, DeWitt DP (2011) Fundamentals of heat and mass transfer, 7th edn. Wiley, Hoboken

Bower AF (2009) Applied mechanics of solids. CRC Press, Boca Raton

Chen HM, Huang JC, Song SX, Nieh TG, Jang JSC (2009) Flow serration and shear-band propagation in bulk metallic glasses. Appl Phys Lett 94:141914

Cohen MH, Turnbull D (1959) Molecular transport in liquids and glasses. J Chem Phys 31:1164–1169

De Hey P, Sietsma J, van den Beukel A (1998) Structural disordering in amorphous $Pd_{40}Ni_{40}P_{20}$ induced by high temperature deformation. Acta Mater 46:5873–5882

Dieter GE, Bacon DJ (1986) Mechanical metallurgy. McGraw-Hill, New York

Dubach A, Dalla Torre FH, Löffler JF (2009) Constitutive model for inhomogeneous flow in bulk metallic glasses. Acta Mater 57:881–892

Dunne F, Petrinic N (2005) Introduction to computational plasticity. Oxford University Press, Oxford

Ekambaram R, Thamburaja P, Yang H, Li Y, Nikabdullah N (2010) The multi-axial deformation behavior of bulk metallic glasses at high homologous temperatures. Int J Solids Struct 47:678–690

Ekambaram R, Thamburaja P, Nikabdullah R (2011) Shear localization and damage in metallic glasses at high homologous temperatures. Int J Struct Changes Solids 1:15–29

Flores KM, Suh D, Dauskardt RH, Asoka-Kumar P, Sterne PA, Howell RH (2011) Characterization of free volume in a bulk metallic glass using positron annihilation spectroscopy. J Mater Res 17:1153–1161

Gao YF (2006) An implicit finite element method for simulating inhomogeneous deformation and shear bands of amorphous alloys based on the free-volume model. Model Simul Mater Sci Eng 14:1329

Gao YF, Yang B, Nieh TG (2007) Thermomechanical instability analysis of inhomogeneous deformation in amorphous alloys. Acta Mater 55:2319–2327

Gao YF, Wang L, Bei H, Nieh TG (2011) On the shear-band direction in metallic glasses. Acta Mater 59:4159–4167

Greer AL, Cheng YQ, Ma E (2013) Shear bands in metallic glasses. Mater Sci Eng R 74:71–132

Guan P, Lu S, Spector MJB, Valavala PK, Falk ML (2013) Cavitation in amorphous solids. Phys Re Lett 110:185502

Henann D, Anand L (2008) A constitutive theory for the mechanical response of amorphous metals at high temperatures spanning the glass transition temperature: application to microscale thermoplastic forming. Acta Mater 56:3290–3305

Henann DL, Anand L (2009) Fracture of metallic glasses at notches: effects of notch-root radius and the ratio of the elastic shear modulus to the bulk modulus on toughness. Acta Mater 57:6057–6074

Huang R, Suo Z, Prevost JH, Nix WD (2002) Inhomogeneous deformation in metallic glasses. J Mech Phys Solids 50:1011–1027

Ichitsubo T, Matsubara E, Kai S, Hirao M (2004) Ultrasound-induced crystallization around the glass transition temperature for $Pd_{40}Ni_{40}P_{20}$ metallic glass. Acta Mater 52:423–429

Inoue A, Zhang T, Masumoto T (1989) Al-La-Ni amorphous alloys with a wide supercooled liquid region. Mater Trans, JIM 30:965–972

Inoue A, Zhang T, Masumoto T (1990) Zr-Al-Ni amorphous alloys with high glass transition temperature and significant supercooled liquid region. Mater Trans, JIM 31:177–183

Inoue A, Kato A, Zhang T, Kim SG, Masumoto T (1991) Mg-Cu-Y amorphous alloys with high mechanical strengths produced by a metallic mold casting method. Mater Trans, JIM 32: 609–616

Jia HL, Liu FX, An ZN, Li WD, Wang GY, Chu JP, Jand JSC, Gao YF, Liaw PK (2014) Thin-film metallic glasses for substrate fatigue-property improvements. Thin Solid Films 561:2–27

Jia HL, Zheng LL, Li WD, Li N, Qiao JW, Wang GY, Ren Y, Liaw PK, Gao YF (2015) Insights from the lattice-strain evolution on deformation mechanisms in metallic-glass-matrix composites. Metall Mater Trans A 46:2431–2442

Jiang Y, Sun L, Wu Q, Qiu K (2017) Enhanced tensile ductility of metallic glass matrix composites with novel microstructure. J Non-Cryst Solids 459:26–31

Klement W, Willens RH, Duwez P (1960) Non-crystalline structure in solidified gold-silicon alloys. Nature 187:869–870

Langer JS (2006) Shear-transformation-zone theory of deformation in metallic glasses. Scripta Mater 54:375–379

Lewandowski JJ, Greer AL (2005) Temperature rise at shear bands in metallic glasses. Nat Mater 5:15–18

Li WD, Bei H, Tong Y, Dmowski W, Gao YF (2013) Structural heterogeneity induced plasticity in bulk metallic glasses: from well-relaxed fragile glass to metal-like behavior. Appl Phys Lett 103:171910

Li WD, Gao YF, Bei H (2015) On the correlation between microscopic structural heterogeneity and embrittlement behavior in metallic glasses. Sci Rep 5:14786

Li WD, Bei H, Gao YF (2016a) Effects of geometric factors and shear band patterns on notch sensitivity in bulk metallic glasses. Intermetallics 79:12–19

Li WD, Gao YF, Bei H (2016b) Instability analysis and free volume simulations of shear band directions and arrangements in notched metallic glasses. Sci Rep 6:34878

Li WD, Liaw KP, Gao YF (2016c) Fracture resistance of high entropy alloys: a review. Intermetallics 99:69–83

Liu MC, Huang JC, Chen KW, Lin JF, Li WD, Gao YF, Nieh TG (2012) Is the compression of tapered micro- and nanopillar samples a legitimate technique for the identification of deformation mode change in metallic glasses? Scripta Mater 66:817–820

Murali P, Guo TF, Zhang YW, Narasimhan R, Li Y, Gao HJ (2011) Atomic scale fluctuations govern brittle fracture and cavitation behavior in metallic glasses. Phys Rev Lett 107:215501

Pan J, Zhou HF, Wang ZT, Li Y, Gao HJ (2015) Origin of anomalous inverse notch effect in bulk metallic glasses. J Mech Phys Solids 84:85–94

Peker A, Johnson WL (1993) A highly processable metallic glass: $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10.0}Be_{22.5}$. Appl Phys Lett 63:2342–2344

Ruan HH, Zhang LC, Lu J (2011) A new constitutive model for shear banding instability in metallic glass. Int J Solids Struct 48:3112–3127

Rudnicki JW, Rice JR (1975) Conditions for the localization of deformation in pressure-sensitive dilatant materials. J Mech Phys Solids 23:371–394

Sarac B, Wilmers J, Bargmann S (2014) Property optimization of porous metallic glasses via structural design. Mater Lett 134:306–310

Schuh CA, Nieh TG (2003) A nanoindentation study of serrated flow in bulk metallic glasses. Acta Mater 51:87–99

Schuh CA, Hufnagel TC, Ramamurty U (2007) Mechanical behavior of amorphous alloys. Acta Mater 55:4067–4109

Singh I, Guo TF, Murali P, Narasimhan R, Zhang YW, Gao HJ (2013) Cavitation in materials with distributed weak zones: implications on the origin of brittle fracture in metallic glasses. J Mech Phys Solids 61:1047–1064

Singh I, Guo TF, Narasimhan R, Zhang YW (2014) Cavitation in brittle metallic glasses – effects of stress state and distributed weak zones. Int J Solids Struct 51:4373–4385

Singh I, Narasimhan R, Ramamurty U (2016) Cavitation-induced fracture causes nanocorrugations in brittle metallic glasses. Phys Rev Lett 117:044302

Slipenyuk A, Eckert J (2004) Correlation between enthalpy change and free volume reduction during structural relaxation of $Zr_{55}Cu_{30}Al_{10}Ni_5$ metallic glass. Scripta Mater 50:39–44

Song SX, Nieh TG (2009) Flow serration and shear-band viscosity during inhomogeneous deformation of a Zr-based bulk metallic glass. Intermetallics 17:762–767

Song SX, Wang XL, Nieh TG (2010) Capturing shear band propagation in a Zr-based metallic glass using a high-speed camera. Scripta Mater 62:847–850

Spaepen F (1977) A microscopic mechanism for steady state inhomogeneous flow in metallic glasses. Acta Metall 25:407–415

Steif PS, Spaepen F, Hutchinson JW (1982) Strain localization in amorphous metals. Acta Metall 30:447–455

Su C, Anand L (2006) Plane strain indentation of a Zr-based metallic glass: experiments and numerical simulation. Acta Mater 54:179–189

Suryanarayana C, Inoue A (2017) Bulk metallic glasses. CRC Press, Boca Raton

Tandaiya P, Narasimhan R, Ramamurty U (2007) Mode I crack tip fields in amorphous materials with application to metallic glasses. Acta Mater 55:6541–6552

Tandaiya P, Ramamurty U, Narasimhan R (2009) Mixed mode (I and II) crack tip fields in bulk metallic glasses. J Mech Phys Solids 57:1880–1897

Telford M (2004) The case for bulk metallic glass. Mater Today 7:36–43

Thamburaja P, Ekambaram R (2007) Coupled thermo-mechanical modelling of bulk-metallic glasses: theory, finite-element simulations and experimental verification. J Mech Phys Solids 55:1236–1273

Wang G et al (2007) Nanoscale periodic morphologies on the fracture surface of brittle metallic glasses. Phys Rev Lett 98:235501

Wang G, Chan KC, Xu XH, Wang WH (2008) Instability of crack propagation in brittle bulk metallic glass. Acta Mater 56:5845–5860

Wang L, Bei H, Gao YF, Lu ZP, Nieh TG (2011) Effect of residual stresses on the hardness of bulk metallic glasses. Acta Mater 59:2858–2864

Wang Y, Li M, Xu J (2016) Toughen and harden metallic glass through designing statistical heterogeneity. Scripta Mater 113:10–13

Wang Y, Li M, Xu J (2017a) Free volume gradient effect on mechanical properties of metallic glasses. Scripta Mater 130:12–16

Wang Y, Li M, Xu J (2017b) Mechanical properties of spinodal decomposed metallic glass composites. Scripta Mater 135:41–45

Xi XK, Zhao DQ, Pan MX, Wang WH, Wu Y, Lewandowski JJ (2006) Periodic corrugation on dynamic fracture surface in brittle bulk metallic glass. Appl Phys Lett 89:181911

Yang B, Liaw PK, Wang G, Morrison M, Liu CT, Buchanan RA, Yokoyama Y (2004) In-situ thermographic observation of mechanical damage in bulk-metallic glasses during fatigue and tensile experiments. Intermetallics 12:1265–1274

Yang B, Morrison ML, Liaw PK, Buchanan RA, Wang G, Liu CT, Denda M (2005) Dynamic evolution of nanoscale shear bands in a bulk-metallic glass. Appl Phys Lett 86:141904

Yang Q, Mota A, Ortiz M (2006) A finite-deformation constitutive model of bulk metallic glass plasticity. Comput Mech 37:194–204

Yavari AR, Moulec AL, Inoue A, Nishiyama N, Lupu N, Matsubara E, Botta WJ, Vaughan G, Michiel MD, Kvick A (2005) Excess free volume in metallic glasses measured by X-ray diffraction. Acta Mater 53:1611–1619

Yousfi MA, Hajlaoui K, Tourki Z, Yavari AR (2013) Serrated flow model for metallic glasses under compressive loading. Acta Metall Sin 26:503–508

Zhao M, Li M (2011a) A constitutive theory and modeling on deviation of shear band inclination angles in bulk metallic glasses. J Mater Res 24:2688–2696

Zhao M, Li M (2011b) Local heating in shear banding of bulk metallic glasses. Scripta Mater 65:493–496

Zhao M, Li M (2012) Comparative study of elastoplastic constitutive models for deformation of metallic glasses. Metals 2:488

Zhao M, Li M, Zheng YF (2011) Assessing the shear band velocity in metallic glasses using a coupled thermo-mechanical model. Philos Mag Lett 91:705–712

Fundamentals of Organic-Glass Adhesion

83

Sushmit Goyal, Hyunhang Park, Sung Hoon Lee,
Mathew McKenzie, Aravind Rammohan, Hyunbin Kim,
John C. Mauro, Kyoungmin Mim, Eunseog Cho, Venkatesh Botu,
Haregewine Tadesse, and Ross Stewart

Contents

S. Goyal (✉) · M. McKenzie · A. Rammohan · V. Botu · H. Tadesse · R. Stewart
Science and Technology Division, Corning Incorporated, Corning, NY, USA
e-mail: goyals@corning.com; McKenzieME@corning.com; RammohanA@corning.com;
BotuV@Corning.com; TadesseH@corning.com; StewartRJ@corning.com

H. Park · S. H. Lee · H. Kim
Corning Technology Center Korea, Corning Precision Materials Co., Ltd., Asan, Republic of
Korea
e-mail: HyunHang.Park@corning.com; SungHoon.Lee@corning.com;
HyunBin.Kim@corning.com

J. C. Mauro
Department of Materials Science and Engineering, The Pennsylvania State University, University
Park, PA, USA
e-mail: jcm426@psu.edu

K. Mim · E. Cho
Samsung Advanced Institute of Technology, Suwon, Gyeonggi-do, Republic of Korea
e-mail: kyoungmin.min@samsung.com; eunseog.cho@samsung.com

Abstract

In this work, we review recently developed methods for quantifying the fundamentals of organic-glass adhesion. This chapter will explore modeling methods, which serve to further develop the known scientific knowledge available. The problem of building more advanced models for characterizing adhesion is approached at two different scales, viz., electronic and atomistic levels. Implementing a multiscale approach, we present our findings on the role of polymer chemistry, polymer rigidity, structure of inorganic substrates, and inorganic surface attributes such as the role of hydroxylation and surface roughness. This work is focused on physisorbed systems consisting only of nonbonded interactions between the organic and glass interface. We find that hydrogen bonding, surface roughness, electrostatics, and the flexibility of organic molecules affect adhesion the most. This result is consistent across length scales for polymers ranging from the subunit or molecular fragment scale to the full polymer film thickness. Based on this work, we propose a few salient steps to drive the modeling capabilities towards assessing functional performance of organic inorganic multilayers, which are increasingly becoming an integral part of various electronic devices.

1   Introduction

1.1   Importance of Adhesion of Organics Onto Glass

Organic-inorganic adhesion is critical for numerous practical applications, such as protective coatings, adhesives, thin-film transistors. Each organic and inorganic surface is unique and confers its own material attributes (e.g., Young's modulus or coefficient of thermal expansion) to the interface. Inorganic glass substrates including high-purity fused silica, ion-exchanged aluminosilicate glass, glass-crystal composites, soda lime silicate, and borosilicate glasses are commercially available and are widely used in industry. Thermally stable organic polymers such as polyimide (e.g., Kapton, UPILEX) are often used in high-temperature applications such as flexible display substrates, protective coatings, and adhesives because of their excellent mechanical, thermal, and chemical-resistant properties (Escott et al. 1990; Liaw et al. 2012).

Although conventional plastic materials are shock-resistive, lightweight, chemically stable, easy to form into various shapes, suitable for continuous process, insulating, and optically transparent; their poor thermal stability has limited their use as flexible substrates for display applications (Whittingham 2014; Wong and Salleo 2009). Specifically, during flexible display manufacturing, the processing temperature can rise above 300 °C, above which most plastic materials decompose (Wong and Salleo 2009). Due to these limitations, Polyimide (PI) materials have generated significant technical interest due to their thermal stability, mechanical properties, and electrically insulating nature (Chen et al. 1995; Cheng et al. 1991; Liaw et al. 2005; Ghosh and Mittal 1996). These properties have led to PI being used for a wide range of engineering applications, including laminated automobile windshields, polymer-coated optical communication fibers, and pharmaceutical glass labware with antifriction polymer coatings (Ghosh and Mittal 1996; Ni et al. 2015; Fryd and Mittal 1984). In particular, PI is one of the most favorable substrate materials used in the production of flexible displays as they largely determine performance and reliability of the display (Ni et al. 2015; Choi et al. 2010). Due to the aforementioned reasons, this chapter is dedicated to understanding the interface between PI and glass. We choose silica as the glass of interest as it helps us simplify the chemistry and understand the effect of structure and surface properties on the PI-glass interface.

The properties of glass and polymers undergo changes during processing and the unique interfacial characteristics could potentially introduce variability to both the process and the product attributes. A practical issue to contend with, during the process of flexible display manufacturing, is that PIs tend to delaminate from the supporting glassy material. This potentially increases the risk of structural failure at the organic glass interface. For these systems, the adhesion between polyimide and glass needs to be understood to avoid structural failure (Juhl et al. 2014; Chaudhury et al. 1987). The failure in these composite systems can occur at the interface between polyimide and glass, known as adhesive failure, or within the polyimide layer or the underlying inorganic layer, known as cohesive failure. Adhesive failures is one group of the dominant failure modes in electronics (Klee and Hocker 1999) and hence is the focus of this chapter.

Adhesion is the key interfacial phenomenon observed in the assembly and packaging process for electronic devices, and this interfacial adhesion is complex and is driven by many different factors such as the physical (e.g., roughness) and chemical nature of the glass surface, types of polymers, rigidity, thermal behavior, humidity, chemistry of the coupling layer between the glass and the polymer (Lacombe 2006; Awaja et al. 2009). Specifically, adhesion between polymer and glass can be influenced by the inorganic glass surface roughness (Peterson et al. 2011), surface and bulk composition (Gellert and Turley 1999), density of hydroxylated groups (Choi et al. 2008), zeta potential (Klee and Hocker 1999), and physisorbed water layers (Holbery and Houston 2006); all of which, in turn, could affect the overall functionality of the glass-polymer application.

For a functional flexible display substrate, efficient control of the adhesion between the PI substrate and the carrier glass is particularly important. This is because the adhesive interaction must be strong enough to maintain structural

integrity while also being weak enough to allow for separation of the PI from the glass; thus, the adhesion energy should be neither too strong nor too weak. If it is too strong, the detachment process causes unrecoverable damage at the interfacial region, resulting in structural failure. If it is too weak, then the glass carrier cannot perform its function of holding PI and the other components placed on its surface during assembly. Hence, developing an understanding of the adhesion mechanism and determining the governing parameters at the interface of PI and glass is critical for improving the manufacturing process.

1.2   Characterizing Adhesion Via Experiments

The degree of adhesion can be measured by experiments and is referred to as practical adhesion (Pizzi and Mittal 2003). Multiple direct measurement techniques can be used to obtain the practical adhesion (Ferreira et al. 2005; Friedrich et al. 2005; Chen et al. 2007; Mathieson and Bradley 1996; Lacombe 2006; Beake et al. 2002; Dai et al. 2006), and the most commonly used techniques are the peel and pull-out tests. In these tests, a tape or a stud is placed on the sample and pressure is applied until failure. This force or pressure required to cause failure is used to calculate the practical adhesion. While these tests provide quantitative measures of adhesion, plastic deformation in film, patterns of mechanical loading, and other factors cause inaccuracies in the measured adhesion values (Piggot 1997; Favre et al. 1997; Hampe and Marotzke 1997). An experimental study for aromatic PI films on SiO_2 suggests that no chemical bond is observed at the interface and the nonbonded interaction between the materials would play a key role in determining the adhesion strength (Miwa et al. 1993). In addition, a surface structure of PI modified by longer curing time and UV-exposure can increase the adhesion strength between PI and Si based substrates (Grady et al. 2014).

Many experimental studies exist on the topic of understanding polymer adhesion to substrates, such as glass, metals, and ceramics (Brown 2000). The literature suggests that there is continuous development of processes for improving experimental measures of adhesion (Zeng et al. 2016). Adhesion of polymer to substrates without surface treatment may be good initially; however, adhesion can deteriorate when temperature and humidity changes. Therefore, in most cases, surface modification is important to enhance adhesion of polymer to solid surfaces. Some surface treatment methods may include industrial technologies such as UV, ozone, plasma, as well as chemical modifications. In addition, microstructures on the substrate or on the film could enhance adhesion through mechanical interlocking (Larsson and Ahmad 2006). Another work highlighted the importance of van der Waals and capillary interactions for enhanced adhesion. Studies on Geckos showed the nature of adhesion of submicron features. Authors discussed that each gecko hair could produce a minuscule force $\sim 10^{-7}$ N (Geim et al. 2003). Polymer-substrate, specifically polymer-glass interactions and strength, is typically dominated by hydrophobic, hydrophilic interactions and hydrogen bonds. When coupling reagents are applied, the agents could react with adsorbed water on the surface, reducing its surface energy and by reacting with the surface hydrogen and oxides to form

covalent bonds. In some cases, physical entanglement between three chemical layers – surface, adhesion promoter, and polymer – could take place. End groups on most coupling agents such as silane could determine the type of interaction between polymer and glass. Atomic Force Microscopy (AFM) and contact angle results indicate that the type of chemical functional group in the coupling agents play a key role in determining the strength and nature of polymer-surface interactions (Dvir and Gottlieb 2009). For chemical-based treatments, typically aminosilane coupling agents are reported as effective treatment for polyimide (PI). The most popular silane, for PI, is GAPS, γ-aminopropyltrialkoxysilane that bonds to glass and PI. Typically, for chemical interactions, monolayers of silanes are preferred. For the case of thicker adhesion promoters however, adhesion could yet improve by creating an interpenetrating network in the interfacial region and the polymer matrix leading to mechanical interlocking. It is recommended to maintain the thickness of the coupling agent within 50 nm or lower for good adhesion. Overall, both chemical and physical (mechanical) interlocking have been shown to enhance adhesion of polyimide to inorganic substrates like glass (Ghosh and Mittal 1996).

As the application of glass broadens to a wider range of display devices and microelectromechanical systems (MEMS), it is important to fundamentally understand the key factors at a molecular level that determine the bonding and debonding of polymers to/from substrates. There is limited information in the literature that shows the impact of substrate-film compositions and environmental conditions such as humidity, temperature, and time. Sufficient experimental data are also lacking on morphological and surface roughness effects on adhesion. It is also relevant to understand interfacial failure modes of polymer films at the time of debonding.

1.3 Computation of Adhesion and the Need for Development

Computational modeling is a powerful tool that can be used to study the adhesion behavior of organic molecules on inorganic surfaces. Among the many tractable computational approaches for the organic-inorganic interfaces, molecular dynamics with classical atomistic force fields has been used to study a variety of inorganic glass systems (Gruenhut et al. 1997; Yang et al. 2016; Priezjev 2016; Luo et al. 2016; Zhao et al. 2016; Mantisi et al. 2015; Deb Nath 2013) and their properties, including chemical diffusivity, structure, and mechanical properties. Similarly, this approach has been used to elucidate polymer attributes, including packing ability and structure factors (Sliozberg et al. 2016; Jung et al. 2016; Aoyagi 2015), adsorption processes (Mashio et al. 2016; Shoravi et al. 2016), and rheological properties (Kang et al. 2010; Sirk et al. 2016; Sablic et al. 2016; Jabbarzadeh 2006). When combining the glass and polymers together in a molecular dynamics simulation however, a deficiency in the available tools and parameters prevents the accurate modeling of the interfacial properties (e.g., adhesion). There is some published work on characterizing certain interfacial attributes of silica nanoparticles in a polyimide matrix (Marotzke and Hampe 1997; Pizzi and Mittal 2003; Cheng and Wagner 2009). These papers do not capture the effects of electronic structure of the chemical

environment of these atoms or how they interact with the inorganic surface atoms in a concerted manner, to either enhance or limit the overall adhesion of the monomer – and ultimately the polymer – onto the surface. To address such questions at this level, we need to use more detailed methods such as density functional theory (DFT) or ab initio molecular orbital theory.

In the literature, DFT has been used to investigate different glassy systems containing atoms in the low hundreds (<400) to address problems such as the structural evolution of germanium melts (Bouzid et al. 2015), embedded clusters in silica (Du and Kroll 2010), bioactivity mechanisms of soda-lime phospho-silicate glasses (Berardo et al. 2013), interface of silicon surfaces with water (Pham et al. 2014), densification of silica (Izvekov and Rice 2012), molten glass structures (Ohkubo et al. 2016), and nanocomposite materials for solar energy collection (Wippermann et al. 2014). These studies used B3LYP (Becke 1993), PBE0 (Adamo and Barone 1999), and PW91 (Perdew and Wang 1992) functionals that have shown good predictions for the structure, effects of compositional changes on structure and electronic band structure and band gap. Using a similar set of functionals, DFT studies on polymers have included prediction of polyvinyl refractive optical properties (Yu et al. 2007), entropy of mixing and phase transitions of homopolymers (McMullen and Freed 1990), stability and reactivity of polyimides for fuel cells (Lopez-Chavez et al. 2014). The literature is very limited when it comes to the characterization of the organic-inorganic interfacial system. The closest references include carbon dioxide (Cui et al. 2016), hydrogen (Lahmer 2015), or other small molecules (Muhich et al. 2015; Gleizer et al. 2014; Ren et al. 2009) reacting on metallic transition metal surfaces.

To understand adhesion phenomena in the interfacial system comprehensively, the computational model should be capable of capturing complete atomistic details while spanning nanometer length scales which are the typical thicknesses of such interfaces. From this point of view, molecular dynamics (MD) is ideal to generate a microscopic understanding of the interaction mechanisms at the interface. Typically, for simulating inorganic materials like silica glasses, nonbonded or partially bonded force fields like Pedone et al. (2006, 2008) and BKS (Beest et al. 1990) have been used to study glass structures. Simulations with these nonbonded force fields have been used to understand the bulk properties, structure of silica, and the interfacial properties with other inorganic materials like silicon (Lee et al. 2007). These force fields, while allowing for dynamic bond breakage and formation as reflected by changes in local structure, cannot accurately describe the structure of organic molecules, which require explicitly defined angles and dihedrals. Hence, for organic molecules, bonded force fields such as PCFF (Sun et al. 1994), CVFF (Dauber-Osguthorpe et al. 1988) and COMPASS (Sun 1998) have commonly been used. While these bonded force fields have elucidated the bulk properties, structure, and changes in interfacial density distribution of the polyimide at the interface between polyimide and silica (Yang et al. 2012; Komarov et al. 2013; Odegard et al. 2005), they have not been used to systematically study the adhesion properties at the interface.

This gap in understanding of the organic-glass interfaces is partially due to the absence of reliable force fields that can accurately predict their interfacial

interaction. Since we are primarily interested in the interfacial interactions of an inorganic surface with a polymer, it is important to use a force field that accurately describes them. Potentials such as PCFF, CVFF, and COMPASS predict polymer properties well but are not optimized for interfacial interactions and surface properties such as surface tension and binding energy. Recently two force fields, ReaxFF (Mueller et al. 2010; Chenoweth et al. 2008; ACTv et al. 2001) and Interface FF (Heinz et al. 2013; Emami et al. 2014a), have shown promise in addressing these problems. Both of these force fields have their own advantages and disadvantages. ReaxFF, a force field based on bond order, is able to model reactions by describing bond formation and breakage between the molecules. The large computational time with this force field however limits the system size of the interface under study. On the other hand, the Interface FF requires less simulation time and accurately predicts properties such as contact angle, heat of immersion (Emami et al. 2014a), and adsorption of peptides (Emami et al. 2014b) but cannot be used to model reactive surfaces.

For example, a simulation study for the interfacial properties within a glassy polymer reveals that the adhesion is governed by the relation between the strength of adhesive (attraction between different materials) and cohesive (attraction within the same materials) interactions (Xia et al. 2014). This study suggests that the cohesive interaction can be affected by incorporation of adhesion agents or altering the degree of the side chain functionalization. For the adhesive interaction, its strength depends on the interface coupling and relaxation of chains at the adherent layer. In addition, Tam et al. suggest moisture as another adhesion-determining factor, which can deteriorate the adhesion strength of epoxy-silica interfaces (Tam and Lau 2015). Despite several experimental and computational investigations, comprehensive understanding of adhesion determining factors is still lacking.

The current computational literature in the area of polyimide interactions with glass does not have detailed studies of how the individual subunits constituting the organic monomers interact with the glass surface. Further, there is currently limited or no effort to understand how these interactions at the subunit level scale up to the monomer scale. Finally, the effect of adhesion of polymer chains of these polyimides on glass has also not been addressed in the literature.

1.4   Focus of This Chapter

In this chapter, we share computational methods to study the fundamentals of the organic glass interactions at three different length scales.

Electronic scale simulations for organic subunits interacting with glassy surfaces: Given the gap in understanding the detailed interactions between polyimide monomers and silicate surfaces, we focus on addressing this specific interfacial system. The composite model of polymers on a glass surface requires large system sizes to explore this question. A typical polyimide monomer, for example, BPDA 1,4,4-APB, has around 60 atoms and a small inorganic surface has around 350 atoms, hence to consider the entire system would result in a fairly large system (>400 atoms) by DFT/ab initio standards. On the other side of the interface, the

glass requires a reasonably large system size, for example, at the very least 300 atoms (for its nonperiodicity), making the interfacial system with the added polymer computationally expensive. To address the system size problem, we used two approaches to scale down the model to access more accurate quantum mechanical approaches to probe the interfacial interactions properly. To handle the polymer length, we propose an initial approach of dividing the monomer into its constituent organic subunits/molecules and probe the interaction of each of these subunits with the inorganic surface. While this simplified approach cannot address the larger scale morphological physics of the polymer material on a surface, it should allow us to capture the unique interactions between the materials more accurately than classical approaches. The learnings from this model could then potentially be used to either perform a few select studies for a larger system or develop better parameters to then study the larger system with a classical force field. To further reduce the system size, we identify surrogate crystal structures like a two dimensional (2D) silica bilayer which has certain attributes of glassy silica. Then, we systematically look at different crystalline variants (with compositions representative of industrial glasses) like kaolinite and hydroxylated albite where we investigate the effect of surface structure and composition (atomically smooth vs. more opened hydroxylated surface), compositional variation like role of Si, Al, Na, and effect of hydroxylation of the surface in controlling the adhesion. By understanding the interactions between the five different inorganic surfaces with five different organic subunits (that constitute the BPDA 1,4,4 APB polyimide), we achieve significant fundamental understanding of the key factors that drive the stickiness of inorganic surfaces to certain types of organic molecules. In our analysis, we will show a level of theory comparison for different DFT methods, explore an adhesion of single and double aromatic ring molecules across all glassy silica surfaces, a closer inspection of their adhesion to hydroxylated surfaces, and analysis of the overall trends of all subunits and surfaces.

Organic monomer interactions with glassy surfaces: At the monomer glass scale to better understand the complex nature of the interactions, in this chapter, we focus on elucidating the molecular origins of this adhesion, the effect of silica structure and hydroxylation, and the effect of polyimide structure and chemistry in greater detail. To address these questions, we present a framework to study this adhesion systematically. Specifically, we begin with careful calibration for the ability of the Interface FF to describe the bulk properties of polyimide and the surface properties of silica. We then use Interface FF to demonstrate the use of the Steered Molecular Dynamic (SMD) approach to study the effect of polyimide monomer adhesion onto crystalline and glassy silica with and without hydroxylation for four different polyimide monomers. Using the developed framework, we elucidate the role of surface structure, degree of hydroxylation, and polyimide chemistry.

Interaction of polymer chains with glassy surfaces: Finally, at the polymer chain glass scale, we attempt to understand fundamental aspects of the interfacial adhesion phenomena for various types of PIs placed on SiO_2 glass. Among several methodologies for measuring adhesion properties, the pulling test is chosen to mimic the experimental detaching process of PI from the carrier glass during the

manufacture of the display. The pulling method is applied within the framework of the Steered Molecular Dynamics (SMD) method, which has been successfully implemented in simulations by employing the same physical protocols of pulling in experiments (Tam and Lau 2015; Yang et al. 2013; Büyüköztürk et al. 2011; Lee et al. 2014). During a pulling test, adhesion determining properties such as the potential of mean force (PMF), pulling distance, and detaching force are calculated with the reactive force-field (ReaxFF). Furthermore, characteristic responses depending on the thermal properties of the polymer chains are obtained, using both coefficient of thermal expansion (CTE) values from experiments and supplemented qualitatively through simulation of variations in the chain conformation. The change in chain conformation during detachment is also examined in detail, particularly at the interfacial region. In addition, each energy term in the potential is decomposed to clearly determine which term contributes significantly to deformation during the pulling process. Finally, an investigation of the failure mechanism is performed to determine whether adhesive or cohesive failure occurs at the interface.

2 Electronic Methods

We calculate the energy of the combined system (i.e., organic subunit in the lowest energy configuration on the surface) and subtract the total of the energy of the inorganic surface and the organic subunit in vacuum. This difference is what is reported as the adhesion energy for the rest of our study.

2.1 Description of the Inorganic Surfaces

A glass surface has many attributes such as the degree of hydroxylation, density and types of cations, and type of hydroxylated species (e.g., Si vs. Al). To systematically understand the role of these different glass surfaces' attributes on the binding of organic molecules, we construct a series of surfaces that start with the simplest two dimensional silica bilayer (Sirk et al. 2016) and gradually increase the complexity by hydroxylating and adding cations. The progression of simple to complex surfaces considered for this study is shown in Fig. 1 where, following the left branch in the figure, the 2D silica bilayer is first translated into a full three dimensional (3D) silica structure, and by a hydroxylation procedure, generates the hydroxylated quartz surface. With the introduction of sodium and aluminum, we obtain a hydroxylated albite surface. Similarly, following the right branch in the figure and again starting with the 2D silica bilayer, we introduce a 3D silica structure along with the aluminum cation resulting in the kaolinite system. Kaolinite $(Al_2Si_2O_5(OH)_4)$ presents two different surfaces to the organic subunits, an aluminum and silica rich side. To distinguish these sides, we will use the notation Al-kaolinite or Si-kaolinite to indicate the respective surface being studied. Kaolinite does have aluminols as part of its crystal structure. With the addition of sodium and modification of the hydroxyl groups of kaolinite, one obtains the hydroxylated albite surface. The
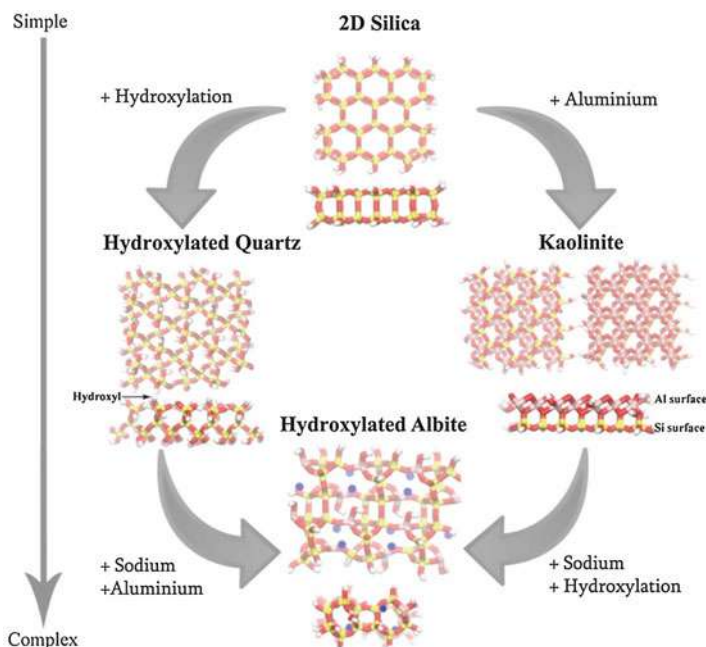
Fig. 1 Progression of surfaces, simple (2D silica) to complex (hydroxylated albite) used in this study to understand the role of surface composition. The coloring scheme that will be used for all figures is as follows: sodium (blue spheres), silicon (yellow), aluminum (pink), oxygen (red), and hydrogen (white)

Table 1 Surface number densities per nm^2 of each cluster

Surface	Hydroxyl	Sodium	Aluminum	Silicon
2D Silica	0	0	0	17.5
Kaolinite	0	0	8.8	11.3
Hydroxylated Quartz	4.9	0	0	13.0
Hydroxylated Albite	3.5	3.5	4.2	12.1

dimensions of these surfaces are roughly 20 Å × 20 Å × 10 Å, containing ~180 atoms for each system. The surfaces are net neutral in charge (by end capping with hydrogen atoms) and are obtained from the American Mineralogist Crystal Structure Database. The number densities of each elemental species are provided in Table 1.

2.2 Description of the Organic Molecules

We focus on polyimides and specifically the BPDA 1,4,4-APB since this is a commonly used system. A single monomer of BPDA 1,4,4-APB has roughly 60 atoms. To study the adhesion of a single monomer onto the inorganic surface, the

Polyimide (BPDA 1,4,4-APB)

Computational
Decomposition

(i)Phthalimide
H-bond Acceptor (ii)N-phenylmaleimide
H-bond Donor   H-bond Acceptor
                Rotatable Bond

(iii) Diphenyl Ether
Weak H-bond Acceptor
Rotatable Bond

(iv)Phenol
H-bond Acceptor
H-bond Donor

(v)Benzene
Weak pi-bonds
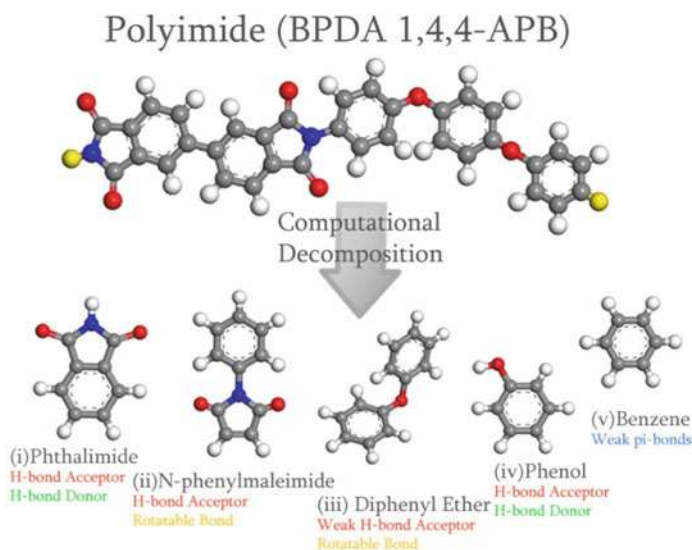
Fig. 2 Representation of polyimide monomer as a combination of 5 key smaller organic subunits to enable computations with a higher level of theory. Yellow atoms indicate the repeat bridge. Coloring scheme is oxygen (red), carbon (grey), nitrogen (blue), and hydrogen (white). Same atom coloring scheme is used throughout the paper

overall system would have over 300 atoms. These calculations are computationally expensive for the different levels of theory. To manage the computational cost while maintaining reasonable accuracy for interfacial interactions, we computationally decompose the polyimide monomer into five different subunits as shown below in Fig. 2 as explained in the Introduction. The five subunits can be divided into two categories: one-aromatic ring molecules (benzene and phenol) and two-ringed molecules (phthalimide, N-phenylmaleimide, and diphenyl ether). The differences in the chemical nature of these subunits are listed below in the Fig. 2, where it can be seen that all of the molecules besides benzene are either hydrogen bond (H-bond) donors or acceptors, suggesting that they can participate in hydrogen bond formation with the silicate surfaces. This is important since the differences in the strength of the interactions between these are almost always directly related to their hydrogen bonding tendencies.

2.3   Level of Theory Calibration and Selection of Computational Approach

We investigate both periodic and cluster systems to calibrate the optimum level of theory required for the simulation. Using the ab initio methods, we are restricted to working with a cluster model. On the other hand, DFT is used to study both cluster and periodic systems to ensure accuracy of the cluster model and of the DFT

comparison to ab initio results. By grounding our cluster calculations with the ab initio MP2 (Møller and Plesset 1934) method, we intrinsically expect to capture the van der Waals interaction more accurately and therefore better compare among all levels of theory. Additionally, we studied the DFT method with the Grimme correction (Grimme 2006a), DFT-D2, which is meant to capture the effect of van der Waals more accurately than with ordinary DFT (Grimme 2006a, b). Similar to our study, simple organic molecule adsorption on metal surfaces used DFT-D2 and was found to overestimate the adhesion energy (Tonigold and Grob 2010; Andersson 2013). The proper treatment of DFT-D parameters of metal atoms at this point remains an open question. Due to these concerns, we do not attempt to optimize the DFT-D2 correction parameters.

2.4   Initial Model Setup

For the cluster calculations, we need a good starting configuration using a periodic system to determine the important surface groups and their configuration to carve out the cluster model. First, we generate 100 initial configurations using a standalone script. This script applies rotations and translations to the organic subunits with respect to the surface. These configurations are used for geometry optimization in VASP (Kresse and Hafner 1993) using PBE (Perdew and Wang 1992) with PAW (Blochl 1994) and a plane wave cut-off energy of 250 eV with a vacuum size of 30 Å. From these 100 different configurations, a certain fraction (typically 10%) was discarded due to unfavorable contacts with the surface. We identify the lowest energy configuration of each structure to create the cluster system, which is used for the level of theory calculations and other analyses.

2.5   Differences in Adhesive Behavior of Phenol and Phthalimide on the Five Different Inorganic Surfaces

In this section, we compare two subunits (single ring phenol and double ring phthalimide) across all the different surfaces. From Fig. 2 we recall that phenol has a hydroxyl group that can participate in hydrogen bond formation. The phthalimide has two carbonyl oxygen groups, which can again participate in hydrogen bond formation. The other structural difference to note is that phenol has one aromatic ring, while phthalimide has one aromatic ring in conjunction with a five-membered ring containing nitrogen. The presence of nitrogen also contributes to the increased electronegativity, which could promote further interactions.

Given these structural differences, we see that the adhesion of both of these molecules increases when progressing from 2D silica to hydroxylated albite. For phenol, moving from the 2D silica to Si kaolinite surface the adhesion energy increases only slightly from −11.6 kJ/mol to −11.7 kJ/mol. Although the Si kaolinite surface structure does have an aluminol embedded, its surface is very similar to 2D silica including the adhered. When we consider the interaction between phenol

and the Al kaolinite surface, the adhesion increases to -12.1 kJ/mol. Phenol adhered structures are much more planar on the Al kaolinite surface than on the Si kaolinite. The phenol distance from the surface is similar between both Si and Al kaolinite (since this measurement is taken from the plane of the Al (for Al kaolinite) or Si (for Si kaolinite)). The hydroxyl hydrogens in kaolinite are there for charge compensation of the alumina layer. Due to this charge compensation and confined Al geometry, upon adhesion, these hydrogens are not able to rotate as much as a hydroxyl group found in other hydroxylated surfaces (e.g., hydroxylated quartz).

On hydroxylated quartz, we see that the H-bond distance of the phenol is shorter than on Al kaolinite with an angle of 35° vs. 15° (more inclined and less planar), clearly suggesting that the increased adhesion (by 3.2 kJ/mol compared to Al kaolinite) is being driven by the enhanced hydrogen bonding between both the surface hydroxyl groups and that of phenol. Finally, on hydroxylated albite, which presents Si/Al/Na/O and OH groups on the surface, we see that the adhesion increases by 8.7 kJ/mol with respect to 2D silica. Furthermore, when we compare the adhesion with respect to Si kaolinite and Al kaolinite, we see an increase of 7–8 kJ/mol. Comparing hydroxylated quartz with hydroxylated albite, we see an increase of ∼5 kJ/mol. We also observe that the distance of phenol from the inorganic surface is greatest for the hydroxylated albite substrate; however, the hydroxylated albite surface has inherent roughness and hence the distance is less well defined compared to the other smooth surfaces likes 2D silica. The orientation of these organic molecules are also less planar on hydroxylated albite, but now two hydrogen bonds are observed, one being mediated by the surface hydroxyl groups and the other interestingly by the sodium atoms. This allows us to conclude that the presence of sodium atoms on the surface enhances the adhesion by close to 5 kJ/mol and is largely due to the electrostatic nature of the interaction between the positively charged cation (Na) and the negatively charged hydroxyl group (OH) on phenol. The presence of surface sodium also brings the surface hydroxyl groups closer to the sodium, which potentially strengthens the hydrogen bonding between these hydroxyl groups and that on phenol.

Turning our attention to the adhesion of phthalimide, we see some similarities in the behavior with phenol and some interesting differences. As with phenol, the adhesion energy of phthalimide progressively increases from 2D silica to hydroxylated albite. For both the kaolinite surfaces, these two subunits adhere in a more planar configuration due to the same two reasons: (1) Si kaolinite is atomically smooth like 2D silica and (2) the rigid noninteracting aluminols prevent any direct hydrogen bonding. However, the degree of adhesion of phthalimide onto Si kaolinite and Al kaolinite is weaker compared to phenol on the same two surfaces. We attribute this difference to the fact that phthalimide is a larger molecule containing two rings, which enhances its rigidity and steric hindrances, especially for less planar surfaces. Notice the difference between phenol and phthalimide on the hydroxylated quartz surface. Here, the adhesion energy is similar (phenol -15.3 kJ/mol vs phthalimide -16.3 kJ/mol) but the angle, which the subunit makes with respect to the surface, is very different (35° for phenol vs. 16° for phthalimide). This difference in orientation is driven by the fact that phthalimide makes three

hydrogen bonding contacts with the surface while phenol makes only one. These additional bonds restrict the orientation on the surface increasing the planarity of the molecule.

When we examine the behavior of phthalimide on hydroxylated quartz and hydroxylated albite, we see that phthalimide finds these two surfaces highly adhesive. On hydroxylated quartz, the presence of the surface hydroxyl groups clearly mediates at least 3 hydrogen bonds, resulting in a much flatter orientation of the phthalimide with respect to the surface. This change in orientation clearly allows more of the phthalimide groups to interact with the surface, which results in an increase of adhesion energy by 1 kJ/mol compared to phenol. On hydroxylated albite, the presence of the positively charged cation appears to increase the overall electrostatic interactions between the two negatively charged carbonyl oxygens on phthalimide and the electronegative nitrogen atom and contribute to an 8 kJ/mol increase in the adhesion. This is greater than the 5 kJ/mol increase we saw with phenol's adhesion moving from hydroxylated quartz to hydroxylated albite.

These two comparisons bring out some of the fundamentals of adhesion of organic molecules onto surface. There is a balancing act between the polar groups' nonbonded interactions and the subunit's planarity/rigidity that drives how close the adhered molecule can approach the surface. From the surface perspective, composition changes can alter both structure and electrostatic environment as seen from both the Si kaolinite and the Al kaolinite surfaces. Understanding these adhered configurations with respect to surface and subunit characteristics, one could possibly formulate an optimized adhesive material by manipulating either the inorganic surface attributes like type of cation or tailoring the organic moieties like presence/absence of ring like structure, nature of chemical groups, size, etc., through computer-aided material design techniques.

2.6   Differences in Adhesive Behavior of the Five Different Organic Molecules on Hydroxylated Inorganic Surfaces

2.6.1   Behavior of Single Ringed Organic Molecules

When we compare the behavior of benzene and phenol adhesion to the two hydroxylated surfaces, it is to be expected that phenol binds more strongly on both these surfaces (due to hydrogen bonding) and further overall adhesion to hydroxylated albite is tighter. This is in line with the analysis presented in the earlier section where the presence of surface hydroxyl groups, sodium cations, and aluminum were all involved with enhancing the adhesion energy of hydroxylated albite. The increase in adhesion energy between phenol and benzene is -12.2 kJ/mol on hydroxylated quartz, while on hydroxylated albite it is -12.5 kJ/mol. The overall adhesion of both benzene and phenol individually increased on hydroxylated albite, while the net difference in adhesion was comparable to that observed on hydroxylated quartz.

2.6.2   Behavior of Double Ringed Organic Molecules

With respect to the double-ringed structures, we explained earlier that phthalimide has three key atoms that play a role in enhancing the adhesion: the two carbonyl oxygen atoms and the nitrogen atom. When we compare the N-phenylmaleimide structure with the phthalimide structure, we note that they both contain these two carbonyl oxygen atoms and the nitrogen atom. However, N-phenylmaleimide's nitrogen serves as the linker atom between the 5 membered ring structure (forming tertiary amine) and the phenyl group while in phthalimide the nitrogen atom is freely available (forming secondary amine) and is bonded to a hydrogen atom. This small but subtle change from a tertiary to secondary amine is likely to result in differences in the adhesive behavior because the nitrogen can access the aromatic ring's delocalized π cloud (lowering nitrogen's affinity to any surface group) and this is a planar sp^2 imide nitrogen.

Diphenyl ether has two aromatic rings that are linked by the ether bond through the oxygen atom. This oxygen has reasonable access to the delocalized pi electrons from either of the aromatic rings, and the ether bond is known to be less rigid than the nitrogen-carbon bond seen in N-phenylmaleimide. This is likely to confer overall a slightly larger flexibility for the two aromatic rings that constitute the ether, which could potentially handle steric hindrances a bit more effectively. When we look at N-phenylmaleimide and diphenyl ether, one might have expected that the N-phenylmaleimide should have had stronger adhesion given that functionally it has the two carbonyl oxygens and the nitrogen atom. We however observe that the adhesion energy is comparable to that of diphenyl ether and perhaps just a bit higher (-0.8 kJ/mol) on hydroxylated quartz while actually being a bit lower on the albite surface (-0.5 kJ/mol). These results in light of the earlier analysis reinforces that the electronegativity of the N-phenylmaleimide while helping adhesion is perhaps being limited by its intrinsic rigidity or limited conformational flexibility.

The chemical nature of the organic molecules plays a large role in modulating the adhesion. Specifically, certain locations can alter the chemical environment significantly. For instance, N-phenylmaleimide has the same three chemical groups as phthalimide, but changes in rigidity and limited availability of all the atoms to participate in the surface interaction limit the overall adhesion. In addition, the size of molecules affects adhesion and smaller molecules have lower steric hindrances and hence greater access to the surface, but the presence of certain functional groups like the carbonyl oxygens and nitrogens can enable better interactions with surface groups, which could offset the steric hindrances with heightened adhesive interactions. In other words, larger molecules may experience larger steric hindrance but can adhere more tightly due to the increased interactions of the additional functional groups. These trends help in explaining the fundamental nature of what drives the interactions and what can be leveraged to modulate the surface interactions.

2.7   Summary and Conclusions

The overall result summary is shown in Fig. 3. Benzene is found to have the weakest adhesion which is understandable given that it is an aromatic molecule with very little to no groups that can have nonbonded interactions with our inorganic surface either through hydrogen bonding or electrostatic interactions. Having an easily rotatable bond, phenol has relatively strong adhesion on all surfaces and does follow the progressive trend from 2D silica to hydroxylated albite. Being able to both hydrogen bond donate and accept is a key attribute to increase adhesion on hydroxylated surfaces as seen by phthalimide. From the most complex in compositional space and being hydroxylated, albite was the most adhesive surface. To summarize, we observe four points that mediate interactions at the organic-inorganic interface;

1. Hydrogen bond formation is a critical factor in mediating the nonbonded interactions.
2. Electrostatics can play a significant role as seen from the role of sodium in the presence of negatively charged groups like hydroxyl.
3. Changes in surface composition can alter both structure and electrostatic environment as seen from both the kaolinite surfaces which hydrogen bond accepting subunits are favored to bind on the alumina side.
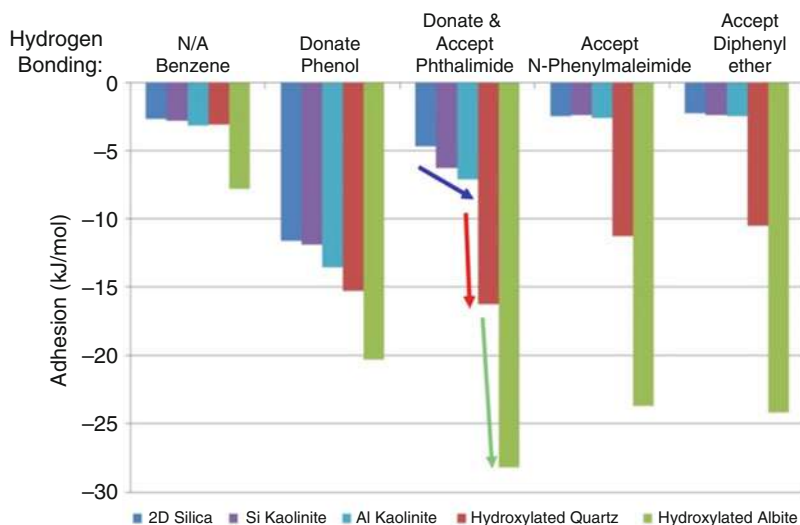
Fig. 3 Plot showing the combined trends of hydroxylation, aluminol addition, and cation across all surfaces and subunits. Arrows indicate the adhesion trends going from 2D silica to kaolinite (blue arrow), to hydroxylated quartz (red arrow), then to hydroxylated albite (green arrow)

4. Rigidity of the organic subunit can enhance the adhesive contact by lowering the conformational degrees of freedom, which allows for additional stability with respect to the surface interactions. For instance, N-phenylmaleimide has the same three groups as phthalimide, but changes in rigidity limit the availability for all the atoms to participate in the surface interaction thereby limiting the overall adhesion.

Lastly, our analysis shows good agreement between classical and DFT-based reconstituted monomer adhesion ratios of Kapton to BPDA 1,4,4-APB, which also gives us some confidence in the classical Interface force field (Brown 2000). Our future work will involve using DFT to characterize the torsional potentials for the nitrogen groups in the Interface potential. We believe this can potentially help better characterize the stacking observed with these kinds of polyimides.

3   Atomistic Methods

3.1   Monomer-Glass Adhesion

3.1.1   Simulation Details

The specific steps that we follow in setting up our simulations are described in this section. First, we create and validate silica structures. We then combine polyimide monomers with silica structures to form the interfacial system and use Steered Molecular Dynamics (SMD) to calculate the adhesion between them. We describe the details of the methods for each of these steps in the sections below.

Force Field

The interface force field, developed by the Heinz group in the past few years, is a classical force field for organic, biomolecules, and inorganic components at the 1–100 nm scale (Heinz et al. 2013). The core of this force field is based on the consistency of classical Hamiltonians for organic and inorganic force fields like PCFF, CVFF, CHARMM, etc., and has been extended to accurately model interfaces for surface energy (Heinz et al. 2013), biomaterials (Patwardhan et al. 2012), and polymers (Emami et al. 2014a). This extension has been carried out by tuning force field parameters to match experimental and ab initio results for surface phenomenon like hydration energy of a silica surface in water (Xia et al. 2014), adsorption isotherms of water on silica (Xia et al. 2014), and selective peptide recognition and binding to silica surfaces (Emami et al. 2014b). The Interface potential consists of Lennard Jones and Columbic interactions as nonbonded interactions, harmonic bonds to mimic covalent bonds, and harmonic angles to model the structure of polymers more accurately.

In our work, we have used the PCFF-interface potential, which is an extension of the well-established PCFF potential. This is a class II potential and has the form

$$
\begin{aligned}
E_{pot} = &\sum_{ij, nonbonded\ 1,2\ and\ 1,3\ excluded} \varepsilon_{ij} \left[2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^9 - 3\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right] \\
&+ \frac{1}{4\pi\varepsilon_0} \sum_{ij, nonbonded\ 1,2\ and\ 1,3\ excluded} \frac{q_i q_j}{r_{ij}} + \sum_{ij, bonded} k_{r,ij}\left(r_{ij} - r_{0,ij}\right)^2 \\
&+ \sum_{ijk, bonded} k_{\theta, ijk}\left(\theta_{ijk} - \theta_{0,ijk}\right)^2
\end{aligned}
$$

More details on the formulation and the description of individual terms can be found in the original paper (Heinz et al. 2013).

Creating Silica Structures

The Interface FF requires the bonds and angles between atoms to be explicitly defined, which makes generating a silica structure slightly more complex than when using a nonbonded potential. To ensure that our approach can readily be extended to other inorganic materials, we start by generating the silica glass structure using the well-established nonbonded Pedone potential (Pedone et al. 2006, 2008) and a simulated annealing procedure (Tandia et al. 2012). A flat surface is created by opening the simulation box and hydroxylating the under-coordinated oxygen atoms exposed to the surface to achieve a surface density of ~ 4.7 OH/nm^2. A corresponding nonhydroxylated surface was made by deleting –H and –OH pairs and introducing Si-O-Si bonds, followed by energy minimization. This method mimics the extreme condition of dehydration and provides a surface that is free of high-energy dangling bonds. The crystalline silica under study is the α-quartz structure obtained from the model structure database shared with the interface potential (Emami et al. 2014a). This structure is generated by cleaving bulk α-quartz along the $(00\bar{1})$ plane, which creates a perfect Q3 surface (where each silicon at the surface is bonded to three bridging oxygen atoms), and on hydroxylation of these silica surfaces, a system with 4.7 OH/nm^2 hydroxyl density is obtained. The size of the glassy silica surface is 4.03 nm $\times$4.14 nm and for α-quartz it is 3.33 nm $\times$3.34 nm.

Polyimides Under Study

We choose four different polyimide monomers for this study. Three of them have the same $3,3'4,4'$-biphenyl tetracarboxylic dianhydride (BPDA) with one of three different APB isomers: 1,3-bis(3-aminophenoxy)-benzene, 1,3-bis(4-aminophenoxy)-benzene, or 1,4-bis(4-aminophenoxy)-benzene. BPDA has been chosen as the benchmark reference case for its bulk properties and microstructure is available in the literature. The final polyimide under study is pyromellitic dianhydride and 4, $4'$-oxydiphenylamine (Kapton) which is one of the most commonly used polyimide. Throughout this chapter, they will be abbreviated as BPDA 1,3,3-APB, BPDA 1,3, 4-APB, BPDA 1,4,4-APB, and Kapton, respectively. Among them, the first three BPDA series have the same chemical compositions except for the oxygen linkage connecting the benzene rings in the APB section. BPDA 1,3,4-APB has one meta linkage and two para linkages, while 1,3,3-APB and 1,4,4-APB have three meta
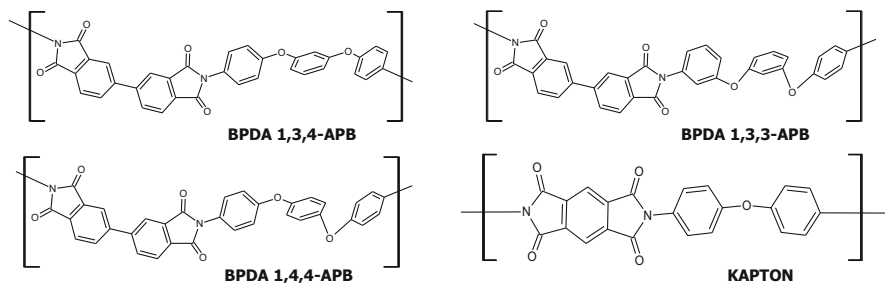
Fig. 4 Repeat units for the Polyimide monomers for three BPDAs (left) and Kapton (right)

and three para linkages, respectively. Kapton has a different chemical structure from the BPDA series (See Fig. 4 for the structure in detail). By choosing these polyimide sets, we expect to see the effect of conformational variability (BPDA) and in chemical composition (BPDA vs. Kapton) on adhesion. Details for creation of the bulk polymer are provided in literature (Goyal et al. 2016).

Modeling the Composite System
We model the polyimide-silica composite to consist of the silica surface with and without hydroxylation (0% and 100%) and a polymer phase. The silica surfaces considered are the α-quartz crystalline surface and the glassy silica surface. The polyimide monomers considered are BPDA 1,3,3-APB, BPDA 1,3,4-APB, BPDA 1,4,4-APB, or Kapton. We use thermal annealing to equilibrate the polyimide monomers on the silica. In order for the monomer on the silica to arrive at the equilibrium conformation completely, the thermal annealing and relaxation procedures are performed by MD simulation with the Interface-CVFF force field. The composite structure is relaxed at 500 K for 500 ps followed by annealing to 300 K for 500 ps with the NVT ensemble using the Nosé-Hoover thermostat. Then, the final equilibration is performed at 300 K for 500 ps with both NVT and NPT ensembles. This procedure helps the monomer to overcome the local minima in the free energy landscape and thus quickly reach the global equilibrium state. The time step was 1 fs, the cutoff for van-der-Waals interactions was 12 Å, and the Particle-Particle Particle-Mesh (PPPM) solver was used for the summation of long-range Coulomb interactions with the precision of 10^{-4}.

3.1.2   Steered Molecular Dynamics
Calculations for the work of adhesion as well as the investigation of the local chain relaxation require performing simulations for a system with more than several hundred atoms as well as considerable relaxation time. In addition, the adhesion behavior between polymer and glass is highly affected not only by their interfacial interactions, but also by the strength of interactions within polymer chains themselves. Within these considerations, the classical Molecular Dynamics (MD) method is a suitable tool for the examination of dynamics at the atomic level

accounting for energy contributions of the interfacial strength and chain dynamics. Within the MD framework, the SMD methodology can be used to examine the chain conformational behaviors at the interfacial region (Park et al. 2003; Jarzynski 1997; Park and Schulten 2004). The SMD method has been widely utilized for investigation of various materials such as polymeric systems and bio-materials (Hamdi et al. 2008; Isralewitz et al. 2001; Davis et al. 2009; Chabria et al. 2010). It is particularly useful to provide insights for deformation-induced conformational changes and subsequent energetic variations in the system.

During SMD simulations, atoms are attached to so-called dummy atoms via a virtual spring. The dummy atoms are displaced by applying a constant velocity (v) to them. A harmonic spring model is used to generate the potential (U), which is defined as

$$U = \frac{1}{2}k\left[vt - \left(\overrightarrow{r} - \overrightarrow{r}_0\right) \cdot \overrightarrow{n}\right]^2 \tag{1}$$

where k is the spring constant, t is time, $\overrightarrow{r}$ and $\overrightarrow{r_0}$ are the current and the initial positions of the atoms, and $\overrightarrow{n}$ is a unit vector in the direction of pulling. While the pulling test is being performed, the atoms are still allowed to relax. The generated force ($\overrightarrow{F}$) due to the applied velocity is defined as,

$$\overrightarrow{F} = -\nabla U \tag{2}$$

The averaged work of adhesion is based on Jarzynski's equality, which relates the equilibrium quantity (i.e., the potential of mean force (PMF)) to the nonequilibrium process (Park et al. 2003; Jarzynski 1997; Park and Schulten 2004). The PMF value is computed from the following Eq. (Ghosh and Mittal 1996).

$$PMF = -\frac{1}{\beta} \log\left\langle e^{-\beta W}\right\rangle \tag{3}$$

where $\beta = 1/(k_B T)$ with the Boltzmann constant kB and the temperature T of the system, W is the work done during the pulling test, and the angular bracket indicates the ensemble average of the given quantity. T is the temperature which is 300 K for our simulations. In our simulations, we use a spring constant, *k*, equal to 100 kcal/mol/Å^2.

To obtain accurate evolution of the PMF with polyimide on silica, the SMD simulation should be maintained in the quasi-equilibrium state during the entire process. In this regard, as stated above, testing the effect of pulling velocity is of great importance for the precise calculation of the adhesion between the polyimides and silica substrates. It is particularly true for the monomer structures, which are very sensitive to any driving force toward a nonequilibrium state.

System	Pulling distance at detachment (Å)
c-SiO$_2$ with 100% OH	13.9
c-SiO$_2$ with 0% OH	13.0
g-SiO$_2$ with 100% OH	15.1
g-SiO$_2$ with 0% OH	13.6
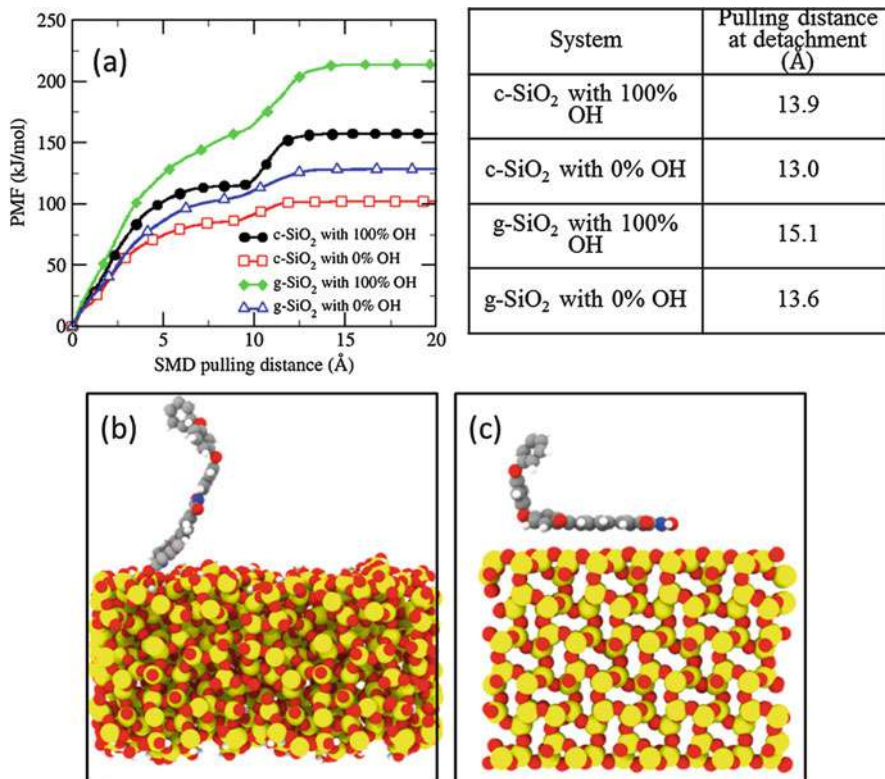
Fig. 5 (**a**) The distance at which BPDA separates from the silica surface. We observe that hydroxylated and glassy surfaces need higher separation for completely detaching. The reason for the larger separation distance is the extended hydrogen bonding which can be observed in hydroxylated surfaces as shown in (**b**) but is absent in nonhydroxylated surfaces as shown in (**c**)

3.1.3   Effect of Surface Hydration and Silica Structure

We start our analysis by studying the adhesion of BPDA 1,4,4-APB on various silica structures and surface hydroxylation. First, we analyze the effect of hydration on adhesion for both the crystalline and glassy silica surfaces. We observe that hydroxylated silica has higher adhesion by 50 kJ/mol for α-quartz and 100 kJ/mol for glassy silica; the PMF curve is shown in Fig. 5a. This trend is independent of silica structure as this result holds true for both crystalline silica as well as glassy silica. We find that the enhanced adhesion of hydroxylated silica vs. nonhydroxylated silica can be explained by the extended interactions between the silica surface and the polyimide. The presence of surface hydroxyl groups causes an increase in hydrogen bonding between the polyimide and surface hydroxyls, resulting in an increased adhesion, as well as a longer pulling distance.

In addition, we observe that the adhesion for glassy silica is always higher than that for their crystalline counterparts by $\sim$75 kJ/mol for fully hydroxylated surfaces

and ~25 kJ/mol for nonhydroxylated surfaces. To understand the detachment process, we next analyze the distance at which the monomer completely detaches from the silica surface. From Fig. 5a, we can see that a longer separation distance is required to detach the monomer from glassy silica when compared to crystalline silica, and for hydroxylated structures when compared to nonhydroxylated structures. This is illustrated in the snapshots of the systems shown in Fig. 5b and c where due to the presence of hydroxyl layers and glassy structure, the polyimide monomer clearly has a stronger interaction with silica. Furthermore, the glassy silica structure has a higher adhesion and a larger saturation distance than crystalline silica. We speculate this occurs due to the increased sites for polymer-surface interaction for the random surface of glassy silica as compared to a uniform surface for the crystalline α-quartz. We find that these results are influenced by the silica structure and hence do not deviate upon changing polyimide chemistry.

3.1.4 Effect of Polyimide Chemistry

To understand the effect of polymer chemistry on adhesion, we compare the adhesion of BPDA 1,4,4 and Kapton monomers on various silica structures with and without hydroxylation. The results are shown in Fig. 6 which shows the results
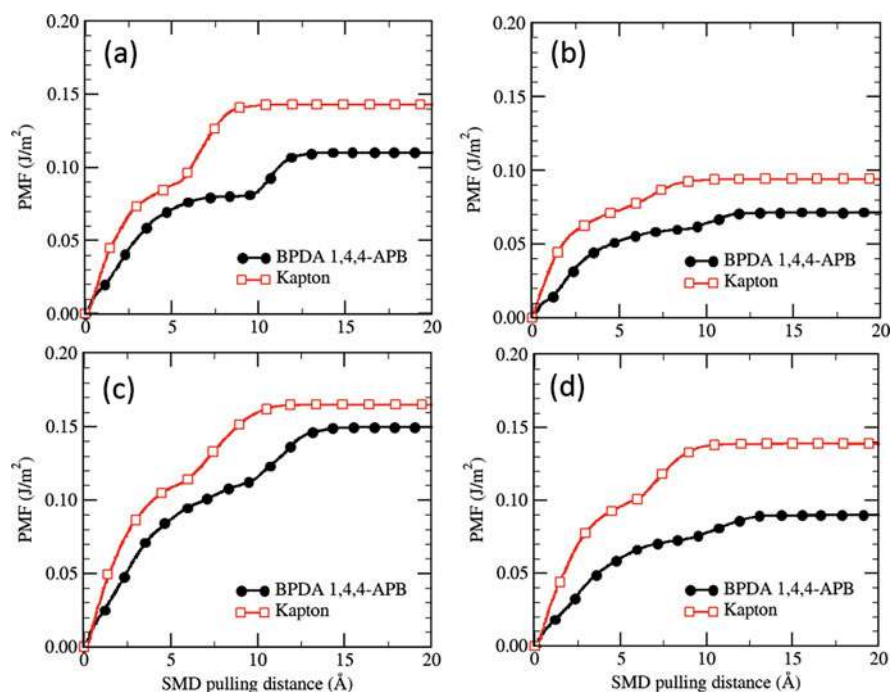
Fig. 6 Comparison of adhesion of polyimide and silica for BPDA (black) and Kapton (red) per projected surface area for (**a**) crystalline silica fully hydroxylated, (**b**) crystalline silica with no hydroxylation, (**c**) glassy silica with full hydroxylation, (**d**) glassy silica with no hydroxylation

for the normalized adhesion and Kapton adheres more strongly than BPDA for every silica structure interaction.

To understand the molecular origin of this discrepancy, we split the polyimide + hydroxylated silica system into multiple subsections and track the potential energy change during the SMD process. This allows us to understand the evolution of interactions between the silica and polyimide surfaces during the detaching process. The system is broken up into the following subgroups: (1) the bulk silica system, (2) the surface atoms for hydroxylated silica, (3) carbon backbone of polyimide, (4) hydrogen attached to polyimide, (5) nitrogen in polyimide, and finally (6) oxygen on the polyimide. It should be noted that all oxygen, carbon, and hydrogen atoms on the polymer are not identical, and hence this analysis is only useful to obtain a qualitative contribution to adhesion by each element. To understand the effect of monomer constituents on adhesion, we normalize the potential energy change by the net surface area of the respective groups of atoms, calculated using the van-der Waal radii. The results are shown in Fig. 7, and we find that the major contributor to the potential energy change per unit area is in the following order: oxygen > carbon ~ nitrogen > surface hydroxyls > hydrogen in polyimide. This clearly suggests that for a given surface area of polymer, the higher the number of oxygen atoms, the higher will be the net adhesion with silica. On comparing BPDA to Kapton, we find that the concentration of oxygen is significantly higher in Kapton (almost 1.5 times), which leads to a higher adhesion per unit area when compared to BPDA. The higher net
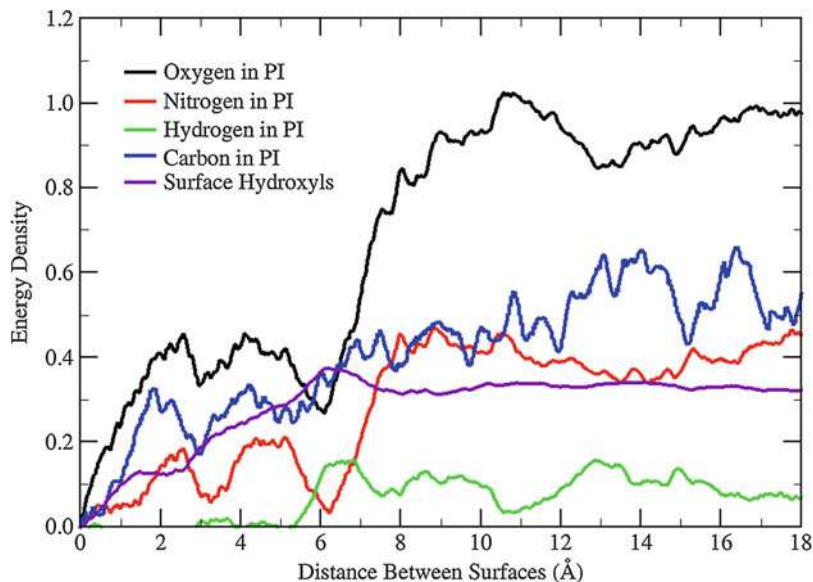
Fig. 7 The adhesion per atom per unit area for each of the subgroups studied in Fig. 8. Note that bulk silica has been omitted due to its minimal effect on adhesion. We can clearly see that oxygen has the largest effect on adhesion followed by carbon, nitrogen, surface hydroxyls, and finally hydrogen

adhesion of BPDA is due to a longer carbon chain (34 for BPDA vs. 22 for Kapton) which is the second highest contributor to potential energy. However, the number of carbon atoms per unit area is similar between Kapton and BPDA, and therefore carbon does not contribute to the overall trend for adhesion per unit area. Hence, we conclude that due to the higher concentration of oxygen in Kapton, the adhesion per unit area is higher than BPDA.

3.2   Polymer Glass Adhesion: REAXFF

We next study the adhesion between silica glass and polymer film to understand the effect of chain rigidity.

3.2.1   Simulation Details with REAXFF

The Reactive Force-Field (ReaxFF) was proposed by van Duin and has been recognized to accurately define the bond formation and dissociation with consideration of bond order at every time step (Liaw et al. 2012; Chen et al. 1995). Hence, this potential is expected to capture the complex adhesion behavior at the interfacial region of PIs and SiO$_2$. ReaxFF describes the connectivity between atoms by the following energy terms:

$$E_{system} = E_{bond} + E_{val} + E_{tors} + E_{over} + E_{under} + E_{pen} + E_{conj} + E_{H-bond}$$
$$+ E_{vdW} + E_{Coul} \tag{4}$$

Each term in Eq. (4) represents bond (E_{bond}), valence-angle (E_{val}), torsion (E_{tors}), over-coordination (E_{over}), under-coordination (E_{under}), penalty (E_{pen}), conjugation effect (E_{conj}), hydrogen bonding (E_{H-bond}), van der Waals interaction (E_{vdW}), and Coulomb interaction (E_{Coul}) energies, respectively.

Initial atomic structures of PIs are constructed using the Amorphous Cell generator in Materials Studio. All PIs are built from five chains of linear PI, each six monomers long. The PIs and glassy SiO$_2$ structures are confined in the z-direction to have a surface area of 28.51 Å by 28.51 Å in x-, y-direction with periodic boundary conditions along those axes. The layer thickness of both structures is around 20 Å. After PI construction, it is placed onto the surface of SiO$_2$ with a vacuum distance of 100 Å in the z-direction, followed by structural relaxation. The hybrid structure is first relaxed with the NVT ensemble for 1 ns at 300 K then with the NPT ensemble for another 1 ns at 1 atm and 300 K. The NPT ensemble is applied only along the x- and y-directions to preserve the vacuum in the z-direction. MD simulations are conducted using the LAMMPS package (Gruenhut et al. 1997). The Nosé-Hoover thermostat and barostat are applied for the temperature and pressure control with damping parameters of 100 fs and 1000 fs, respectively. The velocity-Verlet time integration scheme is used with a time step of 0.5 fs. After structural relaxation, the pulling test using SMD is performed under the NVT ensemble with a constant
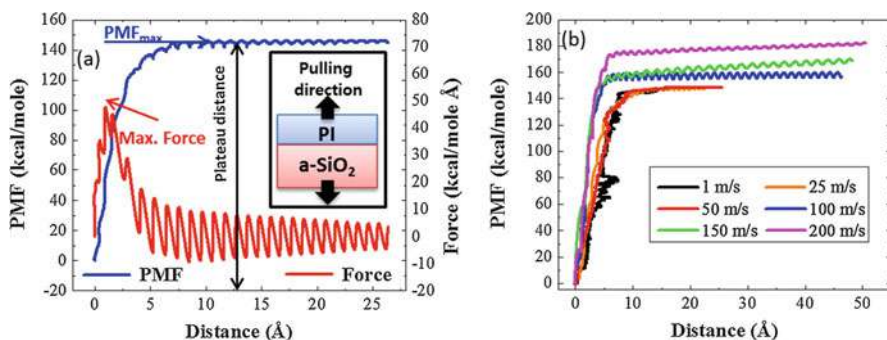
Fig. 8 (**a**) The schematic description for adhesion properties of PMFmax, maximum force, and plateau distance. (Inset) Schematic view of pulling process. (**b**) The PMF convergence as a function of SMD rate for DHBZ-BTDA on SiO$_2$

velocity of 50 m/s. The pulling force is applied perpendicular to the interface of SiO$_2$ and PI with a spring constant of 100 kcal/mol/Å^2.

The adhesion behavior at the interface can depend on the velocity used for the SMD simulation; thus, it is critical to ensure that the adhesion properties are converged under various ranges of velocities. Here, the convergence behavior of the PMF for BTDA is obtained when an applied velocity is decreased from 200 m/s to 1 m/s, as shown in Fig. 8. At velocities slower than 50 m/s (i.e., 25 m/s and 1 m/s), PMF responses are observed to be independent of SMD rates. Consequently, a velocity of 50 m/s is chosen as an optimal value and applied to the rest of the simulations.

3.2.2 Structural Properties

Initial (bulk) structural properties for all PI structures are shown in Table 2. An initial density is computed to be 1.25–1.32 g/cm^3 for all structures. In addition, the radius of gyration (Rg) and the end-to-end distance of a chain (REE) are calculated to compare differences in initial chain conformation between PI structures. Rg is defined as

$$R_g = \sqrt{\frac{1}{N+1} \sum_{i=0}^{N} \left\langle (\vec{r_i} - \vec{r_G})^2 \right\rangle} \tag{5}$$

$$r_G = \frac{1}{N+1} \sum_{i=0}^{N} r_i \tag{6}$$

where N is the number of atoms, r$_i$ is the location of the ith atom, and r$_G$ is the center of mass.

REE is defined as REE $= |r_N - r_n|$, where r$_N$ and r$_0$ is the coordinate of the atom at both ends in the polymer chain. R$_g$ and REE values are averaged over all individual chains.

Table 2 Temperature dependent adhesion properties for DHBZ-BPDA and DHBZ-DTDA on a-SiO$_2$

Structure	[DHBZ-BPDA] + SiO$_2$			[DHBZ-DTDA] + SiO2		
	300 K	450 K	550 K	300 K	450 K	550 K
Adhesion energy [kcal/mole•Å^2]	0.1787	0.0983	0.1023	0.1673	0.1132	0.1031
Plateau distance [Å]	5.01	8.37	8.24	22.46	17.59	15.20
Maximum force [kcal/mole•Å]	76.12	49.35	48.807	39.60	37.949	32.48
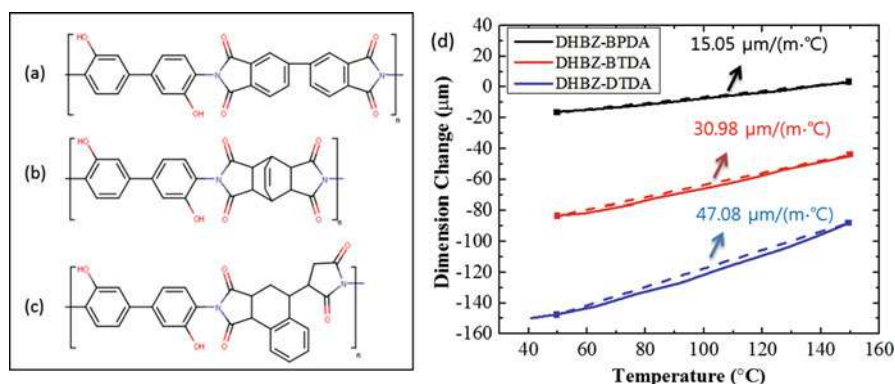
Fig. 9 The molecular structure of (**a**) DHBZ + BPDA, (**b**) DHBZ + BTDA, and (**c**) DHBZ + DTDA. (**d**) The coefficient of thermal expansion (CTE) for polyimide structures from experiment. The solid line is the experimental data and the dashed line is the fitted data to obtain CTE values between 50 °C and 150 °C

R$_g$ and REE values provide qualitative information that can be used to describe flexibility/rigidity of polymer chains (Yang et al. 2016; Priezjev 2016). For example, BPDA exhibits the largest R$_g$ and REE, indicating that its chain is easily bent or twisted under external force when compared to other types of polymer chains having smaller values of Rg and REE. The rigidity of polymers can be further characterized by experimental CTE values shown in Fig. 9, since polymers with higher rigidity tend to exhibit lower CTE values (Luo et al. 2016; Zhao et al. 2016). As expected, BPDA has the smallest CTE value, which is consistent with its larger values of R$_g$ and REE. Finally, to explain the rigidity difference in polyimides, Young's modulus (E) is obtained based on the stress-strain curve.

E values at 300 K for BPDA, BTDA, and DTDA are computed to be 8.2 ± 0.5, 5.3 ± 0.3, and 3.2 ± 0.2 GPa, respectively. This means the structures with lower CTE values exhibit the superior mechanical property. This demonstrates the correlation between CTE, Rg, REE, and the mechanical property in a consistent

manner. Accordingly, we associate these four properties to represent the chain rigidity and explain subsequent differences in adhesion properties for PIs.

3.2.3   Adhesion Properties and SMD Rate Dependency

Calculation of the binding energy at the PI/glass interface is a common way of defining how strongly they interact. However, it cannot provide atomistic information to predict the actual dynamic behavior such as how the chain conformation would evolve when entangled polymer chains are being separated from the glass. In this regard, it is essential to introduce properties that can broaden the understanding of interfacial behaviors. For a quantitative comparison of the adhesion behavior between PI and SiO_2 glass, three descriptive properties are introduced in this study. These are shown in Fig. 10a and include adhesion energy (EAdh), maximum force (FMax), and pulling distance (D_p) at the plateau of EAdh. The EAdh is defined as the potential of mean force (PMF) divided by the projected interfacial area when PI and SiO_2 are fully separated from each other. FMax is the maximum force that the system reaches during detachment. D_p is the center-of-mass distance between PI and SiO_2 when they are completely detached.

The adhesion behavior at the interface could be dependent on the velocity used for the SMD simulation; thus, it is critical to capture the converged value of the adhesion properties under various ranges of velocities. Here, the convergence behavior of PMF for BTDA is obtained when an applied velocity is decreased from
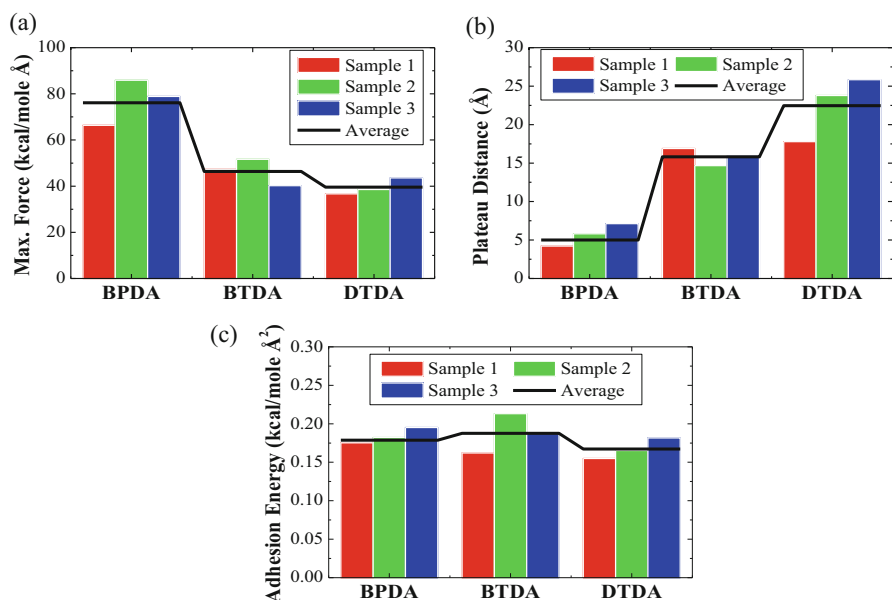
Fig. 10 (**a**) Maximum force, (**b**) plateau distance, and (**c**) adhesion energy for all structures of the polyimide on SiO2. The black lines represent the average values

200 m/s to 1 m/s, as shown in Fig. 10b. At velocities slower than 50 m/s (i.e., 25 m/s and 1 m/s), PMF responses are observed to be independent of SMD rates.

Consequently, a velocity of 50 m/s is chosen as an optimal value and applied to the rest of simulations. Calculated adhesion properties for different types of PIs are presented in Fig. 10c. For better statistical results, pulling tests are repeated for three samples of each PI structure. First, it clearly indicates that results from the same type of the PI exhibit consistency for all adhesion properties (i.e., no outliers appear). More importantly, each adhesion property shows a distinctive trend as the CTE value is lowered (the order of CTE values: BPDA < BTDA < DTDA). For example, the largest value of FMax is required to detach BPDA from the SiO2 surface, while the shortest value of required DP for BPDA. On the other hand, the DTDA structure displays the smallest value of FMax but the largest pulling distance when two substrates are completely detached. To summarize, a more rigid PI needs a larger force for detachment (FMax: BPDA > BTDA > DTDA), while a less rigid PI requires a longer pulling distance (Dp: DTDA > BTDA > BPDA). It means that the PI with a lower value of CTE requires a larger force to detach from SiO2, but detachment occurs at the shorter distance. Since the structure with a larger CTE value has the weaker interaction between chains based on the nature of the CT complex, it is reasonable of DTDA exhibiting a larger pulling distance (Choi et al. 2010; Chaudhury et al. 1987). The larger pulling force shown in BPDA can be explained from investigation on Coulomb energy changes during detachment, which is covered in Sect. 3.3. However, in terms of the adhesion energy, it is interesting to note that its value is almost equivalent among all structures.

For further analysis of the temperature effect, adhesion properties at 450 K and 550 K for BPDA and DTDA structures are obtained (Table 2). It shows that even at higher temperatures, the same adhesion behavior, i.e., a shorter plateau distance and a larger maximum force in BPDA, and comparable adhesion energy for both structures, is obtained comparing to the case at 300 K. It is also important to mention that the adhesion energy at higher temperatures is lower than that at 300 K, which is associated with a smaller force required to detach PIs. This could be because the atomic thermal vibration at the interface is amplified at the higher temperature. Therefore, an important perspective can be obtained from the above results that although the same work of adhesion (PMF) is required for full detachment of any structures considered here, the underlying mechanism is different depending on structural characteristics; thus, the further analysis performed provides detailed insight on the nature of adhesion.

3.2.4 Structural Changes During Pulling Process

During the pulling process, it is highly probable that polymer chains located closer to the interfacial region would play an important role in determining adhesion properties. Hence, understanding the underlying mechanism of adhesion between PIs and SiO2 glass at the atomistic level is of importance because the inter- and intra-material interaction inside bi-materials governs the adhesion failure mode at the interface. Accordingly, it is critical to perform an analysis for the response of

each chain in detail to elucidate relevant physics and dynamics. For comparison, the most thermally stable/high rigidity (BPDA) and the least thermally stable/least rigid (DTDA) PIs are chosen.

First, an overall outlook of the detachment process can be provided from atomic snapshots. For a case of BPDA, geometric configurations in polymer chains at the interface are not strongly affected during detachment. On the other hand, for DTDA structure, polymer chains located nearer to the SiO2 surface are stretched in the direction of pulling and experience significant modification in the chain conformation. A detailed insight into the variation of chain conformation is examined from computing the variation of Rg and REE during pulling process as shown in Fig. 11. This can provide direct information how an individual chain in PIs is deformed during detachment. First, initial Rg and REE values and initial conformation from each chain in BPDA and DTDA are presented in Fig. 10. It should be mentioned that for all chains, averaged values of initial Rg and REE obtained from DTDA are smaller than those from BPDA. This result indicates that DTDA is more deformable/flexible, which is consistent with results from the bulk system of PIs. In addition, the standard deviation of Rg (2.72 Å) and REE (12.03 Å) from DTDA is more than three times larger than those from BPDA (it is 0.87 Å and 3.00 Å for Rg and REE, respectively), which could be another indication of the higher chain flexibility in DTDA. Further analysis can be found in our publication (Kyoungmin et al. 2016).
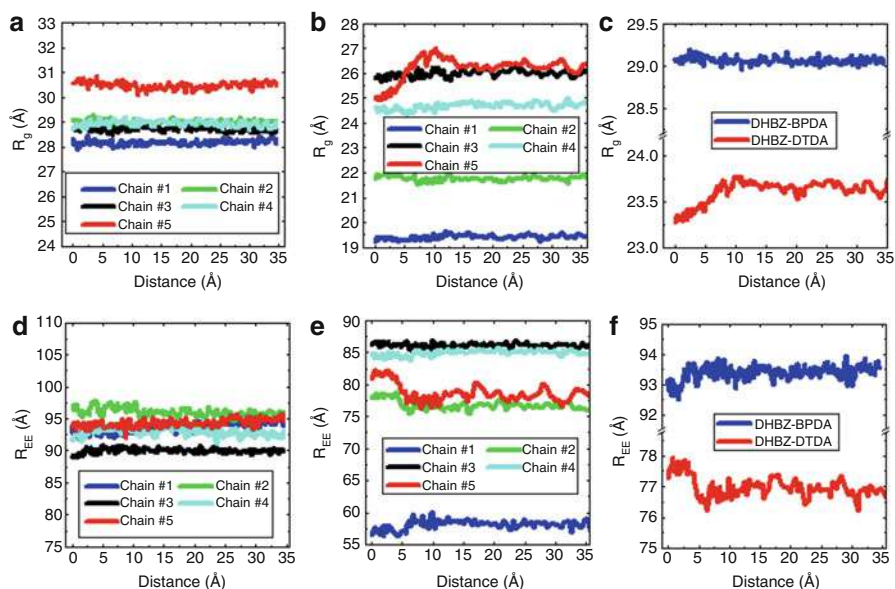
Fig. 11 Radius of gyration (Rg) and chain end-to-end distance (REE) variation in individual polyimide chains during pulling process for the structure of (**a**), (**d**) DHBZ-BPDA and (**b**), (**e**) DHBZ-DTDA on SiO2. Averaged values for (**c**) Rg and (**f**) REE are plotted. The number of chain is increased when the chain is located closer to the SiO2 surface

3.2.5   Failure Mode

Finally, the failure mode at the bi-material interface is investigated to determine whether it exhibits adhesive or cohesive failure over the course of detachment. In general, adhesive failure occurs when two materials are completely detached without internal failure in either material; hence, a "clean" surface at the interface is formed. In contrast, cohesive failure occurs when a polymer experiences internal chain breaking at failure due to the overly strong adhesion at the interface. In order to examine the failure mode, we modified the region where the pulling force is applied, i.e., the force is applied to only the half region from the top of PI in z-direction (denoted case H for "half"), whereas for all previously discussed systems, the force was applied to all atoms in PI (denoted case A for "all"). For case H, PI chains near the surface are being pulled by the other PI chains at the top. Therefore, it is possible that PI chains would experience cohesive failure if their inter-chain interactions are weak or their adhesion to the SiO_2 surface is too strong. However, for the current study, the cohesive failure is not observed for either BPDA or DTDA structures; a clean interface is always formed when PIs are fully detached. It should be noted that all PIs in this study possess very low CTE values compared to other types of polymers. Considering the fact that the charge transfer complex between PI chains is very strong, there are strong attractive forces between chains, preventing cohesive failure (Choi et al. 2010; Chaudhury et al. 1987).

3.3   Effect of Glass Roughness on Adhesion

Creating flat surfaces for molecular modeling is perhaps trivial. For cleaved crystalline surfaces with small unit cells, the manual operation of adding hydrogen to the surface oxides can be done without much difficulty (Mishra et al. 2013). In most studies flat surfaces are primarily used, there are several advantages of flat surfaces: They lend themselves well to the use of density profiles, as most species form orderly structures as a function of distance from the surface, and the explanation that surface roughness features tend to be perceived at wavelengths longer than the simulation length. When considering a glass or other amorphous structure, even the flattest surface has heterogeneities that cause a certain degree of atomic level roughness. With amorphous structures and nanoscale roughness, it becomes less obvious how to create or manage the surface. In some of these cases, the surface configuration with minimum energy can be used, some can be found with temperature gradient protocols (Mishra et al. 2013, 2014). But without a reacting force field, such as reaxFF (Manzano et al. 2012), these approaches are of limited use. The underlying chemical nature of the structure must be understood to properly select for minimum energy surfaces.

3.3.1   Method for Creating Surface Roughness on Glass

To create roughness on the surface, we cut a bulk glassy silica with a predefined surface. As shown in Fig. 12, we firstly define a shape of the surface by a specific mathematical function and then cut an initial bulk structure along the surface.
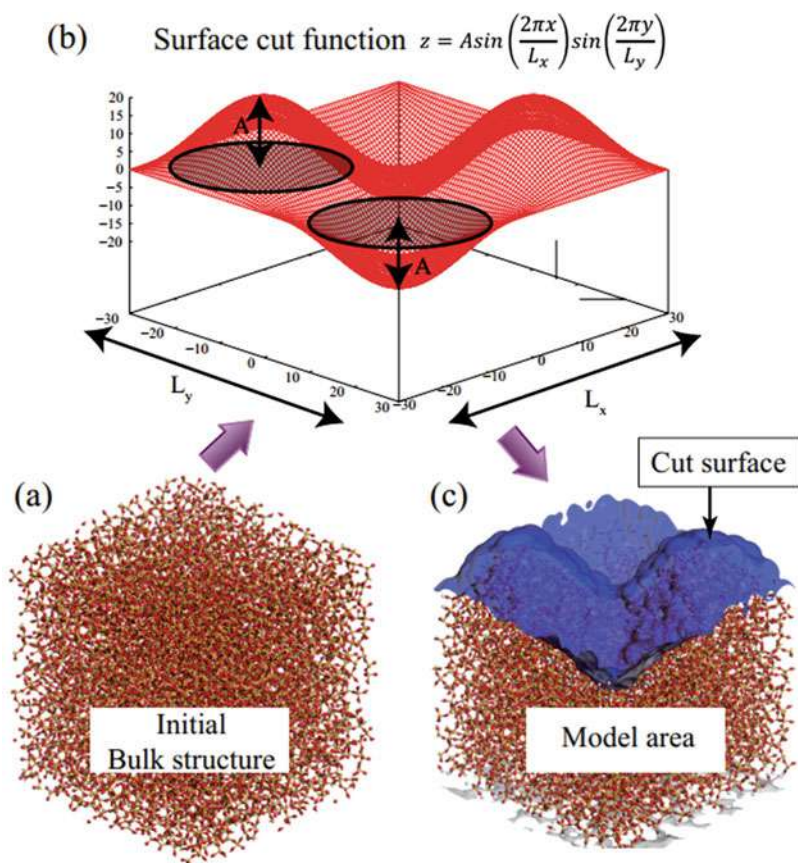
Fig. 12 Schematic procedure of generating rough surface. (**a**) Initial bulk glassy silica, (**b**) predefined surface with sine function, $z = A\sin\left(\frac{2\pi x}{L_x}\right)\sin\left(\frac{2\pi y}{L_y}\right)$, (**c**) prepared rough glassy silica surface, blue shaded area represents cut surface by sine function

Roughness can be adjusted depending on the defined surface function. Since nano-sized roughness is formed by varying amplitudes and spacing (Chen et al. 2007), we applied a sinusoidal function to adjust both parameters as follows,

$$z = A\sin\left(\frac{2\pi x}{L_x}\right)\sin\left(\frac{2\pi y}{L_y}\right) + \bar{z} \tag{7}$$

where A, Lx, and Ly represent roughness amplitude and spacing of each axis, respectively. Various amplitudes (A= 0, 5, 10, 15, 20 Å) and spacings (Lx = Ly = 20, 30, 60, 120 Å) are considered to analyze the effect of each parameter in this work. Experimentally, the amplitude parameter is usually represented as the height variation of the surface compared to a reference plane, and arithmetic average, Ra, is the most common parameter to represent the amplitude of the roughness. Thus, Ra

is measured for the prepared structure and listed as an amplitude parameter instead, A in this work. Ra is calculated using atomic coordination of surface atoms with the following equation,

$$R_a = \frac{1}{n} \sum_{i=1}^{n} |z_i - \bar{z}| \tag{8}$$

where z_i is the z coordinate of each element and $\bar{z}$ is average value. Due to the symmetric shape of a sine function, Ra is almost half of the amplitude value (A). As the amplitude values increase from 0 to 5, 10, 15, and 20 Å, Ra varies from 1.0 Å to 2.4, 4.4, 6.5, and 8.4 Å, respectively. The spacing of roughness corresponds to the period of a sine wave (Lx or Ly in the function) and hence, period is listed as the spacing parameter of the roughness. More specific details on this approach can be found in literature.

3.3.2   Results

For the systematic analysis of roughness, both amplitude and spacing parameters of roughness are considered with various hydroxylation densities. Figure 13a presents the adhesion energy variation with different Ra and hydroxylation densities. As the Ra or hydroxylation density increases, adhesion energies increase. When hydroxylation density is increased from 2.4 OH/nm2 to 4.8 OH/nm2, the number of hydroxyl groups becomes two times larger, i.e., 100% increase, but increased adhesion energies by 10 ~ 20% for any given Ra value. On the other hand, when R_a is changed from 1.0 Å to 8.5 Å with the same hydroxylation density of 4.8 OH/nm2, adhesion energy is increased about 89%. Adhesion energy continuously increases with Ra higher than 8.5 Å, and thus we can conclude that the amplitude of roughness, Ra, is more effective to enhance adhesion between glassy silica and polymer than hydroxylation density in nano-scale roughness.
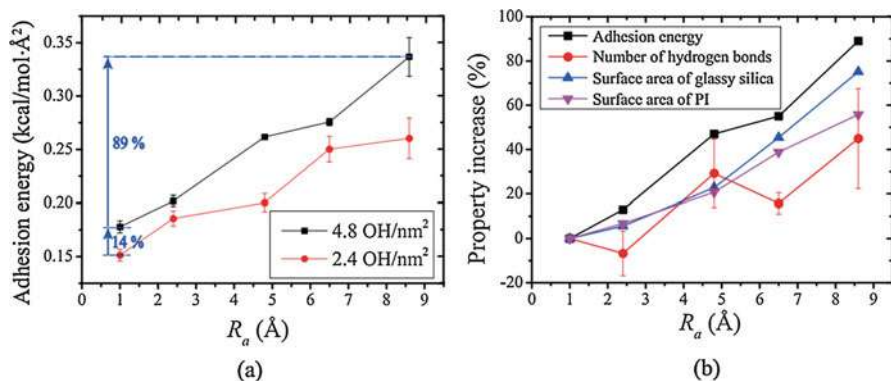
Fig. 13 (**a**) Variation of adhesion energy as a function of Ra (1.0 ~ 8.4 Å) and hydroxylation density (2.4 and 4.8 OH/nm2), (**b**) increment (%) of adhesion energy, number of hydrogen bonds, surface area of glassy silica, and PI

For the analysis of the hydrogen bond contribution to the adhesion, we counted hydrogen bonds at the interface (exhibited as red circles in Fig. 13b). Hydrogen bonds gradually increase with Ra due to the expansion of the surface area, but the amount of increased adhesion energy is always higher than that of increased hydrogen bonds. When the adhesion energy is enhanced about 89%, the number of hydrogen bonds increases only 45%. Thus, the effect of Ra on adhesion is the sum of the increased hydrogen bonds and additional nonbonded interactions by surface expansion. Adhesion energy variation depending on Ra follows two steps. As shown in the figure, both glassy silica and PI surface areas (blue and purple triangles) match well with each other when Ra is lower than 4.4 Å and hence, the effect of roughness might be maximized within this range. However, when Ra is larger than 4.4 Å, the two surfaces begin to mismatch due to the high amplitude of roughness. PI is not completely attached on the glassy silica surface, and thus the effect of roughness on the adhesion declines and adhesion energy saturates at a higher Ra value.

Figure 14 represents the calculated adhesion energy variation on Ra and period. Considering the slope change of the adhesion, it is concluded that period is less effective on the adhesion than Ra within the considered conditions. Adhesion energy can be enhanced by expanding surface area with a reduced period, but maximum adhesion is observed with the longest period (120 Å) in this work. Thus, for an increment of surface area with a reduced period may not be directly related with the adhesion enhancement within the considered period range. For the detailed analysis of reducing adhesion with decreasing period, vacant volume at the interface is measured and presented in Fig. 14b. When the period is 120 Å, the vacant volume is negligible. However, as the period decreases, there exists a distinctively increasing amount of vacant pores at the interface, and due to the vacant pores, contact area between the two surfaces is lowered and hence, adhesion energy is decreased.

Figure 15 shows a schematic trend of adhesion energy between glassy silica and PI depending on period. The adhesion energy trend with period is affected by both surface area and vacant volume at the interface. As the period decreases, surface area expands, and thus adhesion energy increases (blue solid line). However, if
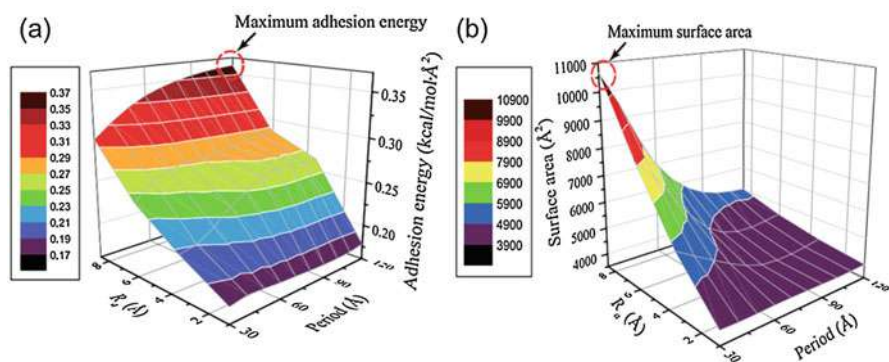
Fig. 14 Variation of (**a**) adhesion energy and (**b**) surface area as a function of Ra and period of roughness
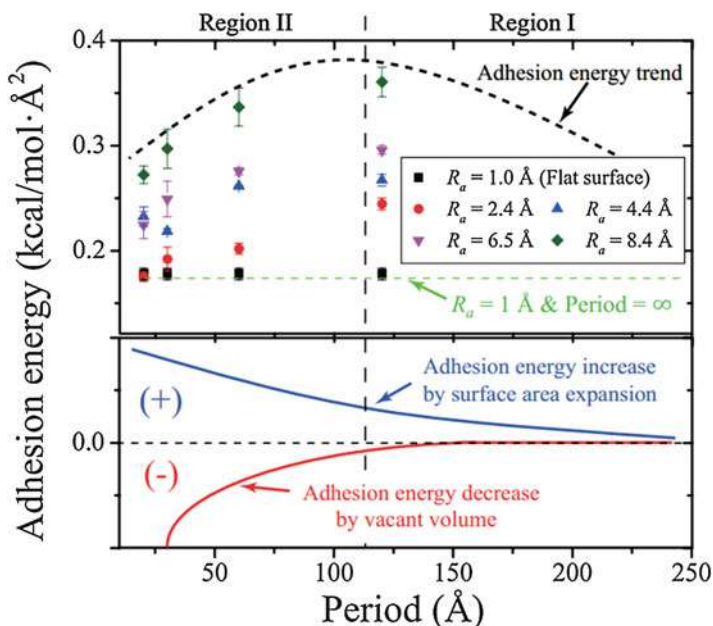
Fig. 15 Schematic trend of adhesion energy as a function of roughness period. Black dotted line represents variation of adhesion energy, and blue and red lines represent variation of adhesion energy due to surface area or vacant volume, respectively. Green dotted line represents adhesion energy for flat surface (Ra = 1 Å with infinite period). Adhesion energy increases at Region I and decreases at Region II

vacant volume exists at the interface, adhesion energy decreases (red solid line). The overall adhesion energy reflects both effects, and hence one can predict the trend of adhesion energy (black solid line) in two separate regions as follows. In region I, adhesion energy increases with decreasing period which implies the surface area expansion is dominant, and adhesion energy decreases in region II which indicates that vacant volume is critical for these periods. At the border of the two regions, there is a specific period which provides the maximum adhesion energy of a system. A black dotted line in Fig. 15 presents the adhesion energy trend of this work, and maximum adhesion energy is observed with a period of 120 Å. The existence of vacant volume varies depending on Ra and flexibility of polymer, and hence maximum adhesion energy can change with these conditions. If we introduce higher Ra than 8.5 Å or a more rigid polymer than the current PI, the period for maximum adhesion will be longer than 120 Å.

3.4   Conclusion

In this book chapter, we have presented a multiscale modeling approach which established methods, and some guiding principles for modulating adhesion between

organic. Electronic structure calculations revealed the importance of hydrogen bonding and of sodium cation in increasing adhesion. Furthermore, the organic subunit rigid can enhance the adhesion by decreasing the conformational degree of freedom.

On investigating monomer glass adhesion, we find that hydroxylated silica surfaces have higher adhesion than nonhydroxylated surfaces. We also find that glassy silica has a higher adhesion than crystalline silica. The reason for this appears to be the additional hydrogen bonding interactions possible in the presence of surface hydroxylation and due to a rougher surface for silica glass allowing more interactions. On decoupling the interaction forces, we find that adhesion with silica for polyimides with higher oxygen density would be higher.

Scaling up to polymer interaction on glass, we find that PI materials with lower CTEs require a larger maximum force to detach them from the glass substrate. In addition, the pulling distance for complete detachment is shorter for the lower CTE polymers. A detailed analysis for the chain conformation change confirms that during the pulling process, the polymer chains near the surface region mainly determine the interfacial adhesion behavior. When pulled, the conformation of BPDA chains at the interface is not significantly affected, while those of DTDA experience substantial changes. This is the cause of a larger pulling distance required for complete detachment for DTDA compared to BPDA. Furthermore, the variation in decomposed energy terms during detachment indicates that the main factors determining the adhesion behavior are bond and coulomb interactions. Finally, the current systems under study are always expected to experience adhesive rather than cohesive failure when detached.

Finally, upon investigating the role of nanoscale roughness on glassy silica, and the calculated results reveal that R_a exhibits the highest effect on the adhesion change among the parameters considered. The energetic contribution to the adhesion is analyzed with energy decomposition during the pulling process, and it is revealed that not only the increase of hydrogen bond but also the increase of Coulomb energy contributes to the adhesion energy variation by R_a. The contribution of Coulomb energy becomes more significant when R_a is higher. From the structural analysis, the trend of adhesion energy variation depending on R_a and period is predicted. R_a can enhance adhesion energy, but the impact of R_a on the adhesion is also determined by the period. When the period is the shortest, normalized adhesion energy by PI surface is similar for any value of R_a.

4   Potential Vistas in Interface Modeling

We have so far discussed adhesion of physisorption of nanometer thick polymer films on glasses, which are bound by columbic and van der Waals interactions. There is a pertinent need to extend this study to be able to incorporate reactions at the interface and to perform simulations that scale larger length scales. To achieve this, there is a need to develop coarse grained models for interfaces which capture accurate polymer and glass physics, along with chemistry of interface evolution.

Recently developed has focused on incorporate bonding at the interface in classical MD and evaluating the potential of coarse-grained simulations (Bulacu 2008; Yang and Qu 2014; Xia et al. 2014) to traverse our understanding to larger length and timescales.

References

Adamo C, Barone V (1999) Toward reliable density functional methods without adjustable parameters: the PBE0 model. J Chem Phys 110:6158

Andersson MP (2013) Density functional theory with modified dispersion correction for metals applied to self-assembled monolayers of thiols on Au(111). J Theor Chem, 2013:327839

Aoyagi T (2015) Computational modeling of polymer with hydrogen bonds. Polymer 64(2):91–93

Awaja F, Gilbert M, Kelly G, Fox B, Pigram PJ (2009) Adhesion of polymers. Prog Polym Sci 34(9):948–968

Beake BD, Zheng S, Alexander M (2002) Nanoindentation testing of plasma-polymerised hexane films. J Mater Sci 37:3821–3826

Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98(7):5648

Beest BWH, Kramer GJ, Santen RA (1990) Force fields for silicas and aluminophosphates based on ab-initio calculations. Phys Rev Lett 64:1955–1958

Berardo E, Pedone A, Ugliengo P, Corno M (2013) DFT modeling of 45S5 and 77S soda-lime phospho-silicate glass. Langmuir 29(19):5749–5759

Blochl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953

Bouzid A, Le Roux S, Boero M, Massobrio C (2015) Origin of structural analogies and differences between the atomic structures of GeSe$_4$ and GeS$_4$ glasses: a first principles study. J Chem Phys 143:034504

Brown HR (2000) Polymer adhesion. Mater Forum 24:49–58

Bulacu MI (2008) Molecular dynamics studies of entangled polymer chains. PhD thesis, University of Groningen, Groningen

Büyüköztürk O, Buehler MJ, Lau D, Tuakta C (2011) Structural solution using molecular dynamics: Fundamentals and a case study of epoxy-silica interface. Int J Solids Struct 48(14–15):2131–2140

Chabria M, Hertig S, Smith ML, Vogel V (2010) Stretching fibronectin fibres disrupts binding of bacterial adhesins by physically destroying an epitope. Nat Commun 1:135

Chaudhury MK, Gentle TM, Plueddemann EP (1987) Adhesion mechanism of polyvinyl chloride to silane primed metal surfaces. J Adhesion Sci Tech 1(1):29–38

Chen TA, Jen AKY, Cai Y (1995) Facile approach to nonlinear optical side-chain aromatic polyimides with large second-order nonlinearity and thermal stability. J Am Chem Soc 117(27):7295e7296

Chen MA, Li HZ, Zhang XM (2007) Improvement of shear strength of aluminium-polypropylene lap joints by grafting maleic anhydride onto polypropylene. Int J Adhes Adhes 27:175–187

Cheng IC, Wagner S (2009) Overview of flexible electronics technology. In: Wong WS, Salleo A (eds) Flexible electronics: materials and application. Springer, New York, pp 1–28

Cheng SZD, Wu Z, Mark E, Steven LCH, Frank WH (1991) A high-performance aromatic polyimide fibre: structure, properties and mechanical-history dependence. Polym (Guildf) 32(10):1803e1810

Chenoweth K, Duin ACT, Goddard WA III (2008) ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. J Phys Chem A 112:1040–1053

Choi MC, Hwang JC, Kim C, Ando S, Ha CS (2010) New colorless substrates based on polynorbornene-chlorinated polyimide copolymers and their application for flexible displays. J Polym Sci Part A Polym Chem 48(8):1806–1814

Choi M, Kim Y, Ha CS (2008) Polymers for flexible displays: from material selection to device applications. Prog Polym Sci 33(6):581–630

Cui C, Han J, Zhu X, Liu X, Wang H, Mei D, Ge Q (2016) Promotional effect of surface hydroxyls on electrochemical reduction of CO_2 over SnO_x/Sn electrode. J Catal 343:257–265. https://doi.org/10.1016/j.jcat.2015.12.001

Dai CA, Liao CC, Tsui TA, Chien HC, Liu MW (2006) Adhesion measurement of interfaces between gelatin and poly(ethylene terephthalate) using microscratch technique. J Appl Polym Sci 99:1960–1974

Darwent B deB (1970) Bond dissociation energies in simple molecules NSRDS-NBS 31. US National Bureau of Standards, Washington, DC

Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT (1988) Structure and energetics of ligand binding to proteins: E. Coli dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins 4:31–47

Davis DA, Hamilton A, Yang J, Cremar LD, Van Gough D, Potisek SL, Ong MT, Braun PV, Martinez TJ, White SR, Moore JS, Sottos NR (2009) Force-induced activation of covalent bonds in mechanoresponsive polymeric materials. Nature 459(7243):68–72

Deb Nath SK (2013) Study of the effect of sizes on the structural properties of SiO_2 glass by molecular dynamics simulations. J Non-Cryst Solids 376:50–60

Du J, Kroll P (2010) Electronic structure and interfacial properties of Ge nanoclusters embedded in amorphous silica. J Non-Cryst Solids 356(44–49):2448–2453

Duin ACT, Dasgupta S, Lorant F, Goddard WA III (2001) ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A 105:9396–9409

Dvir H, Gottlieb M (2009) Effect of silane sizing on polymer-glass adhesion. http://www.iccm-central.org/Proceedings/ICCM17proceedings/Themes/Materials/INTERPHASES%20&%20INTERFACES/D7.15%20Gottlieb.pdf. Accessed 22 Feb 2017

Emami FS, Puddu V, Berry RJ, Varshney V, Patwardhan SV, Perry CC, Heinz H (2014a) Force field and a surface model database for silica to simulate interfacial properties in atomic resolution. Chem Mater 26:2647–2658

Emami FS, Puddu V, Berry JR, Varshney V, Patwardhan VS, Perry CC, Heinz H (2014b) Prediction of specific biomolecule adsorption on silica surfaces as a function of pH and particle size. Chem Mater 26:5725–5734

Escott R, Harris FW, Hergenrother PM, Makino D, Satou H, Sroog CE, St. Clair TL, Stenzenberger H, Suzuki H, Takekoshi T, Wilson D, Young PR, Wilson D, Stenzenberger HD, Hergenrother PM (eds) (1990) Springer Netherlands

Favre JP, Desarmot G, Sudre O, Vassel A (1997) Were McGarry or Shiriajeva right to measure glass fibre adhesion? Compos Interfaces 4:313–326

Ferreira L, Evangelista MB, Martins MCL, Granja PL, Esteves JL, Barbosa MA (2005) Improving the adhesion of poly (ethylene terephthalate) fibers to poly (hydroxyethyl methacrylate) hydrogels by ozone treatment: surface characterization and pull-out tests. Polymer 46:9840–9850

Friedrich J, Mix R, Kuhn G (2005) Adhesion of metals to plasma-induced functional groups at polymer surfaces. Surf Coat Technol 200:565–568

Fryd M, Mittal KL (1984) Polyimides: synthesis, characterization, and applications, vol 1. Springer US, Boston, pp 377–383

Geim AK et al. (2003) Microfabricated adhesive mimicking gecko foot-hair. Nat Mater 2:461–463

Gellert EP, Turley DM (1999) Seawater immersion ageing of glass-fibre reinforced polymer laminates for marine applications. Compos A 30(11):1259–1265

Ghosh MK, Mittal KL (1996) Polyimides: fundamentals and applications. Marcel Dekker, Inc., New York. Technology & Engineering

Gleizer A, Peralta G, Kermode JR, De Vita A, Sherman D (2014) Dissociative chemisorption of O_2 inducing stress corrosion cracking in silicon crystals. Phys Rev Lett 2014(112):115501

Goyal S, Park H, Lee SH ea (2016) Characterizing the fundamental adhesion of polyimide monomers on crystalline and glassy silica surfaces: a molecular dynamics study. J Phys Chem C 41(120):23631–23639

Grady ME, Geubelle PH, Sottos NR (2014) Interfacial adhesion of photodefinable polyimide films on passivated silicon. Thin Solid Films 552:116–123

Grimme S (2006a) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comp Chem 27:1787–1799

Grimme S (2006b) Semiempirical hybrid density functional with perturbative second-order correlation. J Chem Phys 124:034108

Gruenhut S, Amini M, Macfarlane DR, Meakin P (1997) Molecular dynamics glass simulation and equilibration techniques. Mol Simul 19(3):139–160

Hamdi M, Ferreira A, Sharma G, Mavroidis C (2008) Prototyping bio-nanorobots using molecular dynamics simulation and virtual reality. Microelectron J 39(2):190–201

Hampe A, Marotzke C (1997) The energy release rate of the fiber/polymer matrix interface: measurement. J Reinf Plast Compos 16:341–352

Heinz H, Lin TJ, Mishra KR, Emami FS (2013) Thermodynamically consistent force fields for the assembly of inorganic, organic, and biological nanostructures: the INTERFACE force field. Langmuir 29:1754–1765

Holbery J, Houston D (2006) Natural-fiber-reinforced polymer composites in automotive applications. JOM 58(11):80–86

Isralewitz B, Gao M, Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol 11(2):224–230

Izvekov S, Rice BM (2012) Mechanism of densification in silica glass under pressure as revealed by a bottom-up pairwise effective interaction model. J Chem Phys 136:134508

Jabbarzadeh A, Tanner RI (2006) Molecular dynamics simulation and its application to nanorheology. Rheol Rev:165–216

Jarzynski C (1997) Nonequilibrium equality for free energy differences. Phys Rev Lett 78(14):2690–2693

Juhl KM, Bovet N, Hassenkam T, Dideriksen K, Pedersen CS, Jensen CM, Okhrimenko DV, Stipp SL (2014) Change in organic molecule adhesion on α alumina (sapphire) with change in NaCl and CaCl₂ solution salinity. Langmuir 30:8741–8750

Jung IS, Lee YJ, Jeong D, Graf R, Choi TL, Son WJ, Bulliard X, Spiess HW (2016) Conformational analysis of oxygen-induced higher ordered structure of A, B-alternating poly (arylene vinylene) copolymers by solid-state NMR and molecular dynamics simulations. Macromolecules 49(8):3061–3069

Kang J-H, Kim K-S, Kim K-W (2010) Molecular dynamics study on the effects of stamp shape, adhesive energy, and temperature on the nanoimprint lithography process. Appl Surf Sci 257(5):1562–1572

Klee D, Hocker H (1999) Polymers for biomedical applications: improvement of the interface compatibility. In: Eastmond GC, Hocker H, Klee D (eds) Biomedical application polymers blends, vol 149, pp 1–57

Komarov PV, Mikhailov IV, Chiu YT, Chen SM, Khalatur PG (2013) Molecular dynamics study of interface structure in composites comprising surface-modified SiO₂ nanoparticles and a polyimide matrix. Macromol Theory Simul 22:187–197

Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. Phys Rev B 47:558

Kyoungmin M, Kim Y, Goyal S et al (2016) Interfacial adhesion behavior of polyimides on silica glass: a molecular dynamics study. Polymer 98:1–10

Lacombe R (2006) Adhesion measurement methods theory and practice. CRC/Taylor & Francis, Boca Raton

Lahmer MA (2015) Hydrogen sensing properties of the ZnO(0001) surface enhanced by Be doping: a first principles study. Sensors Actuators B Chem 221:906–913

Larsson MP, Ahmad MM (2006) Improved polymer–glass adhesion through micro-mechanical interlocking. J Micromech Microeng 16:S161–S168

Lee MB, Kuranaga T, Munetoh S, Motooka T (2007) Surface nucleation of the (111) plane of excimer Laser annealed Si on SiO₂ substrates: a molecular dynamics study. J Appl Phys 101:054316

Lee S, Park J, Yang J, Lu W (2014) Molecular dynamics simulations of the traction separation response at the interface between PVDF binder and graphite in the electrode of Li-ion batteries. J Electrochem Soc 161(9):A1218–A1223

Liaw DJ, Chang FC, Leung M, Chou MY, Muellen K (2005) High thermal stability and rigid rod of novel organosoluble polyimides and polyamides based on bulky and noncoplanar naphthalene biphenylydiamine. Macromolecules 38(9):4024e4029

Liaw DJ, Wang KL, Huang YC, Lee KR, Lai JY, Ha CS (2012) Advanced polyimide materials: syntheses, physical properties and applications. Prog Polym Sci 37(7):907–974

Lopez-Chavez E, Pena-Castaneda YA, de la Portilla-Maldonado LC, Guzman-Pantoja J, Martinez-Magadan JM, Oviedo-Roa R, de Landa C-AF, Cruz-Torres A (2014) Role of sulfonation in the stability, reactivity, and selectivity. J Mol Model 20:2325

Luo J, Vargheese KD, Tandia A, Hu G, Mauro JC (2016) Crack nucleation criterion and its application to impact indentation in glasses. Sci Rep 6:23720

Mantisi B, Bauchy M, Micoulaut M (2015) Cycling through the glass transition: evidence for reversibility windows and dynamic anomalies. Phys Rev B 92(13):134201/1–134201/11

Manzano H, Pellenq RJ, Ulm FJ, Buehler MJ, van Duin AC (2012) Hydration of calcium oxide surface predicted by reactive force field molecular dynamics. Langmuir 28(9):4187–4197

Marotzke A, Hampe C (1997) The energy release rate of the fiber/polymer matrix interface: measurement. J Reinf Plast Compos 16:341–352

Mashio T, Ohma A, Tokumasu T (2016) Molecular dynamics study of ionomer adsorption at a carbon surface in catalyst ink. Electrochim Acta 202:14–23

Mathieson I, Bradley R (1996) Improved adhesion to polymers by UV/ozone surface oxidation. Int J Adhes Adhes 16:16–29

McMullen WE, Freed KF (1990) A density functional theory of polymer phase transitions and interfaces. J Chem Phys 92(2):1413–1426

Mishra RK, Fernández-Carrasco L, Flatt RJ, Heinz H (2014) A force field for tricalcium aluminate to characterize surface properties, initial hydration, and organically modified interfaces in atomic resolution. Dalton Trans 43(27):10602–10616

Mishra RK, Flatt RJ, Heinz H (2013) Force field for tricalcium silicate and insight into nanoscale properties: cleavage, initial hydration, and adsorption of organic molecules. J Phys Chem C 117(20):10417–10432

Miwa T, Tawata R, Numata S (1993) Relationship between structure and adhesion properties of aromatic polyimides. Polymer 34(3):621–624

Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46:618–622

Mueller E, Duin ACT, Goddard WA III (2010) Development and validation of ReaxFF reactive force field for hydrocarbon chemistry catalyzed by nickel. J Phys Chem B 114:4939–4949

Muhich CL, Qui J, Holder AM, Wu YC, Weimer AW, Wei WD, McElwee-White L, Musgrave CB (2015) Solvent control of surface plasmon-mediated chemical deposition of Au nanoparticles from alkylgold phosphine complexes. ACS Appl Mater Interfaces 7(24):13384–13394

Ni H, Liu J, Wang Z, Yang S (2015) A review on colorless and optically transparent polyimide films: chemistry, process and engineering applications. J Ind Eng Chem 28:16–27

Odegard G, Clancy T, Gates T (2005) Modeling of the mechanical properties of nanoparticle/polymer composites. Polymer 46:553–562

Ohkubo T, Tsuchida E, Takahashi T, Iwadate Y (2016) Ab initio molecular dynamics simulations and GIPAW NMR calculations of a lithium borate glass melt. J Phys Chem B 120(14):3582–3590

Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Chem Phys 119(6):3559–3566

Park S, Schulten K (2004) Calculating potentials of mean force from steered molecular dynamics simulations. J Chem Phys 120(13):5946–5961

Patwardhan VS, Emami FS, Berry RJ, Jones SE, Naik RR, Deschaume O, Heinz H, Perry CC (2012) Chemistry of aqueous silica nanoparticle surfaces and the mechanism of selective peptide adsorption. JACS 134:1–14

Pedone A, Malavasi G, Menziani MC, Cormack AN, Segre U (2006) A new self-consistent empirical interatomic potential model for oxides, silicates, and silica-based glasses. J Phys Chem B 110:11780–11795

Pedone A, Malavasi G, Menziani MC, Segre U, Cormack AN (2008) Molecular dynamics studies of stress−strain behavior of silica glass under a tensile load. Chem Mater 20:4356–4366

Perdew JP, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B 45:13244–13249

Peterson AM, Jensen RE, Palmese GR (2011) Thermoreversible and remendable glass–polymer interface for fiber-reinforced composites. Compos Sci Technol 71(5):586–592

Pham TA, Lee D, Schwegler E, Galli G (2014) Interfacial effects on the band edges of functionalized Si surfaces in liquid water. J Am Chem Soc 136(49):17071–17077

Piggot M (1997) Why interface testing by single-fiber methods can be misleading. Compos Sci Technol 57:965–974

Pizzi A, Mittal KL (eds) (2003) Handbook of adhesive technology. Marcel Dekker, New York

Priezjev N (2016) Reversible plastic events during oscillatory deformation of amorphous solids. Phys Rev E 93:013001

Ren J, Zhou GF, Guo ZC, Zhang W (2009) Density functional theory study on the surface reaction mechanism of atomic layer deposited Ta_2O_5 on Si(100) surfaces. Chem J Chin Univ 30(11):2279–2283

Sablic J, Praprotnik M, Delgado-Buscalioni R (2016) Open boundary molecular dynamics of sheared star-polymer melts. Soft Matter 12(8):2416–2439

Shoravi S, Olsson GD, Karlsson BC, Bexhorn F, Abghoui Y, Hussain J, Wiklander JG, Nicholls IA (2016) In silico screening of molecular imprinting prepolymerization systems: oseltamivir selective polymers through full-system molecular dynamics-based studies. Org Biomol Chem 14(18):4210–4219

Sirk TW, Karim M, Lenhart JL, Andzelm JW, Khare R (2016) Bi-modal polymer networks: viscoelasticity and mechanics from molecular dynamics simulation. Polymer 90:178–186

Sliozberg YR, Kroger M, Chantawansri TL (2016) Fast equilibration protocol for million atom systems of highly entangled linear polyethylene chains. J Chem Phys 144(15):154901/1–154901/12

Sun H (1998) COMPASS: an ab initio force-field optimized for condensed-phase applications – overview with details on alkane and benzene compounds. J Phys Chem B 102:7338–7364

Sun H, Mumby SJ, Maple JR, Hagler AT (1994) An Ab initio CFF93 all-atom force field for polycarbonates. J Am Chem Soc 116:2978–2987

Tam L, Lau D (2015) Moisture effect on the mechanical and interfacial properties of epoxy-bonded material system: an atomistic and experimental investigation. Polymer 57:132–142

Tandia A, Vargheese D, Mauro JC, Varshneya AK (2012) Atomistic understanding of the network dilation anomaly in ion-exchanged glass. J Non-Cryst Solids 358:316–320

Tonigold K, Grob A (2010) Adsorption of small aromatic molecules on the (111) surfaces of noble metals: a density functional theory study with semiempirical corrections for dispersion effects. J Chem Phys 132:224701

Whittingham MS (2014) Ultimate limits to intercalation reactions for lithium batteries. Chem Rev 114(23):11414–11443

Wippermann S, Voros M, Gali A, Gygi F, Zimanyi GT, Galli G (2014) Solar nanocomposites with complementary charge extraction pathways. Phys Rev Lett 112:106801

Wong WS, Salleo A (eds) (2009) Flexible electronics: materials and applications. Springer, New York

Xia W, Hsu DD, Keten S (2014) Dependence of polymer thin film adhesion energy on cohesive interactions between chains. Macromolecules 47(15):5286–5294

Yang S, Choi J, Cho M (2012) Elastic stiffness and filler size effect of covalently grafted nanosilica polyimide composites: molecular dynamics study. Appl Mater Interfaces 4:4792–4799

Yang S, Gao F, Qu S (2013) A molecular dynamics study of tensile strength between a highlycrosslinked epoxy molding compound and a copper substrate. Polymer 54(18):5064–5074

Yang S, Qu J (2014) An investigation of the tensile deformation and failure of an epoxy/Cu interface using coarse-grained molecular dynamics simulations. Model Simul Mater Sci Eng 22(6):65011

Yang Z, Wang B, Cormack AN (2016) The local structure of Fe in Li (Al, Fe)Si_2O_6 glasses from molecular dynamics simulations. J Non-Cryst Solids 444:16–22

Yu X, Yi B, Wang X (2007) Prediction of refractive index of vinyl polymers by using density functional theory. J Comp Chem 28(14):2336–2341

Zeng H et al. (2016) Adhesion and detachment mechanisms between polymer and solid substrate surfaces: using polystyrene–mica as a model system. Macromolecules 49(14):5223–5231

Zhao W, Cheng JL, Feng SD, Li G, Liu RP (2016) Intrinsic correlation between elastic modulus and atomic bond stiffness in metallic glasses. Mater Lett 175:227–230

Design and Modeling of High-Strength, High-Transmission Auto Glass with High Sound Transmission Loss

84

Mehmet C. Onbaşlı

Contents

Abstract

Development of high-strength, scratch-resistant and highly transparent glass with good sound isolation properties is necessary for automotive, industrial and architectural applications. While major progress has been achieved in developing the mechanical strength, scratch resistance and elasticity of highly transparent glass compositions, achieving broadband sound isolation in glass has remained a challenge due to the intrinsic thicknesses and the phonon band structure of the glass layers used. Since phonon band structure of high-strength glass limits the tunability of its acoustic transmission spectrum, composite multilayer glass designs need to be used to enhance sound transmission loss while maintaining the mechanical and optical merits. In this chapter, the design of high-strength, highly transparent and sound isolating glass is posed as a composite acoustic

M. C. Onbaşlı (✉)
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Department of Electrical and Electronics Engineering, Koç University, Istanbul, Turkey
e-mail: monbasli@ku.edu.tr

(meta)material design and optimization problem. Sound transmission loss within human audible range, particularly within 1–10 kHz, must be maximized toward 30 dB or more for good sound isolation while maintaining good optical and mechanical properties over large areas. Next, constitutive and wave equation relations for acoustic modeling of sound transmission loss are presented with example results for polymer/glass sandwich layers. Finally, future research opportunities on new glass materials with tunable phononic band structures, new metamaterial topologies and modeling methods are discussed.

1 Introduction

Glass is a multifunctional material that could be made transparent in optical, radio-frequency, or infrared bands, mechanically strong, highly elastic, and thermally insulating over large areas. These functionalities and the low cost of glass make this material be used ubiquitously. Glass is used in large volumes and areas in automotive, industrial, photovoltaic, and architectural applications, which demand glass to be mechanically strong, scratch-resistant, and highly transparent (Pambianchi et al. 2016). Glass must also have sound isolation capabilities for residential or commercial office user comfort, industrial process control, and driver safety. For instance, the windshield and the side and rear glass of a vehicle both protect drivers from particles or wind and isolate from thermal climate from the outside. In addition, these glass pieces must also isolate from noise originating from other vehicles, which may become a road hazard on highways or noisy urban settings (Leonhard et al. 2015). Some of the major glass manufacturers have developed mechanical properties through experimental and data-driven glass composition development (Mauro et al. 2016; ▶ Chap. 81, "Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass"). These examples show that state-of-the-art glass compositions and manufacturing processes can yield scratch resistance and superior strength in large volumes at low cost.

While mechanical properties mentioned above have been developed significantly with respect to those of soda lime glass through composition research (Tandia et al. 2019), sound isolation of glass has not been developed enough for rational design of acoustic transmission spectrum and enhanced sound transmission loss (Shengchun et al. 2010). These properties need to be developed with a materials modeling and engineering methodology while meeting manufacturability and mechanical and optical end property requirements simultaneously. These simultaneous requirements limit the flexibility in engineering sound absorption and reflection properties of glass. Engineering glass compositions do not yet seem to be a feasible direction toward optimizing acoustic phonons. In this chapter, I pose this challenge as a multi-constraint optimization problem and multilayer solution methods and examples are presented for enhancing sound transmission loss, particularly within 1–10 kHz (Leonhard et al. 2015). For good sound isolation, sound transmission loss within human audible range, particularly within 6–9 kHz, must be maximized beyond

30 dB or more. Acoustic constitutive relations and numerical models are presented to describe some of the multilayer solutions to this multi-objective materials optimization problem.

Sound signal is a time-dependent vibration of molecules in space, and this vibration brings a change in particle velocity when sound signal hits molecules. For solids such as a glass slab, due to limited particle displacement in compact molecular arrangement, a pressure gradient through the thickness of glass can be observed. In addition to this pressure gradient, the intrinsic molecular absorption due to acoustic phonon bands of the material leads to phase shift and absorption of the sound energy impinging on the glass slab. As a result, the transmitted acoustic signal is partially reflected and attenuated. This reflection and attenuation determines the fraction of sound energy that is transmitted through glass panes such as windshields and auto glass. The molecular vibration frequencies in the transmitted sound energy constitute the transmission spectrum of sound. When the transmitted sound power intensity is normalized with respect to the incoming sound power, one can determine the frequency response of glass panes. This frequency response is determined on a decibel (dB) scale, and thus sound transmission loss (STL) is defined as $STL = 10 \times \log_{10}(P_{transmitted}/P_{in})$. The purpose in this chapter is to overview composite glass pane designs for enhanced STL. Typically, multilayer planar glass/polymer/glass sandwiches are used for sound isolation. The polymer in most cases is polyvinyl butyral (PVB), which is a transparent resin that binds strongly on glass. Hence, the optical and mechanical properties of laminated composite glass-PVB structure help approach the desired specifications (Leonhard et al. 2015).

The acoustic transmission behavior of glass is determined by the mechanical properties, density, thickness, and elastic moduli of glass materials. The intrinsic acoustic scattering behavior of a single sheet of glass is determined by the electronic and phononic band structure of glass. In particular, acoustic phonons play a key role in absorption and scattering of incoming sound intensity and make up the intrinsic glass behavior. These band structures depend strongly on the composition and microscopic glass topology. Since glass is an amorphous material, rational design of phonon band structure of glass has been difficult. As a result, the intrinsic specific sound transmission loss behavior of a glass sheet is traditionally taken as a function of glass composition. Since glass composition leads to an essentially fixed acoustic transmission behavior, composite multilayer sandwich structures have been investigated more rigorously (Leonhard et al. 2015) for simultaneously optimizing mechanical, optical, and acoustic properties.

A key requirement in automotive glass design for windshields is lightweight, and this requirement is imposed on automotive industry by US regulations FMVSS 205 (ANSI Z26.1) and European requirements as defined in ECE R43 Rev. 3 for windshield applications. As a result, the auto glass panel designs should be as thin as possible. The automotive glass industry has traditionally been using trilayers of glass/polymer/glass-laminated composites for high strength, high optical transparency, and good sound and thermal isolation. Glass layers depend on the specific glass compositions developed by each glass manufacturer, and they can

be annealed soda-lime glass or other proprietary compositions. Polymer layer is typically chosen as standard polyvinyl butyral (SPVB), which binds strongly on glass and retains mechanical strength of the structure while providing good optical transparency. While fundamentally new glass compositions may emerge, end user windshield applications are most likely going to contain glass/polymer composites as the industry is strongly driven by high-volume manufacturability and well-established standards. As a result, successful glass models should be able to solve for sound transmission loss of multilayer composites in time, frequency, and real space domains. New emerging physics of acoustics and wave phenomena indicate that fundamental breakthroughs in sound isolation are also imminent (Yang and Sheng 2017).

In the following part, constitutive material relations and the wave equations used for acoustic modeling of sound transmission loss are presented with example results for polymer/glass sandwich layers. In the final part, future research opportunities on new glass materials with tunable phononic band structures, new (meta)material topologies, new mechanisms, and modeling methods are discussed.

2   Auto Glass Design: Materials and Acoustic Modeling

The second part of this chapter covers the acoustic wave equation formalism, model details, and example results on the sound transmission behavior of laminated glass/polymer composites. In the first subsection of the second part, the equations used for modeling the constitutive relations of glass/polymer composites are presented with a brief discussion of each term in the wave equation for sound transmission, boundary conditions, source definitions, and the details of finite element modeling steps. In the second subsection, example results on the sound transmission behavior of laminated glass/polymer composites are presented and discussed.

2.1   Constitutive Relations and Modeling Sound Transmission Loss

Sound transmission loss of laminated glass composites are modeled with finite element modeling using linear acoustic wave equations and material constitutive relations. In order to understand the acoustic transmission, reflection, and attenuation behavior of materials, one may use commercial acoustic finite element method solvers such as ANSYS MAPDL, ANSYS Acoustics ACT, Siemens LMS, MSC ACTRAN, or COMSOL Multiphysics Acoustics module. In the finite element model, first, the geometry of the structure to be simulated is defined. The sound wave or the pressure oscillates inside the enclosed structure and then exits the structure to the exterior medium.

In linear acoustics, pressure changes are assumed to be small with respect to the mean pressure:

$$p = p_0 + p', \quad p' \ll p_0 \tag{1}$$

Momentum and mass are conserved as described by Eqs. (2) and (3):

$$\frac{\partial \overrightarrow{\mathbf{v}}_a}{\partial t} = -\frac{1}{\rho_0} \nabla p_a + \frac{4\mu}{3\rho_0} \nabla \left(\nabla \cdot \overrightarrow{\mathbf{v}}_a \right) \tag{2}$$

$$\frac{1}{\rho_0 c_0{}^2} \frac{\partial p_a}{\partial t} = -\nabla \cdot \overrightarrow{\mathbf{v}}_a + \frac{Q}{\rho_0} \tag{3}$$

where ρ_0 is the mass density of the medium (kg·m^{-3}), p_a is the acoustic pressure (N·m^{-2}), v_a is the sound velocity (m·s^{-2}), and the Q/ρ_0 term (s^{-2}) is that captures mass loss due to sound wave-medium interaction. SI units are typically used. Here, constant and ambient temperature is assumed. Using Eqs. (2) and (3), the governing equation that describes the spatial and time dependence of linear acoustic pressure profile is found:

$$\begin{aligned}
\nabla \cdot \left(\frac{1}{\rho_0} \nabla p_a \right) &- \frac{1}{\rho_0 c_0{}^2} \frac{\partial^2 p_a}{\partial t^2} + \nabla \cdot \left[\frac{1}{\rho_0} \left(\frac{4\mu}{3} + \mu_B \right) \nabla \left(\frac{1}{\rho_0 c_0{}^2} \frac{\partial p_a}{\partial t} \right) \right] \\
&= -\frac{\partial}{\partial t} \left(\frac{Q}{\rho_0} \right) + \nabla \cdot \left[\frac{1}{\rho_0} \left(\frac{4\mu}{3} + \mu_B \right) \nabla \left(\frac{Q}{\rho_0} \right) \right]
\end{aligned} \tag{4}$$

In the models presented here, while no acoustic fluid-structural interaction is assumed, polymers may undergo glass transition with temperature, and their elastic moduli may change. As a result, temperature variations must also be modeled. This investigation is, however, beyond the scope of this chapter. For a deeper analysis of acoustic fluid-structural interaction with example cases, readers may refer to (Bouayed and Hamdi 2013; Kinsler et al. 2000; Kruntcheva 2007; Todo 2005).

2.1.1 Solution Types

Equations (2), (3), and (4) are solved for each node in the linear finite element acoustic model. The model types can be classified into three categories:

- *Modal solutions*, in which eigenmodes of the system are found. Understanding the modes of a given geometry could be important when unintentional resonances can emerge within the structure. These resonances may localize sound pressure and cause potential nonlinear structural damage. Therefore, one needs to avoid resonances within the audible range inside the glass/polymer composites to eliminate nonlinear sound localization and structural damage.
- *Harmonic solutions*, in which sound pressure level and power distribution at each port are found. This helps understand the sound transmission loss spectrum of a laminated windshield glass. In this chapter, we focus on harmonic solutions.

- *Transient solutions*, in which time evolution of the spatial dependence of sound pressure or acoustic energy density profile is solved. These solutions could be important in automotive glass product development as the windshield of a vehicle or the sound source is a moving part and the actual behavior of sound transmission loss may need to be accounted for using transient simulations of different angular orientations, relative velocities, and temperatures. In transient solutions, the temporal spacing between each sample is typically defined as $\Delta t = 1/(2f_{max})$, where f_{max} is the maximum frequency component of the sound wave incident on the structure. This definition originates from Nyquist theorem, which requires sampling a signal of maximum frequency f_{max} using a sampling rate of $2f_{max}$ in order to avoid aliasing. Aliasing is a misidentification of the frequency of a sound signal.

2.1.2   Material Properties

The material properties that govern sound isolation simulations include mass density, sound speed in the material, and dynamic viscosity. Density and the sound speed are fundamental material properties, which generally depend on temperature and sound frequency. Dynamic viscosity is defined as $\eta = F/(A \times SR)$, where F is the force, A is the area, and SR is the shear rate. Dynamic viscosity is a term in units of Pa·s, and it introduces dissipative effects and sound absorption losses. This viscosity term is both frequency and temperature dependent.

Sound wave propagation in outer medium such as air is modeled as a nonuniform ideal gas. In such media, the spatial and temporal evolution of sound waves is described using Eqs. (5), (6), and (7), which describe the pressure state and speed of sound (Kinsler et al. 2000):

$$\nabla \cdot \left(\frac{1}{\rho_0\left(\overrightarrow{r}\right)} \nabla p \right) - \frac{1}{\rho_0\left(\overrightarrow{r}\right) c^2\left(\overrightarrow{r}\right)} \frac{\partial^2 p}{\partial t^2} = -\frac{\partial}{\partial t} \left(\frac{Q}{\rho_0\left(\overrightarrow{r}\right)} \right) \tag{5}$$

$$c^2\left(\overrightarrow{r}\right) = \gamma RT\left(\overrightarrow{r}\right) \tag{6}$$

$$P_{state}\left(\overrightarrow{r}\right) = \rho\left(\overrightarrow{r}\right) RT\left(\overrightarrow{r}\right) \tag{7}$$

While these equations capture the dynamics of sound transmission, the interaction of pressure gradient with a rigid wall needs to be modeled using boundary conditions. For cases where temperature differences are also considered, viscous-thermal effect (pressure wave interaction with a perfect rigid wall or flexible wall) is introduced by solving the pressure equation coupled with Navier-Stokes equation (Zhao 2017).

2.1.3   Boundary Conditions

The boundary conditions in acoustics are defined by first introducing the following terminology:

- *Perfectly matched layer boundary (PML)*: absorbs incoming pressure waves.
- *Sound-hard surface*: A surface on which particle normal velocity is zero (or Neumann boundary with $\widehat{n} \cdot \nabla p = 0$).
- *Sound-soft surface*: A surface on which sound pressure is constrained (also known as Dirichlet boundary, where $p = p_0$ fixed pressure).
- *Mixed boundary*: The surface whose frequency-dependent wave transmission properties are modeled using a mixed impedance mismatch model, where acoustic impedance is defined as $Z = p/(vS)$, with p being the pressure, v being the sound velocity, and S being the surface area.
- *Port*: An exterior surface on which incident pressure is launched into the model.
- *Transparent port*: The pressure is reflected back to the infinity through this port.
- *Boundary layer impedance model*: Boundary layer thickness is much lower than the thickness of domains (or actual materials). The wavelength is also much greater than these layers (Zhao 2017).
- *Low reduced frequency model*: When the sound wavelength is much greater than the geometric feature sizes and boundary layer thicknesses, one can simplify the sound wave equation into:

$$j\overrightarrow{v}^{pd} = -c_0 \frac{1}{\overline{k}\gamma} \frac{1}{p_0} \nabla^{pd}\overline{p} + \frac{l^2}{s^2}\Delta^{cd}\overrightarrow{v}^{pd}$$

$$\frac{\overrightarrow{p}}{p_0} = \frac{\overline{\rho}}{\rho_0} + \frac{\overline{T}}{T_0}$$

$$l\nabla \cdot \overrightarrow{v} + \overline{k}c_0 \frac{\overline{\rho}}{\rho_0} = 0$$

$$j\overline{T} = \frac{l^2}{s^2\sigma^2}\Delta^{cd}\overline{T} + j\frac{\gamma - 1}{\gamma}\frac{T}{p_0}\overline{p}$$

where $s = l\sqrt{\frac{\rho_0\omega}{\mu}}$, $\overline{k} = \frac{\omega l}{c_0}$, $\gamma = \frac{c_p}{c_v}$, $\sigma^2 = \frac{\mu C_p}{\lambda}$

- *Wave absorbing boundary conditions*: The domain or open medium is terminated with infinite, homogeneous, inviscid fluid (gas), and, thus, outgoing propagating waves keep propagating outward and do not reflect back.

Among these boundary conditions, radiation and absorbing fluid element boundaries are used for all types of analyses (modal, harmonic, and transient), and perfectly matched layers are used for harmonic analysis. In the calculations of pressure profile on boundaries and inside materials, finite element mesh sizes are typically chosen as $\lambda/10$-$\lambda/5$ where $\lambda = c_0/f$.

2.1.4   Definition of Sources

The sound wave source is defined in different forms including incident pressure plane wave $P_{inc} = P_{inc,0}e^{j\varphi}e^{-j\overrightarrow{k}_i(\overrightarrow{r}-\overrightarrow{r_s})}$ and particle velocity $v_{n,f} = \frac{1}{\omega\rho_0}\widehat{n} \cdot \overrightarrow{k}_i p_{inc}$. Here, the mass density and the sound speed in the given environment must be known

for numerical calculation of sound transmission loss. While plane wave is the most fundamental source, other sound pressure sources are also used:

- Monopolar or spherical incident waves for a sphere of radius a

$$P_{inc} = aP_a e^{jka} \frac{e^{-j\vec{k}_i(\vec{r}-\vec{r}_s)}}{|\vec{r}-\vec{r}_s|} \tag{8}$$

- Dipolar waves as a combination of two monopoles with opposite signs and separation of distance d.
- Loudspeakers with monopole or dipole circular pressure radiators.
- Diffuse sound fields where pressure waves are launched with amplitude $A_n(\omega)$, distance d_n along N directions, and phase angles $\Psi_n = 0\text{--}2\pi$

$$p(r, \omega) = \sum_{n=1}^{N} A_n(\omega) e^{j\Psi_n} e^{-j\vec{k}_n d_n} \tag{9}$$

The last source type can also be used for studying the angular dependence of sound transmission loss. With the definitions of sources, we also define acoustic port modes, which are the junctions of entry and exit of the sound pressure waves. For a two-port structure, the amplitudes and phases of the incoming/outgoing sound pressure waves are defined as follows: P_{11} (wave incident from Port 1, reflected from Port 1), P_{21} (wave incident from Port 1 and exits from Port 2), P_{22} (wave incident from Port 2, reflected from Port 2), and P_{12} (wave incident from Port 2 and exits from Port 1). These parameters are used for describing the sound transmission loss in decibel scale by calculating P21 in postprocessing.

2.1.5 Postprocessing

After the simulations are completed, sound pressure level and sound transmission loss are calculated:

$$SPL \ (dB) = 10\log_{10}\left(\frac{\sum_{i=1}^{N} p^2_i}{p^2_{ref}}\right) \tag{10}$$

Here, sound pressure level at the threshold of human hearing is 0 dB, which has the same pressure as the reference pressure, $p_{ref} = 2 \times 10^{-5}$ Pa. Here, p_i terms are pressures at the sampled frequencies. To calculate sound transmission loss, time-averaged sound power through an area S must be calculated:

$$P_W = \frac{1}{2} \iint\limits_{S} p\vec{v}^* \cdot \hat{n} ds \tag{11}$$

$$\text{Sound transmission loss (TL) (dB)} = 10\log_{10}\left(\frac{P_{in}}{P_t}\right) \qquad (12)$$

$$\text{Sound reflection loss (RL) (dB)} = 10\log_{10}\left(\frac{P_{in}}{P_{ref}}\right) \qquad (13)$$

The wave description can also be viewed as a network of acoustic impedances connected to each other, where each layer or mesh's acoustic impedance Z is defined as $Z = p/(vS)$, where p is the pressure, v is the sound velocity, and S is the surface area. The reflections occur due to impedance mismatch between two dissimilar layers.

When reflections are calculated, one can plot the frequency dependence of sound pressure level and transmission loss on a given port based on Eqs. (10) and (12). By using symmetry of the geometry, one can also simplify the calculations and find far-field parameters, reflection, and sound transmission loss outside the finite element mesh region. Symmetry considerations also simplify the calculations of angular dependence of transmission loss, far-field sound power level, and sound pressure levels over multiple octave bands at far-field points.

2.1.6   High-Performance Computing (HPC) Considerations

As sound transmission loss simulations for automotive glass applications involve 3D mesh and calculations, high-performance computation is necessary. The commercial solvers can easily reduce solution times by an order of magnitude with respect to typical workstation computers with single-core central processing units. HPCs can reduce calculation times by using optimal mesh sizes, multiple acoustic models for different parts of the model structure, and efficient job distributions on HPC cores.

2.2   Sound Transmission Loss Models in PVB-Glass Multilayer Composites

Glass industry has well-established fabrication methods of fabricating polyvinyl butyral (PVB) layers on glass for acoustic and thermal isolation. This polymer is an optically transparent resin, which binds strongly on glass surfaces, thus retaining the mechanical strength and optical transparency of multilayer laminate composite materials (Pambianchi et al. 2016). Polymeric inner layers are used for composites because of their density and elastic modulus contrast with glass and because of their lightweight. Automotive emission regulations require each part in the automotive to be lighter, and thus polymers are preferred by glass industry for more than the past five decades. For instance, Corning Incorporated, which is one of the leading glass research and manufacturer companies, develops structures as thin as 3.56 mm using Corning's Gorilla Glass for Automotive (GGfA) applications. These glass layers are combined with annealed soda-lime glass (ASLG) and standard PVB layers (SPVB). Traditionally, glass laminate composite windshields are fabricated

as ASLG/SPVB/ASLG layers with 4.96 mm thickness. By substituting one of the ASLG layers with GGfA, the total glass layer thickness required for achieving similar strength is reduced as GGfA is strengthened using Corning's ion exchange process (Leonhard et al. 2015) as shown in Fig. 1. This process is a glass strengthening method, which prevents crack propagation due to Na/K ion concentration and stress gradient near glass surface. These layers are then tested and qualified to meet US regulations FMVSS 205 (ANSI Z26.1) and European requirements as defined in ECE R43 Rev. 3 for windshield applications.

While state-of-the-art glass composition engineering enables improvements in scratch resistance and mechanical strength of glasses, modeling the sound transmission loss of laminated sandwich panels is necessary in achieving enhanced sound isolation. In the rest of this part, sandwich panel modeling results are presented for laminated sheets of glass and a structural middle layer. These modeling results are representative of common sandwich panel design trade-offs in STL.

In this part, the sound transmission loss spectra of unbounded orthotropic sandwich panels shown in Fig. 2 are investigated. The modeled structure is a three-layer sandwich of glass top and bottom face sheets and a honeycomb structural core. Sound waves impinge on the top face sheet and cause transverse shear deformation in the layers. As a result, the thickness, rigidity, and other mechanical properties of the layers become important. In this model, Young's modulus of the face sheet

Fig. 1 Automotive windshield glass is traditionally designed as sandwiches of annealed soda-lime glass (ASLG) and standard PVB layers (SPVB). Using Corning's ion exchange process, one can strengthen glass and substitute one of the glass layers with Corning's Gorilla Glass for Automotive (GGfA) applications. As a result, the total windshield thickness is reduced from 4.96 to 3.56 mm (Leonhard et al. 2015, reprinted with permission from SAE International)
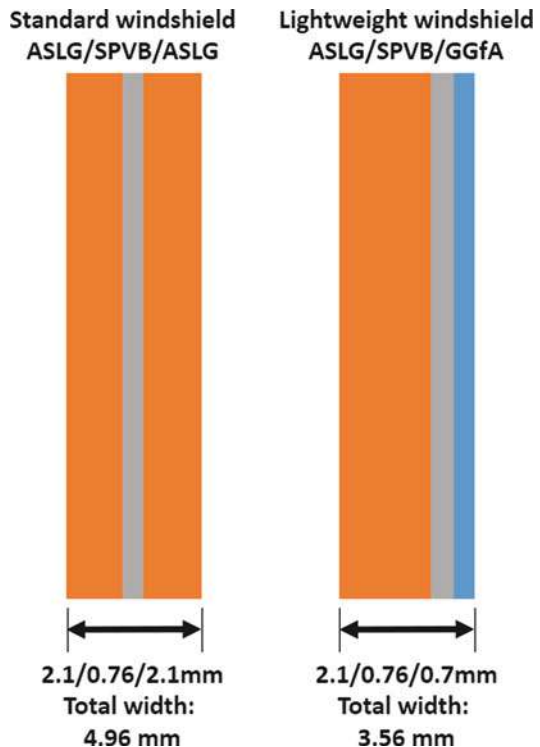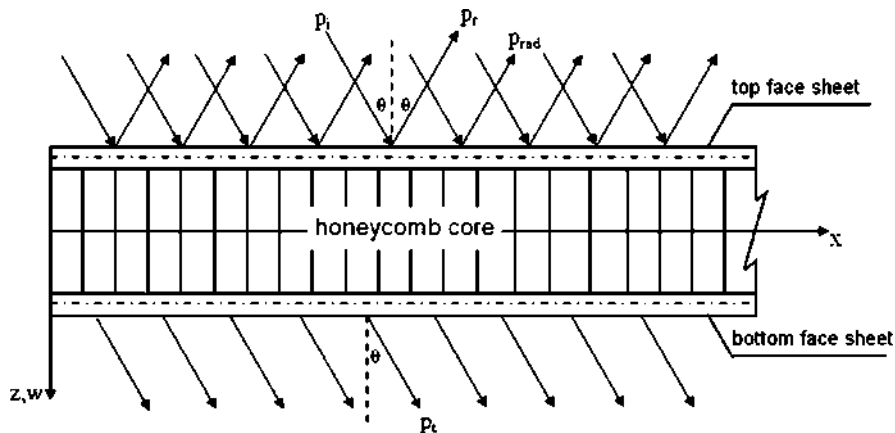
Fig. 2 Structure of the multilayer orthotropic sandwich panel (glass/honeycomb core/glass) under acoustic excitation. Here, sound transmission loss (STL) is defined as STL = $10 \times \log10(P_t/P_{in})$ (Shengchun et al. 2010, reprinted with permission from Elsevier Ltd.)

has been taken as 6.37 GPa, face sheet's density as 630 kg/m^3, face sheet thickness as 6.35 mm, Poisson's ratio of face sheet as 0.33, shear modulus of the core as G_{xz} = 23 MPa and G_{yz} = 50 MPa, core density as 28 kg/m^3, and core thickness as 73.5 mm.

Parametric studies were done to determine the effects of shear rigidity and thickness of the panel face sheets on STL. The acoustic wave equation solutions have been calculated based on FEM discussed in Part 2.1. Shear rigidity of the honeycomb layer transforms an incoming longitudinal acoustic pressure wave into an in-plane transverse stress gradient, and, thus, STL is enhanced at a wide range of frequencies. A caveat of this approach has been found in the low-frequency regime, in which STL decreases to a few dB when shear rigidity increases from 10^6 to 10^7 N/m as shown in Fig. 3. Figure 4 shows the dependence of STL on the top glass sheet thickness, which yields an STL dip for higher frequencies (1–10 kHz) due to resonant coupling and transmission of sound pressure waves. Since these thicknesses are necessary for structural and mechanical strength of the layers, a trade-off in STL at higher frequencies emerges. This trade-off could be overcome by adding further layers into the structure and reducing the resonance strength of the dips shown in Fig. 4.

3   Conclusions and Future Work

3.1   Glass Composite Models for Enhanced Sound Isolation Loss

Glass industry is highly conservative in the manufacturing processes and layer structures for windshields. As a result, glass models for enhancing sound transmis-
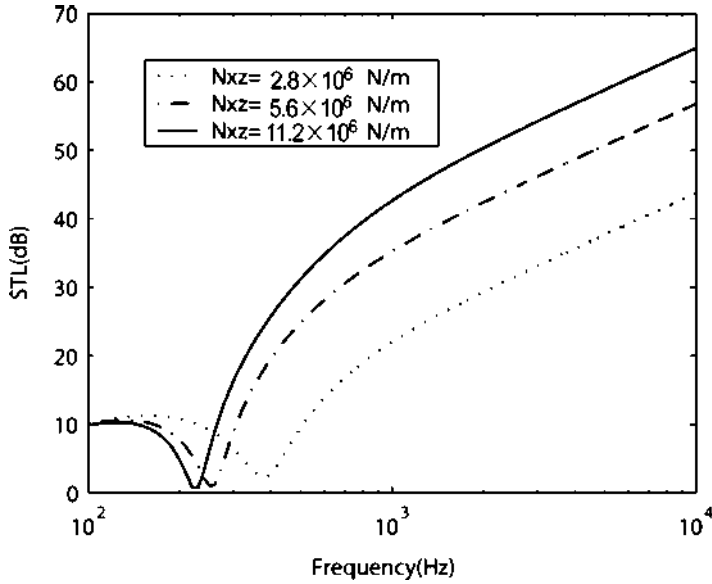
Fig. 3 The effect of increasing shear rigidity of the core on field STL of orthotropic sandwich panel. Higher rigidity shifts STL dip to lower frequencies (Shengchun et al. 2010, reprinted with permission from Elsevier Ltd.)

sion loss must continue to follow laminated glass/PVB/glass or similar multilayer composites. While simplified models may assume infinitely long composite sheets, realistic models need to account for the finite size effects as well. In addition, the angular uniformity of STL of curved composite windshield layers becomes another important figure of merit, which should be maximized. Since both geometric and material properties become important in these laminated composite structures, simultaneous iterative models of both material properties and windshield geometry should be developed.

3.2 Opportunities for New Glass Models and Materials with Improved Sound Isolation

Simultaneous optimization of mechanical strength, scratch resistance, optical transmission, and sound isolation is a challenge due to the intrinsic trade-offs imposed from the phononic band structure, microstructure, and geometric requirements. For high mechanical strength and elastic modulus, the glass composition space must include some key dopant constituents to allow for wide dynamic range for strain in the elastic deformation regime. In addition, since auto glass applications also demand fabrication of clear sheets of glass over large areas, one needs compatibility with large-area and large-volume manufacturing methods such as Corning's fusion
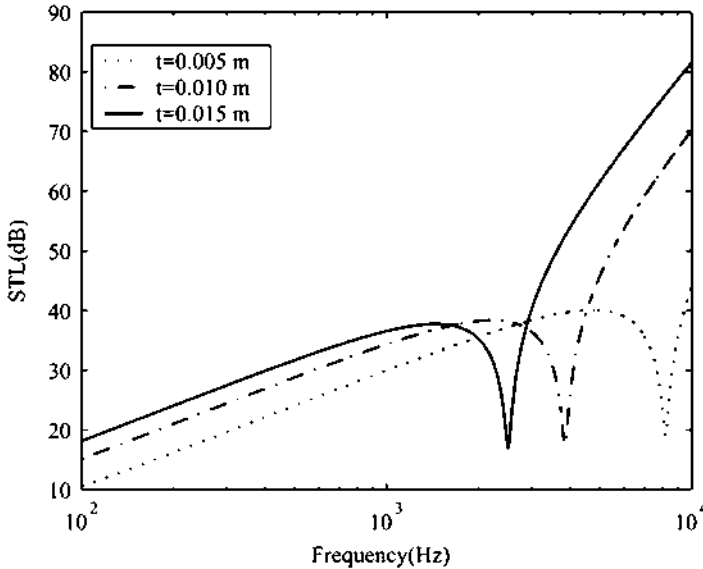
Fig. 4 Effect of face glass sheet thickness on STL of the sandwich panel. Increasing glass panel thickness shifts the STL transmission spectrum dip from about 8 to 2.5 kHz (Shengchun et al. 2010, reprinted with permission from Elsevier Ltd.)

draw process or other large-volume production methods (▶ Chap. 81, "Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass"). While these simultaneous requirements pose a complex challenge, researchers have been addressing some of these complexities by introducing new fundamental composite structures with unique effective mechanical and optical properties as well as design paradigms. Below is a list of fields in acoustics and wave phenomena in general which are promising in developing models and composites that could address the sound-isolating, strong, elastic, and transparent glass design problem with a few constraints at a time.

Acoustic Metamaterials: Metamaterials are composites, which are rationally designed with key geometric features to achieve desired effective Young's moduli and Poisson's ratios not found in nature. In metamaterials, the overall material properties become strong functions of geometry, symmetry, and microscopic composite periodicity than would otherwise be in a regular single-phase structural material. As a result, one can engineer the density (ρ) and compressibility (C) of the effective metamaterial structure. For instance, resonant coupling of materials in a Helmholtz coil configuration may lead to negative dynamic C (C < 0, ρ > 0), or materials could be engineered to oscillate in or out of phase with the force when the drive frequency is below or above the resonance frequency, respectively. Thus, negative dynamic density could be achieved through out-of-phase time harmonic motion of the moving mass. Evanescent acoustic modes can be engineered to have much stronger effects by design to achieve negative C or ρ or both. For further overview,

readers are referred to excellent reviews on this subject (Wu et al. 2018; Yang and Sheng 2017; Fang et al. 2006; Ma and Sheng 2016; Cummer et al. 2016; Maldovan 2013; Greaves et al. 2011; Yamamoto 2018). With resonant metamaterial structures, giant sound loss enhancements could be engineered at the desired sound frequencies by design. As the components of metamaterials could also be optically transparent, metamaterial design strategy could be a solution to the sound isolation loss problem. For engineering of the intrinsic glass properties, readers are referred to the other glass design chapters in this section (▶ Chaps. 83, "Fundamentals of Organic-Glass Adhesion," and ▶ 81, "Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass").

Simultaneous Sound and Heat Isolation: Since acoustic transmission is determined by electron and phonon band structure, transmission of sound and heat waves through glass composites are tightly coupled. Therefore, a natural outcome of designing for high sound isolation loss could be enhanced thermal insulation as well. Previously, Raman and Brillouin scattering studies experimentally indicated the coupling of phonons and photons, but, recently, a mechanism involving the simultaneous localization of mechanical and optical waves by using acoustic and optical cavities has been introduced to significantly enhance sound and light interactions (Maldovan 2013). As a result, one may design and use a new class of acousto-optical crystal that can integrate the combined management of phonons and photons (Maldovan 2013). Acoustic transmission spectra can be controlled using nonlinear acousto-optic metamaterials whose phonon band structures are a strong function of "capacitive and inductive" behavior of the vibration spectrum of the periodic lattice. As a result, one can achieve large-scale thermal diode behavior with large sound isolation loss. An excellent review paper (Maldovan 2013) describes phonon cavity engineering principles and presents examples for bridging mesoscopic and microscopic designs of phononic band structures for the desired sound and heat transmission loss. The engineering of phonon band structure may help simultaneous achievement of strength and sound isolation loss in glass composites.

Active Control of Sound Transmission Loss: Acoustic transmission spectra could also be controlled on demand by applying voltage. If mechanical properties of one of the layers in a glass/polymer composite depend on voltage, then sound transmission loss could be enhanced by voltage bias to stiffen or soften the structural layer within laminated glass sheets. As a result, glass could be operated in "silent mode" with voltage bias (Zheludev and Kivshar 2012). Two key requirements for such a design are, first, a sizable elastic modulus and density change upon voltage for sound isolation enhancement and, second, a negligible or no change in optical transparency. In order to achieve such demanding properties, one may use 2D materials with transparent contacts or functionalized PVB layers (Kraft et al. 2006). While major research on electrochromic windows progressed enough for commercialization (Wang et al. 2016), electroacoustic glasses may become an emerging area for research.

Elastic Wave Manipulation with Nonreciprocity and PT Symmetry: Recent investigations in wave phenomena indicate that one can engineer the photonic,

electronic, phonon, and spin-resolved band structures with certain symmetries. These symmetries include point or planar inversion, parity, time-reversal, rotational, and translational symmetry. Some of these symmetries such as rotational, translational, and inversion symmetries are observed in real space, while time-reversal symmetry is observed in time or wave vector (or frequency) space. When these symmetries are broken, one can control the propagation and attenuation of photons, phonons, spins, and wave particles on demand. For instance, introducing a magnetic dopant into a material and applying magnetic field in a direction allows for asymmetric photon and phonon transport for positive and negative wave vectors of the same frequency, and thus $\omega(k) \neq \omega(-k)$. A time-dependent control of density or elastic modulus of glass layers can also enable time-reversal symmetry breaking. Using these symmetry-breaking principles, one can engineer one of the layers in glass pane composites to yield nonreciprocal sound transport properties, thus yielding high sound transmission loss. Since glass and PVB are amorphous materials, there are no real space lattice symmetries to break. On the other hand, one can use time or frequency domain symmetry-breaking methods such as controlling density or elastic modulus of one of the layers using time-dependent voltage or magnetic field. This control allows for nonreciprocal propagation of sound. Thus, glass/PVB laminate composites could be engineered for high nonreciprocal sound transmission loss and isolation (Ma and Sheng 2016; Haberman and Norris 2016; Yang and Sheng 2017; Fang et al. 2006; Cummer et al. 2016; Popa and Cummer 2014; Greaves et al. 2011; Krushynska et al. 2018; Yan et al. 2018; Kushwaha et al. 1993).

Bioinspired Acousto-optic Structural Materials Engineering: Natural structural materials are formed out of a limited number of organic/inorganic compounds at room temperature. These layers consist of hard and soft materials organized in complex hierarchical architectures, with characteristic dimensions ranging from the nanoscale to the macroscale. As a result, these materials are lightweight and commonly have unique combinations of strength and toughness, although synthetic replication of these materials has been challenging (Fig. 5).

The room temperature formation kinetics, the hierarchical structure, and the unique mechanical properties of biological materials may enable enhanced sound transmission loss.

Hybrid-Nonlinear Acoustic Materials: In this chapter, the intensity of sound pressure level has been assumed to remain small enough with respect to mean pressure as shown with Eq. 1, such that wave transport always remains in the linear regime. On the other hand, when this assumption is relaxed, acoustic transmission spectra can be controlled using nonlinear acousto-optic filtering. In nonlinear acoustic transmission, sound isolation and nonreciprocal sound transmission can be achieved through asymmetric frequency conversion and filtering the fundamental frequency.

To exploit nonlinearities in sound isolation, the medium needs to be a nonlinear acoustic material and a frequency-selective mirror (i.e., a periodic phononic crystal). These two layers need to be used together as shown on Fig. 6. The structure consists of a frequency-selective surface (FSS) and a nonlinear acoustic layer. The mirror
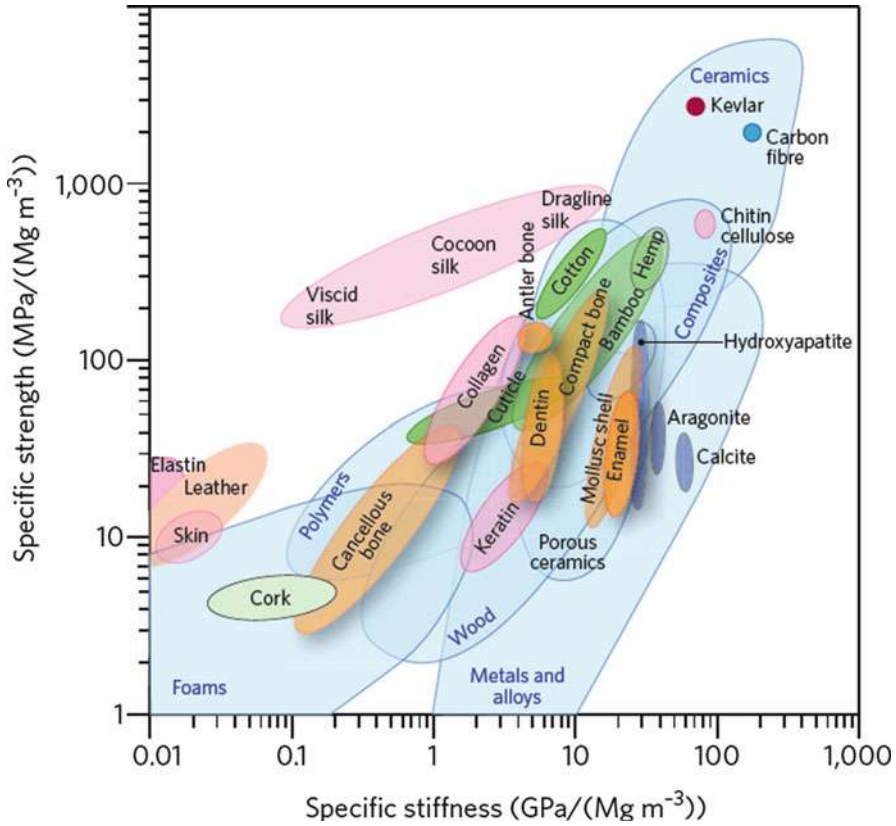
Fig. 5 Ashby plot for classes of structural materials with respect to their specific strength and specific stiffness values. Naturally occurring biological materials (shown in pink) yield unique strength/stiffness combinations (Wegst et al. 2014). (Reprinted with permission from Springer Nature)

reflects the fundamental frequency (f_0) but transmits the second or higher-order harmonic ($2f_0$ or nf_0). The nonlinear acoustic layer is first excited from the left side (i.e., from inside windshield) of the filter at fundamental frequency in the audible range, and sound waves are reflected by the structure as shown on Fig. 6. When the structure is excited from the right-hand side of the structure (from outside), the nonlinear layer converts most of the incoming acoustic energy to the second or higher harmonic frequency, which is transmitted through the frequency-selective mirror. Thus, the sound frequencies transmitted from outside (right) to inside (left) could be tuned to beyond audible range (Fig. 6).

Nonlinear acoustic materials are modeled using non-Hermitian Hamiltonians, which describe systems with loss and/or gain or other higher-order effects, which may upconvert incoming sound frequencies to beyond audible range and help enhance sound isolation below 20 kHz.
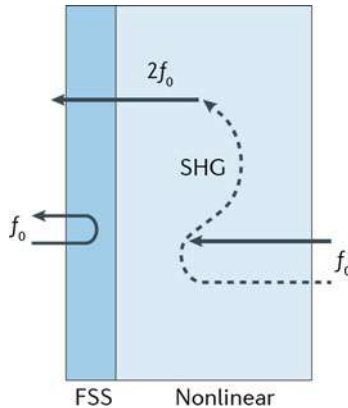
Fig. 6 Bilayer of frequency-selective surface (FSS) and nonlinear acoustic layer helps eliminate sound transmission at audible frequencies from outside (right) to inside (left of the structure). The FSS reflects sound back inside at the fundamental frequency, while the incoming sound from outside is upconverted to second- or higher-order harmonics with high efficiency to avoid audible range (Cummer et al. 2016). (Reprinted with permission from Springer Nature)

References

Bouayed K, Hamdi M (2013) A dynamic response of a laminated windshield with viscoelastic core – numerical vs. experimental results. Proc Mtgs Acoust 19:065025. https://doi.org/10.1121/1.4799567

Cummer SA, Christensen J, Alù A (2016) Controlling sound with acoustic metamaterials. Nat Rev Mater 1:1–13

Fang N, Xi D, Xu J, Ambati M, Srituravanich W, Sun C, Zhang X (2006) Ultrasonic metamaterials with negative modulus. Nat Mater 5:452–456

Greaves GN, Greer AL, Lakes RS, Rouxel T (2011) Poisson's ratio and modern materials. Nat Mater 10:823–837

Haberman MR, Norris AN (2016) Acoustic metamaterials. Acoust Today 12:31–39

Kinsler LE, Frey AR, Coppens AB, Sanders JV (2000) Fundamentals of acoustics. Wiley, New York

Kraft A, Rottmann M, Heckner K (2006) Large-area electrochromic glazing with ion-conducting PVB interlayer and two complementary electrodeposited electrochromic layers. Sol Energy Mater Sol Cells 90:469–476

Kruntcheva MR (2007) Acoustic-structural coupling of the automobile passenger compartment. In: Proceedings of the world congress on engineering WCE 2007, vol II, 2–4, London

Krushynska AO, Galich P, Bosia F, Pugno NM, Rudykh S (2018) Hybrid metamaterials combining pentamode lattices and phononic plates. Appl Phys Lett 113:201901

Kushwaha MS, Halevi P, Dobrzynski L, Djafari-Rouhani B (1993) Acoustic band structure of periodic elastic composites. Phys Rev Lett 71:2022–2025

Leonhard T, Cleary T, Moore M, Seyler S, Fisher WK (2015) Novel lightweight laminate concept with ultrathin chemically strengthened glass for automotive windshields. SAE Int J Passeng Cars – Mech Syst 8(1):95–103. https://doi.org/10.4271/2015-01-1376

Ma G, Sheng P (2016) Acoustic metamaterials: from local resonances to broad horizons. Sci Adv 2:e1501595

Maldovan M (2013) Sound and heat revolutions in phononics. Nature 503:209–217

Mauro JC, Tandia A, Vargheese D, Mauro YZ, Smedskjaer MM (2016) Accelerating the design of functional glasses through modeling. Chem Mater 28:4267–4277

Pambianchi MS, Dejneka M, Gross T, Ellison A, Gomez S, Price J, Fang Y, Tandon P, Bookbinder D, Li M (2016) Corning incorporated: designing a new future with glass and optics. In: Madsen LD, Svedberg EB (eds) Materials research for manufacturing. Springer Series in Materials Science, vol 224. Springer, Cham, pp 1–38

Popa B, Cummer SA (2014) Non-reciprocal and highly nonlinear active acoustic metamaterials. Nat Commun 5:3398

Shengchun W, Zhaoxiang D, Weidong S (2010) Sound transmission loss characteristics of unbounded orthotropic sandwich panels in bending vibration considering transverse shear deformation. Compos Struct 92:2885–2889

Tandia A, Onbasli MC, Mauro JC (2019) Machine learning for glass modeling. In: David Musgraves J, Hu J, Calvez L (eds) Springer Handbook of Glass. Springer, Berlin/Heidelberg, pp 1157–1192

Todo M (2005) New type of acoustic filter using periodic polymer layers for measuring audio signal components excited by amplitude-modulated high-intensity ultrasonic waves. J Audio Eng Soc 53:930–941

Wang Y, Runnerstrom EL, Milliron DJ (2016) Switchable materials for smart windows. Ann Rev Chem Biomol Eng 7:283–304

Wegst UGK, Bai H, Saiz E, Tomsia AP, Ritchie RO (2014) Bioinspired structural materials. Nat Mater 14:23–36

Wu Y, Yang M, Sheng P (2018) Perspective: acoustic metamaterials in transition. J Appl Phys 123:090901

Yamamoto T (2018) Acoustic metamaterial plate embedded with Helmholtz resonators for extraordinary sound transmission loss. J Appl Phys 123:215110

Yan M, Lu J, Li F, Deng W, Huang X, Ma J, Liu Z (2018) On-chip valley topological materials for elastic wave manipulation. Nat Mater 17:993–999

Yang M, Sheng P (2017) Sound absorption structures: from porous media to acoustic metamaterials. Ann Rev Mater Res 47:83–114

Zhao L (2017) Introduction to ANSYS Mechanical Acoustics. https://www.ansys.com/services/training-center/structures/ansysmechanical-acoustics

Zheludev NI, Kivshar YS (2012) From metamaterials to metadevices. Nat Mater 11:917–924

Part XIII
Nuclear Materials

A Decade of Nuclear Materials Modeling: Status and Challenges

85

Jaime Marian and Ram Devanathan

Contents

Abstract

The performance of materials used in nuclear technologies is controlled by mechanisms that have their origin at the atomic level. Materials modeling is needed in conjunction with theory and experimental observations to develop fundamental understanding of energetic processes that span disparate scales in space and time. Modeling of nuclear materials has been in use continuously since the early developments in the 1950s. It is by no means a mature field, with constant advances still being made on many levels, including theory, simulation, and algorithms. This can be ultimately linked to the fact that we do not currently have a complete theory of irradiation damage in materials. Multiscale materials modeling has stepped up to fill that gap and continues to address the research

J. Marian (✉)
Department of Materials Science and Engineering and Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA

Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA
e-mail: jmarian@ucla.edu

R. Devanathan
Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA
e-mail: ram.devanathan@pnnl.gov

needs of the nuclear materials community. This section provides the reader with a rigorous and practical description of the different elements of multiscale modeling that have had and are having the broadest impact on the field.

1   Introduction

The chapters in this section provide a detailed description of methods used to model nuclear materials from the electronic to continuum scales. While modeling of nuclear materials takes advantage of both established and novel techniques being developed within the multiscale modeling framework (Curtin and Miller 2003) for other purposes, the particularities of irradiation exposure and the associated material property degradation effects demand specific modifications – if not full-blown reformulations – that now make it a specific branch within multiscale modeling of materials (MMM). This should come as no surprise, as it is well known that energetic particle irradiation is an inherently multiscale process, acting at the nanometer scale in its primary form and transferring its effects to the engineering scales. Likewise, femtosecond-scale processes propagate themselves through time until their effects are felt after hours and even months and years of continued operation. This complexity is compounded by the fact that the effects brought about by energetic particle irradiation are of strikingly varied nature: alloy chemistry, heat transfer, mechanical behavior, melting, and diffusion, among many others, are all known to play an important role at some point or another along the spatiotemporal sequence (Zinkle and Busby 2009). This multidisciplinarity is an inherent characteristic of multiscale modeling of irradiated materials and is reflected in the group of authors contributing to this section. These contributions integrate well with the chapters of Part VII of MTM, entitled "Modeling of Microstructure Evolution: Mesoscale Challenges" edited by Stan and Sarrao in MTM, with Part XIV "Modeling of Radiation Damage in Materials: Best Practices and Future Directions" edited by Nordlund and Short and other stand-alone publications that provide a foundational understanding of models of radiation-induced changes in microstructure and properties of materials.

2   Contributions

This overview offers a brief commentary on the different chapters to introduce the reader to the field. ▶ Chapter 86, "Density Functional Theory Calculations Applied to Nuclear Fuels," by Andersson focuses on electronic structure calculations in nuclear fuels. This is an extremely challenging problem due to the strongly correlated nature of $5f$ electrons in uranium, which makes the use of standard exchange-correlation schemes within density functional theory (DFT) insufficient to capture the properties of fuels based on UO_2. Advances in this field are critical to push along in the development of advanced fuels such as metallic fuels, UN, or U_3Si_2.

The author discusses in detail calculations of point defects, both under thermal equilibrium and under far-from-equilibrium situations (under irradiation), and the structure and properties of Xe atoms generated as a fission product and its implications for fission gas accumulation in the fuel. This is one of the most important issues in the long-term structural integrity of fuel elements in nuclear reactors, and although semiempirical calculations are gaining traction for their ability to deal with larger systems, DFT and its variants are still a fundamental piece in our modeling portfolio to study the complex structure of U-based compounds.

Given the short length and time scales accessible to DFT-based methods, there is a need to develop empirical potentials to study energetic radiation damage processes in materials using molecular dynamics and Monte Carlo simulations. ▶ Chapter 87, "Interatomic Potentials for Nuclear Materials," by Devanathan provides an overview of commonly used potential functions and reviews the suitability of different potentials to model defects and radiation effects in materials, such as α-Fe, SiC, pyrochlores, and UO_2. This chapter serves as a bridge between Andersson's chapter devoted to DFT and subsequent chapters on molecular dynamics and kinetic Monte Carlo methods. DFT data are often used along with experimental data to fit interatomic potentials. In addition to fitting to elastic constants and lattice constants from DFT, fitting to forces derived from DFT calculations for representative model liquids is a promising strategy to develop reliable interatomic potentials. Several interatomic potentials have been developed for commonly studied materials, such as Fe and UO_2. DFT results for defect properties not used to fit the potential can be used as a standard against which the results of the potentials can be compared. Often, no single potential stands out as being clearly superior for modeling dynamic defect properties due to the limitations of the functional forms used. The combination of high-throughput DFT calculations and machine learning has the potential to identify new functional forms and parameters for potentials to simulate radiation damage. Ultimately, the development of reliable interatomic potentials is closely tied to the reliability of MD simulations of radiation damage presented in the next chapter.

Djurabekova and Nordlund provide a comprehensive overview of the uses of molecular dynamics for simulations of energetic particle irradiation in ▶ Chap. 88, "Molecular Dynamics Simulations of Non-equilibrium Systems." Molecular dynamics (MD) is a fundamental tool in the computational modeling portfolio due to its intrinsic high temporal and spatial resolution, efficiency and scalability, versatility to study all kinds of systems and geometries, etc. As it relates to bulk irradiation damage from high-energy particles such as neutrons or heavy ions, MD has provided key insights into the primary damage production process, including the start of the displacement cascade, the thermal spike, quenching phase, and defect survival phases. For example, MD conclusively showed that primary damage production does not scale linearly with recoil energy, an assumption in place in the irradiation damage community for decades, but that it follows instead a power law due to breakup into subcascades or, in some selected cases, saturates above a threshold energy. As well, it was through MD simulations that cornerstone theories such as the production and dislocation bias models have been verified, a fact that has had very large implications in our understanding of irradiation damage processes.

This would not have been possible without the capability to simulate tens and even hundreds of millions of atom systems, which are now routine in many research groups thanks to very successful optimization efforts over the years to make MD codes efficient and scalable on massively parallel architectures. The authors also discuss the role played by electron-phonon coupling as a viable way of extracting heat from the molten zone following the thermal spike. This is typically modeled via two-temperature models cast over the underlying particle dynamics.

MD simulations have also proven critical to study light-ion deposition processes on material surfaces, e.g., as in the case of fusion first walls exposed to H and He plasma ions. Simulations have given researchers access to the mechanisms behind physical and chemical sputtering in graphite surfaces, erosion in W and B materials, and surface morphology evolution in W. The authors make a convincing case for molecular dynamics to continue being an essential element of the modeler's toolbox for years to come to study irradiation processes in materials.

In ▶ Chap. 89, "Kinetic Monte Carlo Algorithms for Nuclear Materials Applications," Balbuena, Caturla, and Martinez cover the kinetic Monte Carlo (KMC) technique and its application to irradiation damage accumulation in representative structural metals. KMC simulations have established themselves in this field by providing access to time and length scales that range between those covered in molecular dynamics simulations and those accessible with mean field rate theory calculations. As such, they are ideally suited to study post-primary damage kinetics while retaining full three-dimensional spatial resolution. For example, they are able to capture the fine details of defect evolution in stages II, III, and IV of resistivity recovery experiments, including correlated recombination, cascade aging, and the early stages of damage accumulation. Often, KMC is offered in contraposition to rate theory calculations, which, although capable of reaching longer time scales and irradiation doses, fail to capture the highly heterogeneous nature of irradiation damage in metals. One of the reasons hindering progress in the time scale front for KMC simulations is the disjoint nature of defect diffusion in metals, with self-interstitials moving practically athermally along rectilinear trajectories at very high rates and vacancy clusters displaying very low mobilities. This results in time evolutions controlled by high rate processes with slow time progress. Another difficulty intrinsic to the KMC method is the stochastic nature of time evolution, which makes it difficult to take advantage of parallelization, in contrast with almost all other methods. While new algorithmic advances have shown some promise, there are opportunities for more improvement in this topic. The authors also discuss the development of new methods that determine events "on the fly," that is, calculating locally in space and time the different transitions available to a given evolving species. Following new efficiency improvements, these methods have opened the door to unprecedentedly complex systems, such as neutron-irradiated materials with changing chemical inventory as well as multicomponent alloys. The future promises to be bright for KMC practitioners as more improvements are made and the capabilities of the technique are expanded.

However, achieving the high doses required in power devices, or in next-generation reactor concepts, is still an insurmountable task for direct KMC

calculations. For that, one must give up some level of resolution in order to gain computational expediency. In ▶ Chap. 90, "Rate Theory of Radiation Damage," Ghoniem provides a rigorous overview of the "mean field" kinetic theory of defects, identifying the steps taken to arrive at systems of rate equations describing defect accumulation evolution. The mean field representation is accomplished by spatial averaging of defect populations and treatment of those populations under classical nucleation theory. This intrinsically assumes that all cluster species are homogeneously distributed spatially, something dubious in the early stages of cascade evolution, and evolve by either absorption or emission of monomers. This is valid when clusters are assumed to be immobile, implying that growth/shrinkage is mediated by the motion of single vacancies or self-interstitial atoms. The internal microstructure is accounted for in terms of defect sinks such as dislocations, grain boundaries, and/or inclusions and precipitates. The chapter discusses breakthroughs in the application of rate theory models to radiation damage such as the production and dislocation bias models, which have been instrumental in explaining features such as swelling and hardening in nuclear materials. While rate theory methods have become the de facto computational "workhorse" for radiation damage accumulation calculations, they are nonetheless riddled by some strong limitations, such as the absence of fluctuations in a highly heterogeneous scenario or the phenomenon of "combinatorial explosion," i.e., the nonlinear growth in the number of equations with the number of distinct chemical species. This limits their application to situations with a rich chemical diversity (e.g., as in the case of nuclear transmutation). Despite this, these methods continue to be an essential tool to provide us with long-dose estimates of defect evolution.

▶ Chapter 91, "Discrete Dislocation Dynamics Simulations of Irradiation Hardening in Nuclear Materials," by Marian, Fitzgerald, and Po provides a comprehensive review of the use of the dislocation dynamics (DD) technique in the context of irradiated materials. The authors discuss key features of the approach, and specific enhancements both in terms of theory and implementation develop to study hardening in irradiated microstructures. In addition to all the challenges surrounding accurate and efficient DD simulations in unirradiated materials to study strain hardening, the presence of irradiation defects adds yet another layer of complexity to the approach. The main types of defects that emerge during irradiation of materials can be broadly classified into three main groups. Dislocation loops can be treated directly within linear elasticity, and in principle they require the least amount of implementation effort. However, osmotic forces must be introduced to balance elastic attraction forces which can result in artificial loop shrinkage if untreated. As well, faceting of loops – a well-known feature in anisotropic crystals – has been dealt with by developing efficient formulations of anisotropic linear elasticity. The second type of defects considered is precipitates and/or inclusions. These are generally treated using Eshelby's solutions for elliptical inclusions in elastic media and are now routine in most DD codes. Finally, voids and bubbles can be challenging to deal with, as they involve free surfaces and image forces. However, recent finite element implementations, albeit costly, are now also routinely employed for this purpose. Present DD codes are now also capable of simulating plastic flow localization, a key

feature of microstructural evolution in high-dose irradiation in metals. Challenges remain, not the least of which is pushing the envelope in the amount of plastic strain that can be simulated, and to formalize the coupling of DD with crystal plasticity or continuum theories of deformation. From a physics point of view, perhaps, a remaining challenge not yet tackled to the same level of other developments is that of irradiation-enabled climb and its greater implications for creep at high temperature. In all, the chapter highlights the state of the art of dislocation dynamics simulations of irradiated materials as a healthy field that has and is contributing with key findings to our understanding of nuclear and ion-implanted materials.

▶ Chapter 92, "Mesoscopic Modelling of Irradiation Damage Processes: Bridging Many-Body Mechanics and Thermodynamics in Rate Processes," by Woo explores the mesoscopic connection between statistical mechanics and macroscopic thermodynamics. Progress in bridging the quantum mechanical world of discrete energy and the engineering world of temperatures and thermodynamic functions is assessed, and the effects of many-body dynamics and quantum statistics on reaction kinetics are discussed. The discussion is centered on the basic assumptions that underlie traditional efforts to model irradiation damage using continuum theory and their limitations, especially when it comes to treating low activation barrier events. To overcome these limitations, a Langevin equation scheme in the Mori-Zwanzig formalism is presented. The approach is especially relevant to understanding clustering of defects produced by radiation damage. The author identifies new mesoscale science frontiers, such as materials response to extreme environments and defect processes in ferromagnetic metals.

Spencer, Schwen, and Hales offer an overview of the governing equations for modeling materials in nuclear reactor environments in ▶ Chap. 93, "Multiphysics Modeling of Nuclear Materials," by Spencer et al. This chapter begins by reviewing the physics behind the behavior of fuels and structural components in a reactor, presents numerical methods and discretization techniques, and concludes with examples of relevant multiphysics problems. For light water reactor fuel performance modeling, the fuel microstructural evolution is of great interest. Thermal and stress gradients present in the fuel affect the migration of chemical species and defect clusters produced by irradiation. The modeling scheme has to consider ballistic radiation damage, chemical mixing, added fission products, fission gas resolution from bubbles, creep, and fracture. These conditions are clearly far from equilibrium. The fuel centerline temperature is an important quantity of interest because of its influence on neutronics and fuel performance. This temperature depends on the thermal conductivity of the fuel, which is affected by the thermal gradient in the fuel pellet and the extent of radiation damage. The heat transfer equation is thus nonlinear, and there is a need to couple models of thermal conductivity and mechanical behavior. The multiphysics approach includes models for cladding creep and fission gas release. The models must deal with unusual conditions, such as reactivity insertion accidents, localized stresses due to a chipped fuel pellet, and stress corrosion cracking of cladding. In the case of cladding embrittlement due to hydride formation, the modeling must account for volume changes, misfit strains, stored mechanical energy, changes in the chemical potentials

of the diffusing species, local plastic deformation, and microstructure evolution. The chapter provides a good representation of currently used algorithms and multiphysics codes.

When spatiotemporal information is of the essence for microstructural evolution, phase field models provide a very good alternative to standard methods while still keeping computational cost in check. ▸ Chapter 94, "Phase-Field Modeling of Microstructure Evolution in Nuclear Materials," Ahmed and El-Azab describe the potential of the method to simulate microstructure evolution in nuclear materials, without a doubt one of the most challenging applications that phase field simulations have had to tackle yet. Phase fields evolve according to a free energy functional that represents bulk processes as well as interfaces separating spatial fields characterized by different values of the order parameter. This order parameter can represent different solid phases, clusters of defects, bubbles or voids, different grain orientations, etc. By way of their two main variants, the order parameter can evolve according to an Allen-Cahn or a Cahn-Hilliard equation, both of which originate from rigorous thermodynamic principles that satisfy the first and second laws of thermodynamics. Familiar parameters, such as diffusivities, reaction coefficients, absorption coefficients, etc., are all naturally part of the formulation, dictating the time scale as they do in rate theory or kinetic Monte Carlo models. As well, thanks to its thermodynamic formulation, driving forces such as irradiation, mechanical stresses, and thermal gradients can be incorporated into the formulation in a straightforward manner. The combination of all of these features has allowed researchers to study phenomena such as irradiation-induced segregation and precipitation, formation of voids/bubble formation, nucleation, growth, and coarsening of dislocation loops, both in nuclear fuels and irradiated structural materials. Several limitations still remain such as the theoretical treatment of homogeneous nucleation and artifacts introduced by the smoothening of nominally sharp interfaces. In addition, resolution of interfaces via smooth gradients comes at a price in terms of computational cost, although much progress has been achieved over the last few years by way of application of spectral solvers, finite elements, and others. Finally, phase field simulations must be fed parametric information from lower scales, which calls for a tighter coupling with atomistic simulations before the predictive capabilities of the method can be brought to quantitative standards.

In ▸ Chap. 95, "Thermodynamic Modeling of Nuclear Fuel Materials," McMurray and Besmann describe the framework of classical and computational thermodynamics leading to predictions of phase diagrams for complex systems. The problem of predicting stable phases of fuel materials is extraordinarily complex due to its highly dynamical nature, governed by burnup, transmutation, and temperature, and the high number of components present at any given point. The authors first give an overview of classical thermodynamics, with special emphasis on pointing out the corrections and correlations used in practice to calculate free energies of multicomponent systems. They then introduce the CALPHAD approach, a technique that has enjoyed great success over the last few decades to furnish approximate phase diagrams of systems of – practically – arbitrary complexity. While fuels under service are assumed to not reach equilibrium at any given moment, and thus not

expected to attain any one of the phases present in the predicted phase diagrams, knowledge of these is important to define the driving forces that govern the evolution of these materials under irradiation conditions. Of particular interest is the role played by electronic structure calculations in parameterizing CALPHAD, which is bringing in a level of unprecedented predictive potential to these calculations. The authors discuss unary, binary, and ternary diagrams for selected systems, including uranium-based fuels such as U-C, U-N, and UO_2. In all, thermodynamic modeling is a fundamental piece of radiation damage modeling, particularly as it concerns expected states of solid and liquid fuels to further define and refine other multiphysics models.

3   Summary and Outlook

Treatises discussing the state of the art in the field of multiscale modeling of nuclear materials are a relatively recent occurrence. Their genesis can probably be tied to efforts by ASTM in the early 2000s to compile papers presented at specialized conferences into a single reference bound volume for researchers to use (e.g., Kirk and Natishan, 2003). While specific topics related to nuclear materials modeling were included in Parts A and B of the first edition of the *Handbook of Materials Modeling* (Yip, 2005), where multiscale materials modeling at large was covered, (i) the renewed interest in nuclear energy as a mitigatory source of energy of the effects of climate change and (ii) recognition that the field has grown over the last decade to extents deserving of its own section in the handbook are the main drivers behind the present collection of chapters.

The growth of the field is readily appreciable by looking at the number of workshops, symposia, and conferences available to the nuclear materials modeling community today, ranging from general professional conferences, such as TMS and MRS, to significant portions of modeling-specific international conferences such as MMM, to specialized technical meetings such as NuMat. This interest has been galvanized by the community in the form of publications as well, among which probably stands out *Comprehensive Nuclear Materials* (Konings 2012), where modeling techniques are brought into the limelight of fission and fusion reactor material research and accelerator systems. The contents of this section must be seen as one more effort by the community to organize their thoughts and compile the existing state of the art into a series of chapters that can help researchers mount future efforts or maximize ongoing ones.

While there is considerable potential for applying the methods discussed here to extend the life span of existing nuclear reactors, develop accident-tolerant nuclear fuel and cladding, and design new materials for advanced reactors, the contents of this section also reflect the growth of a field over the last several years and the continuing challenges in an era of exceptional computational resources and method development. These remaining challenges are in fact testimony to the tremendous difficulties associated with the field of irradiation damage and the progress made by the community to solve them. In this sense, this section will give the nonspecialist

reader a broad overview of the catalog of methods available for studying the behavior of irradiated materials across the entire spatiotemporal spectrum, while, for the specialist, it can serve to judiciously select from among the methods at our disposal depending on the relevant conditions and material class and to steer them along defined research directions to augment specific capabilities.

References

Curtin WA, Miller RE (2003) Atomistic/continuum coupling in computational materials science. Modelling Simul Mater Sci Eng 25:071004

Kirk MT, Natishan ME (eds) (2003) Predictive material modeling: combining fundamental physics understanding, computational methods and empirically observed behavior, ASTM STP 1429. ASTM International, West Conshohocken

Konings RJM (ed) (2012) Comprehensive nuclear materials. Elsevier, Amsterdam

Yip S (ed) (2005) Handbook of materials modeling. Springer Science & Business Media, Dordrecht

Zinkle SJ, Busby JT (2009) Structural materials for fission & fusion energy. Mater Today 12(11):12–19

Density Functional Theory Calculations Applied to Nuclear Fuels

86

David Andersson

Contents

Abstract

The goal of nuclear fuel performance modeling is to assess the margin to failure, and it is used by researchers, industry, and regulators to develop fuel designs, optimize reactor operation, and evaluate accident scenarios. Accurate models of material properties and the evolution of material properties as a function time in the reactor (burnup) are critical for accomplishing predictive fuel performance simulations. Historically, these models were fitted to reactor test data and they were empirical in nature, which limited their application to the range of conditions covered by the available test data. By developing mechanistic models that resolve the microstructure, its evolution, and impact on properties,

D. Andersson (✉)
Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: andersson@lanl.gov

2121

predictive simulations can be achieved. This requires a multi-scale modeling framework that at the atomistic scale uses density functional theory (DFT) calculations to determine model parameters governing, for example, point defect and fission gas thermodynamics and kinetics as well as thermal conductivity. This chapter reviews DFT methods applicable to these problems and exemplifies results for primarily UO_2, which is the most common fuel in light water reactors. Additionally, advanced fuels such as U_3Si_2, metallic uranium, and UN are briefly discussed. Challenges, opportunities, and limitations as well as the role of DFT calculations in the multi-scale modeling framework are also covered.

1   Introduction

Density functional theory (DFT) calculations start from a quantum mechanical description of the interaction between electrons, and at the most fundamental level, they predict the density distribution of electrons in a material and the corresponding total energy (Jones 2015). The electron density governs bonding. Many other properties are functions of the electron density and the total energy. They can be attained through perturbations of the material, such as changing the volume or shape of the unit cell (which gives access to equilibrium lattice constants and elastic constants), displacing atoms (which enables calculation of phonons and migration barriers), and removing or adding atoms (which provides defect formation energies and information about phase diagrams). Phonons and defect thermodynamics and kinetics are important properties underlying nuclear fuel performance. DFT calculations have become an important and reliable tool for studying these properties, often leveraged in a multi-scale modeling and simulation framework to connect the fundamental properties accessible from this computationally expensive technique with more complex properties governing materials performance in reactor applications (Tonks et al. 2017; Devanathan et al. 2010).

DFT maps the vast quantum mechanical many-body electron problem in a material into a set of single-particle equations, which typically rely on a periodic representation of the material to further reduce the computational cost (Jones 2015). The key theorems underlying DFT state that the ground-state properties of a many-electron system are uniquely determined by the electron density and that this ground-state density minimizes the total energy functional of the system (Hohenberg and Kohn 1964; Kohn and Sham 1965). In this formalism, the difficult many-body effects that arise from the electron-electron interactions are captured by the so-called exchange-correlation potential (Kohn and Sham 1965). Accurate representation of materials properties requires a good approximation for the many-body electron effects. Over the years, several of the improved predictive capabilities of DFT calculations have been accomplished by developing better approximations for the exchange-correlation potential. The local density approximation (LDA) was the first major step (Kohn and Sham 1965), which was followed by gradient corrected versions (the generalized gradient approximation, GGA) (Langreth and Perdew 1980; Langreth and Mehl 1983; Perdew 1985; Perdew and Wang 1986;

Becke 1988; Perdew et al. 1996). Systems that exhibit strong correlations among the electrons, which include actinide materials and nuclear fuels, often need additional model refinement to describe the physical properties correctly. Popular approaches include the Hubbard U method (adding Hubbard U and J terms to the standard Hamiltonian in order to capture electron localization) (Anisimov et al. 1997), hybrid functionals (incorporating a portion of exact exchange from Hartree-Fock theory) (Heyd et al. 2003), and dynamical mean field theory (DMFT) (Georges et al. 1996). Nuclear fuel performance properties are in most cases closely related to defects, which imply large system sizes (number of atoms) extending beyond the crystallographic unit cell. The added computational expense from having to consider strong correlations in nuclear fuels simulations is a major challenge, which favors the Hubbard U model, often labeled as DFT $+$ U, over other potentially more rigorous but computationally costly methods.

DFT and the application of periodic boundary conditions help to make the many-body electron problem tractable. Further efficiency gains are possible by replacing the explicit treatment of core electrons with a smooth pseudopotential (Vanderbilt 1990) or related ideas implemented in the projector augmented wave (PAW) method (Blöchl 1994; Kresse and Joubert 1999). This is important for nuclear fuels since uranium has 92 electrons. Typical DFT calculations for uranium only treat the outermost 14 valence electrons explicitly, which is a huge benefit since the computational expense scales as $\sim N^3$, where N is the number of electrons, and absolutely necessary in order to be able to study defects.

Compared to empirical or semiempirical force-field descriptions of the interaction between atoms, the main benefit of DFT is much better control of the accuracy and having a predictive capability without any need for fitting to experimental data. For UO_2, one important feature of DFT compared to most semiempirical potentials is the ability to treat varying charge states, that is the formation of U^{3+}, U^{4+}, U^{5+}, and U^{6+} ions. The improved accuracy, unfortunately, comes with a much higher computational cost, which limits the number of atoms that can be studied to somewhere between 100 and 1000 depending on the material and the particular flavor of DFT being applied. Even though fitting to experimental data is not needed in DFT, validation against experiments is key for reliable prediction of materials properties.

This chapter focuses on using DFT calculations to model nuclear fuels and in particular a few key properties that are critically important for the performance of nuclear fuels in reactor applications. The most common type of nuclear fuel in the current fleet of light water reactors (LWRs) is uranium dioxide, UO_2. It is a semiconductor with low thermal conductivity, which is undesirable, but it has excellent performance under irradiation and a high melting temperature. Other options include metallic fuels based on uranium with added alloying elements such as zirconium, which have higher thermal conductivity than UO_2 but a lower melting point. They also tend to be less radiation tolerant, which can lead to significant swelling. Metallic uranium reacts very rapidly with water and nuclear reactors that would use metallic fuel thus rely on a different coolant than standard light water reactors, for example, liquid sodium, in order to avoid uncontrolled reaction with

the coolant in case of a clad breach. UO_2 with zirconium cladding has been a very successful fuel in LWRs, and it has been optimized for performance, reliability, and safety. However, there are fundamental limitations that have motivated interest in alternative advanced fuels and reactor designs, which encompass the metal fuels mentioned above as well as silicon (U_3Si_2)-, nitrogen (UN)-, and carbon (UC)-based uranium compounds. These alternatives address the two major drawbacks of UO_2: low uranium density and poor thermal conductivity. The new fuel concepts and the associated reactor designs also aim at providing additional safety margins, leading to better accident tolerance.

The remainder of this chapter will review the role of DFT in modeling nuclear fuel performance and discuss the unique challenges with applying DFT to nuclear fuels, in particular the impact of strong correlations caused by the uranium 5f electron, followed by case studies of using DFT to investigate point defects, fission gas behavior, and thermal conductivity in nuclear fuels. The main focus of our discussion is on UO_2, but other fuels will also be examined.

2   Nuclear Fuel Performance Modeling, Multi-Scale Simulations, and Role of DFT Calculations

The design and choice of materials for nuclear fuels vary between reactor types, but in general fuel elements with fissile material are enclosed in a protective cladding or in some cases a coating in order to prevent release of radioactive materials to the coolant, with exception for certain molten salt reactors where the coolant itself contains the fuel. The coolant transfers the heat generated in the fuel element from the cladding or coating surface to either a steam generator or directly to a turbine for production of electricity. In light water reactors, the fuel is in the form of small UO_2 pellets stacked on top of each other, with cladding made of a zirconium alloy shaped in the form of a long tube (fuel rod), which are assembled in bundles and placed in the reactor vessel. The goal of nuclear fuel performance modeling is, in essence, to assess the margin to failure of the fuel rod (Rashid et al. 2011; Williamson et al. 2012). This analysis includes both the fuel pellet and the cladding separating the fuel from the coolant. Fuel performance simulations are used by the industry to develop fuel designs, optimize reactor operation, and evaluate accident scenarios. These simulations are also used extensively by regulatory authorities. Fuel performance modeling is performed at the finite element level and resolves the full pin dimensions, with legacy codes being 1D (radial) (Berna et al. 1997) or in some cases 2D (radial and axial) (Rashid et al. 2004) and modern codes, such as the Bison fuel performance code, having the capability to also perform full 3D simulations (Williamson et al. 2012).

Accurate models of material properties and the evolution of material properties as function time in the reactor (burnup) are critical for accomplishing predictive fuel performance simulations. The relation and interlinkage between properties and fuel performance is famously complex (Rashid et al. 2011; Lassmann 1980).

Historically, these models were fitted to reactor test data and they were empirical in nature, which limited their application to the range of conditions covered by the available test data. These models should ideally be mechanistic by capturing the microstructure evolution during reactor operation and the effect of the microstructure on the properties of the fuel (Tonks et al. 2017). The idea is that such models would possess a higher degree of predictive power. Due to the challenge and expense of performing irradiation testing, development and parametrization of said models would almost certainly entail a strong component of multi-scale modeling and simulations. This is especially the case for new advanced fuels that have not been through the same historical testing programs as UO_2. There are several efforts around the world working toward the goal of mechanistic fuel performance modeling supported by multi-scale materials modeling. For example, the approach of Tonks et al. (2017) aims at predicting the microstructure and chemical evolution of fuels under irradiation, as described by a set of state variables, and then connecting the microstructure and chemical state of the fuel to materials properties governing fuel performance at the macroscale. This is in contrast to current empirical models that, instead of using the microstructure as the reference state for material properties, correlate them to burnup. The latter is not a unique measure of the actual microstructure, because all burnups do not lead to the same microstructure state. The proposed state-variable approach requires a multi-scale simulation framework starting from the electronic or atomistic scale, then going to mesoscale simulations involving among others Monte-Carlo techniques, cluster dynamics, rate theory, and phase field simulations to finally inform continuum level finite element simulations of macroscale fuel performance (Tonks et al. 2017). To this date, the most important role of DFT calculations in this multi-scale framework has been to study thermodynamic and kinetic properties of point defects (Andersson et al. 2009, 2011; Dorado et al. 2009; Dorado et al. 2011; Dorado et al. 2012; Crocombette 2012; Vathonne et al. 2014) and fission gas/product atoms (Yun et al. 2008; Nerikar et al. 2009; Andersson et al. 2011, 2014, 2015; Thompson and Wolverton 2011; Thompson and Wolverton 2013; Vathonne et al. 2017) in the fuel as well as chemistry manifested by non-stoichiometry and phase evolution (Andersson et al. 2012, 2013; Yu et al. 2009; Geng et al. 2008; Brincat et al. 2015). Thermal conductivity has also been predicted from DFT (Pang et al. 2013), but so far effects of microstructure and defects on thermal conductivity have only been assessed by empirical potentials.

3   DFT Calculations for Nuclear Fuels

Nuclear fuels rely on fission of ^{235}U to produce heat. From a DFT perspective, uranium atoms require special attention since the 5f electrons are strongly correlated and must in many cases be treated with advanced exchange-correlation schemes to capture the properties correctly. For example, both LDA and GGA predict UO_2 to be a metal instead of a semiconductor. This can be corrected by applying, e.g., Hubbard

U (Dudarev et al. 1997), hybrid functional (Kudin et al. 2002), or DMFT (Yin and Savrasov 2008) models. Most existing studies on defects in UO_2 use the DFT + U methodology due to its higher computational efficiency. For UO_2, the values of the U and J parameters, representing the onsite Coulomb and exchange parameters, respectively, are taken from experiments (Dudarev et al. 1997). The DFT + U method is available in many different DFT codes. The Vienna ab initio simulation package (VASP) (Kresse and Furthmüller 1996) is the most popular choice in the literature. The DFT + U method has been shown to reproduce the measured band gap of UO_2 rather accurately, but, as expected, that does not necessarily imply that all thermodynamic and thermophysical properties are perfectly captured. As an example, the lattice parameter is pushed up above the experimental value by GGA + U (Dorado et al. 2009; Sanati et al. 2011), while LDA + U predicts a smaller lattice constant closer to the experimental value (Wang et al. 2013; Sanati et al. 2011) but instead slightly overestimates the elastic constants (Sanati et al. 2011). The tendency of LDA(+U) to overbind and GGA(+U) to push in the opposite direction is well established in DFT. The difference in the predicted lattice constant between GGA + U and LDA + U carries over to defect migration barriers, with GGA + U predicting lower barriers than LDA + U, which has been hypothesized to be a result of the increased space for the migrating ion due to the larger lattice constant predicted by GGA + U (Andersson et al. 2011; Dorado et al. 2012). Even though trends in thermodynamic properties are similar between LDA + U and GGA + U, there are certainly quantitative differences, and the agreement with experimental data is in general qualitative rather than quantitative.

UO_2 is antiferromagnetic below the Néel temperature of 30.8 K. The magnetic structure is non-collinear with a **3 k** pattern and can only be reproduced if spin-orbit coupling is included in the DFT + U calculations (Laskowski et al. 2004). This can be done for perfect UO_2, but it presents severe computational challenges for defect studies involving supercells. Consequently, it is common practice to approximate the ground-state **3 k** structure of UO_2 with a simpler **1 k** structure where the magnetic moments are antiferromagnetically aligned on the (100) planes of the fluorite structure. Of course, at temperatures of relevance to reactor applications, UO_2 is paramagnetic. This state can be approximated by, for example, applying the SQS method to randomly distribute the localized magnetic moments in a supercell (Dorado and Garcia 2013). However, it has been shown that the **1 k** AFM structure is a good approximation of the paramagnetic state with respective to defect properties and that assumption is implicit in a majority of UO_2 defect studies (Dorado and Garcia 2013).

One issue that was revealed in early studies on UO_2 employing the DFT + U method was the propensity to obtain metastable electronic solutions due to different orbital occupations for the uranium 5f electrons (Dorado et al. 2009). This complicated prediction of accurate defect properties and created a surprisingly large spread among defect energies reported in the literature. Several approaches have been developed to control the electronic solution and increase the likelihood of

reaching the lowest energy ground-state orbital configuration. The two most popular choices are probably the occupation matrix control (Dorado et al. 2009) and the U ramping methods (Meredig et al. 2010). The first relies on prescribing the initial orbital occupation matrix, which helps the minimization algorithm to find the correct solution. This requires a search over the full range of possible occupation matrices. For structures with low symmetry that can be a laborious task. Alternatively, the U ramping method relies on a slow increase of the U value from zero to the nominal value for UO_2, which has been shown to increase the chances of finding the correct solution. Even though this does not guarantee that the lowest energy solution is obtained, practical experience has shown that the method is quite reliable.

U_3Si_2 is a metal, which is captured by regular LDA/GGA. However, there is another important issue for U_3Si_2. Standard GGA predicts the ground-state U_3Si_2 crystal structure to be thermodynamically as well as dynamically unstable with respect to alternative phases identified by combining DFT calculations with a structure optimization algorithm (Noordhoek et al. 2016). None of the new phases have been observed in experiments and they are likely artifacts. This behavior can be rectified by introducing a small Hubbard U parameter, although this gives rise to ferromagnetism that is not reproduced in experiments (Noordhoek et al. 2016). However, existing studies have not fully explored the competition between ferromagnetism and paramagnetic structures based on disordered local moments. Even though metastable electronic solutions exist also in the U-Si system, the resulting variation in energy is much smaller than for UO_2, and existing studies do not apply any of the special methods developed for UO_2 to control the orbital occupation.

When used as a nuclear fuel, the high-temperature bcc γ phase of uranium is most relevant. It may also be stabilized by alloying. Metallic uranium exhibits complicated bonding with many allotropes as function of temperature. Most studies apply regular GGA for this material, and the need for schemes accounting for strong correlations is less obvious than for UO_2 and U_3Si_2. However, there are still challenges that complicate DFT studies of this material. Similar to the bcc phase for many transition metals, DFT predicts the bcc phase of uranium to be mechanically unstable due to a negative shear elastic constant at zero Kelvin (Beeler et al. 2010; Beeler et al. 2013). This implies that DFT cannot easily be used for studying the properties of the bcc phase, in particular studies of defects are problematic since they will induce sufficient symmetry breaking for the lattice to spontaneously relax into a new or severely distorted structure. The bcc phase is stabilized by anharmonicity at high temperature. Methods have been developed to account for this physics, mainly in transition metals (Souvatzis et al. 2008). The basic idea is to superimpose the phonons that stabilize the bcc phase at high temperature on the static structure and thus capture the anharmonic effects leading to stability at high temperature (Söderlind et al. 2012). Another option is to run molecular dynamics simulations at high temperature (Hellman et al. 2011), which automatically accounts for the anharmonicity and stabilizes the bcc phase.

4   Thermodynamic and Kinetic Properties of Point Defects in Nuclear Fuels and their Behavior under Irradiation

The behavior of point defects in nuclear fuels governs some of the most important fuel performance properties. First, radiation damage creates a cascade of interstitials and vacancies, some of which recombine with each other and some of which agglomerate into voids and interstitial loops. The latter causes fuel swelling and may lead to the fuel element coming into contact with the cladding, thus impacting mechanical performance. The presence of point defects also reduces the thermal conductivity of the fuel. Moreover, point defects govern diffusion and thus impact any diffusion-controlled phenomena in the fuel, such as creep, fission gas retention/release, and sintering. Consequently, it is essential to be able to predict the equilibrium concentration of point defects as well as their evolution under irradiation.

Prediction of non-stoichiometry and diffusion rates requires a point defect model, and the behavior under irradiation requires a model that handles defect creation and annihilation, such as rate theory or cluster dynamics. The main role for DFT calculations is to provide input to these higher-level models. There are examples of this work flow in the literature, most of them for UO_2 (Andersson et al. 2014, 2015). We refer to the literature for a discussion of the higher-level models and focus our review on the underlying DFT calculations.

4.1   Point Defects under Thermal Equilibrium

In thermal equilibrium the concentration of point defects is determined by the thermodynamic properties of vacancies, interstitials, and possibly anti-site defects (not important for UO_2, though). For semiconductors such as UO_2, it is also necessary to account for the charge state of each defect (Nerikar et al. 2009a; Andersson et al. 2011; Crocombette 2012), which for $UO_{2\pm x}$ corresponds to the number of bound holes and electrons, and ensure that overall charge neutrality is maintained. A thermodynamic model connects the defect properties to concentration. UO_2 can maintain significant non-stoichiometry and it should be expressed as $UO_{2\pm x}$ (Guéneau et al. 2002), with x ranging from slightly negative values to 0.25. The exact value of x, as governed by the oxygen chemical potential or partial pressure, is important since it directly couples to the concentration of oxygen and uranium defects. The range of non-stoichiometry is much smaller for UN and U_3Si_2.

DFT calculations have been used extensively to study properties governing the equilibrium thermodynamics in nuclear fuels, in particular for UO_2. The most straightforward quantities to calculate are defect formation energies and binding energies of small defect clusters. The typical approach applies a supercell setup based on at least a $2\times2\times2$ expansion of the fluorite UO_2 unit cell. Oxygen and uranium vacancies and interstitials are introduced into the lattice, and the structure is then allowed to relax in order to minimize the energy. Most calculations also

relax the volume and shape of the supercell, consistent with zero pressure boundary conditions. Adding an oxygen interstitial to a supercell would create two localized holes in order to maintain overall charge neutrality in the cell, because the oxygen interstitial forms an O^{2-} ion. However, in many cases we are more interested in the oxygen interstitial in itself without any bound holes, which corresponds to a charged defect. For example, the oxygen Frenkel (creation of one oxygen vacancy and interstitial) and Schottky (creation of two oxygen vacancies and one uranium vacancy) defect reactions involve nominally charged defects and should be calculated using charged supercell to avoid formation of holes and electrons. For an oxygen interstitial, the charged defect can be studied by adding two electrons to the supercell according to the standard approach of applying a compensating uniform background charge, which fills the two holes that exist in the neutral calculation. It is important to apply corrections for the long-range interaction between periodic images of the defect and to align the electrostatic potential between the reference defect-free structure and the charged supercell containing the defect. Approaches to handle these corrections have been discussed extensively in the literature on semiconductors (Lany and Zunger 2008; Taylor and Bruneval 2011; Freysoldt et al. 2009, 2014). The calculation of the image charge correction at the monopole level is well-defined, but there are still open questions about how to best apply the potential alignment correction, especially for defect structures that include volume relaxation (Bruneval et al. 2015). For UO_2 close to perfect stoichiometry, all defects are predicted to have formal charges, i.e., an oxygen interstitial has a charge of -2, but if the material deviates from the stoichiometric point or equivalently the Fermi level moves from the mid-gap position, the defects may take nonformal charge states (Vathonne et al. 2014). Other charged defects are studied following the same approach as outlined for oxygen interstitials above.

In agreement with experiments (Guéneau et al. 2002), DFT studies of the UO_2 point defect thermodynamics have shown it is much easier to oxidize (form UO_{2+x}) UO_2 than to reduce (form UO_{2-x}) it (Andersson et al. 2011; Andersson et al. 2014; Cooper et al. 2018). An extra complication for UO_2 is that many point defects, in particular oxygen interstitials in the UO_{2+x} range, have a strong tendency to form clusters, which culminates in the formation of ordered phases at the U_4O_{9-y} and U_3O_7 stoichiometries. These reactions have been investigated by DFT and also in experiments (Andersson et al. 2013; Desgranges et al. 2009, 2011). More work is needed to reconcile the detailed experimental observations with predictions by DFT. Even though oxygen defects are dominant in $UO_{2\pm x}$, U defects are very important for fuel performance since they diffuse slower and thus constitute the rate-limiting step for mass transport.

Defect energies can be obtained from static single-point supercell calculations, while entropies require calculation of phonons, which implies calculation of the dynamical matrix from forces on atoms displaced from their equilibrium positions or from perturbation theory. The finite displacement method has been used to calculate phonon dispersion in perfect UO_2 based on the GGA and GGA $+ U$ for the exchange-correlation potential (Sanati et al. 2011; Pang et al. 2013; Yin and Savrasov 2008; Wang et al. 2013). The predictions using GGA $+ U$ are in

good agreement with inelastic neutron experiments. The low symmetry of defect supercells requires many more displacements and force calculations to obtain the phonon dispersion. This added computational complexity together with the challenge of using DFT $+ U$ in phonon calculations has so far precluded defect entropies from being calculated. Existing studies of defect thermodynamics involving the full free energy have relied on a combination of DFT calculations and empirical potentials to access entropies (Andersson et al. 2014). The main conclusions from studies of the UO_2 point defect thermodynamics are that oxygen defects are more prevalent than uranium defects; that UO_2 is much easier to oxidize (UO_{2+x}) than to reduce (UO_{2-x}) (Cooper et al. 2018), which favors oxygen interstitials and uranium vacancies; and that the oxygen interstitials in UO_{2+x} has a strong tendency to form clusters (Andersson et al. 2013). Charge compensation for oxygen interstitials in UO_{2+x} is predicted to occur by U^{5+} ions (holes) (Cooper et al. 2018), and U^{6+} ions are only form in U_3O_8 and compounds with higher oxygen content (Andersson et al. 2013), which do not form the fluorite structure.

The discussion about point defects under thermal equilibrium conditions above refers to UO_2, but similar work has been performed for other fuels, including U_3Si_2 (Middleburgh et al. 2016), UN (Claisse et al. 2016a, b; Mei et al. 2013), and metal fuels (Beeler et al. 2013; Beeler et al. 2010). U_3Si_2 does not exhibit significant non-stoichiometry, which could be a drawback since the chemical evolution happening due to fission events may result in formation of new U-Si phases with different volume. The fact that UO_{2+x} stays in the fluorite structure with very little volume change is beneficial for fuel performance.

4.2 Point Defects under Irradiation

Under irradiation, defects (vacancies and interstitials) are created by damage events and may annihilate by recombining with each other or with various sinks, such as surfaces, voids, grain boundaries, and dislocations. The balance between defect sources and the reaction rate with sinks governs the irradiation-induced concentration of defects. The reaction rate is a function of the sink strength, which can be a complex function of several parameters of which concentration is always going to be important, and the diffusion rate of defects. At high temperatures the rate of diffusion is sufficiently fast to recover the thermal equilibrium concentrations of point defects. In order to understand the materials response and properties under irradiation below this temperature, the kinetic properties of point defects must be calculated.

DFT calculations are a valuable tool also for the kinetic problem. As for thermodynamics, a supercell setup is used to describe each point defect and the barrier is calculated by using the nudged elastic band method or a variation of this method (Millis et al. 1995; Henkelman and Jonsson 2000; Henkelman et al. 2000). Diffusion of oxygen defects is much faster than for uranium defects. The oxygen defect barriers are in the range of 0.5–1 eV (Andersson et al. 2009, 2012, Dorado et al. 2011), while those for uranium defects are roughly 4 eV or

higher (Dorado et al. 2012), with exception for some clusters of uranium vacancies (Andersson et al. 2014). The fast diffusion rate of oxygen defects implies that their concentrations will be close to thermodynamic equilibrium at reactor operating conditions, while the concentration of uranium defects is increased at all but the very highest temperatures. Interstitial diffusion of both oxygen and uranium ions occurs by interstitialcy mechanisms, which means that, rather than performing a direct jump to a vacant nearest neighbor site, the interstitial atom kicks out a lattice atom into a neighboring interstitial position. Diffusion of uranium vacancies follows a slightly curved path to a nearest neighbor site (Goyal et al. 2015). The curved path reduces the barrier significantly by extending the distance to nearest neighbor ions at the saddle point. The curved mechanism for uranium vacancies also means that clusters of two uranium vacancies may diffuse with a lower barrier than single vacancies, because in nuclear fuels a uranium ion in the saddle point position experiences less repulsion from the nearest neighbors when a second vacancy is nearby.

In addition to the barrier for migration, the diffusion rate includes a factor that corresponds to the attempt frequency, which is related to the ratio between the vibrational frequency for the starting and saddle points. As for the vibrational entropy, the attempt frequency for diffusion in UO_2 has to date only been obtained from empirical potentials (Andersson et al. 2014).

Self-diffusion of uranium and oxygen is determined by the concentration and mobility of point defects. The concentrations are given by the thermodynamic properties discussed in the previous section and the mobilities are obtained according to the procedure reviewed above. For UO_2, this results in uranium vacancies being responsible for diffusion in nearly stoichiometric UO_2 and hyper-stoichiometric UO_{2+x} (Dorado et al. 2011; Andersson et al. 2011), while uranium vacancies with bound oxygen vacancies dominate diffusion in slightly hypostoichiometric UO_{2-x} (Andersson et al. 2011). The concentration of uranium interstitials is too low according to DFT, but experiments have indicated that there could be a contribution for highly substoichiometric samples. Correspondingly, oxygen self-diffusion occurs by interstitials or clusters of interstitials in UO_{2+x} (Andersson et al. 2009, 2012; Dorado et al. 2011), while oxygen vacancies are dominant in UO_{2-x}. Both mechanisms contribute in nearly stoichiometric UO_2. It is interesting that even the large oxygen interstitial clusters can migrate with a low barrier and contribute to diffusion in UO_{2+x} (Andersson et al. 2012).

5   Fission Gas Diffusion in Nuclear Fuels

Fissioning of ^{235}U creates a number of fission gases and fission products, of which Xe is the most important from a fuel performance perspective. Xe atoms are not soluble in any nuclear fuel and thus want to escape from the lattice. This can occur by forming bubbles within the grains or by escaping to the grain boundaries and precipitating bubbles that decorate the boundaries. Finally, the gas may reach a free surface and be released from the fuel. This step typically occurs by the

concentration of gas at grain boundaries increasing until a bubble percolation threshold is reached. When this happens, the gas may reach an already open grain triple junction that vents to a crack or open surface, causing release. Bubbles within grains are usually limited to a few nanometers in size due to continuous resolution from fission fragments, while grain boundary bubbles can grow much larger. If the gas is retained in the fuel element as small intragranular bubbles or, in particular, large grain boundary bubbles, the fuel swells, and if it is released, the pressure increases in the plenum. If there is a clad breach, the gas may escape to the outside environment. Consequently, it is important to be able to predict the rate of gas release. This is a notoriously difficult problem due to the coupling to chemistry, defect dynamics, and microstructure evolution. It is one of the most uncertain and impactful parameters in nuclear fuel performance models. The models are most advanced for UO_2. Even though they are based on years of testing, there is room for improvements. The underlying diffusion mechanisms for Xe atoms in bulk UO_2 are still debated. Atomistic simulations and DFT calculations have been used to develop mechanistic release models for UO_2 (Andersson et al. 2014). For new advanced fuels, fission gas release/retention models are rudimentary because of the lack of test data. Simulations will play key role in the development of new gas release models (Miao et al. 2017; Miao et al. 2018).

The rate of fission gas release is controlled by the steps outlined above. Of these, the first step corresponding to diffusion of single gas atoms through the bulk grain to intragranular bubbles or to the grain boundaries lends itself best to investigation by DFT calculations, simply because the number of atoms required to describe grain boundaries with fission gas atoms in the second step is too large for handling in DFT. Since bulk diffusion of single gas atoms is the slowest step in the fission gas evolution process, this sets the timescale for fission gas release.

The diffusion rate of fission gas atoms is determined by the location of gas atoms in the lattice and how they interact with thermal- and irradiation-induced defects. DFT and empirical potentials have been used to investigate the preferred trap site of Xe atoms in the UO_2 lattice (Andersson et al. 2011, 2014; Ball and Grimes 1990; Grimes and Catlow 1991; Thompson and Wolverton 2011). The calculations are performed by adding a Xe atom to a vacant uranium or oxygen site (or a combination of uranium and oxygen vacancies) or to an interstitial site in a supercell. The stability is then assessed from the energy required to add the Xe atom to this supercell and the energy associated with creating the trap site (vacancies) that the Xe atom occupies. That is, it is the sum of the trap site formation energy and the incorporation energy of Xe into this site, which is referred to as the solution energy. Several studies employing DFT and empirical potentials have shown that large fission gas atoms such as Xe prefer uranium vacancy trap sites (Andersson et al. 2011, 2014; Ball and Grimes 1990; Grimes and Catlow 1991; Nerikar et al. 2009b). These vacancy trap sites may also be coordinated with oxygen vacancies. The number of oxygen vacancies bound to the trap site is a function of the total concentration of oxygen vacancies in the material, which is governed by the oxygen chemical potential or non-stoichiometry, represented as the x in $UO_{2\pm x}$. For nearly stoichiometric UO_2 DFT studies have come to the conclusion that the preferred

trap site is a divacancy (one uranium and one oxygen vacancy), while oxygen-poor conditions (UO_{2-x}) favor a bound Schottky defect (one uranium and two oxygen vacancies) and oxygen-rich conditions favor a single uranium vacancy as the trap site (Andersson et al. 2014). Because UO_2 is a semiconductor, the preferred charge state of each Xe trap site, or equivalently the number of bound holes and electrons, must be calculated. This is done by following the same methodology as for point defects described in the preceding section. For UO_{2-x} and UO_2, all trap sites are in their nominal charge states (no bound holes) at the temperatures of interest for nuclear fuel applications (Andersson et al. 2014), but a transition to partial charges (bound holes) may occur for oxygen-rich conditions (UO_{2+x}) (Andersson et al. 2014). This transition is consistent with the Fermi-level shifting from the top half of the band gap in UO_{2-x} close to the valence band edge in UO_{2+x}. A complete analysis of the preferred trap site for Xe atoms must involve both the energy and the entropy contributions, but just as for point defects, DFT calculations of trap site entropies have so far been deemed too costly or too difficult to pursue. Existing studies of the full free energy have relied on a combination of DFT calculations and calculations based on empirical potentials for the entropies (Andersson et al. 2014). As further discussed below, at low temperatures the high binding energies of a uranium vacancy to the single uranium vacancy trap sites for Xe cause a transition to the uranium divacancy (potentially with additional oxygen vacancies) as the preferred trap site (Andersson et al. 2014).

Diffusion of Xe in an interstitial trap site occurs with a low barrier by temporarily displacing a lattice oxygen ion into an interstitial site (Liu and Uberuaga 2011). However, the Xe interstitial position is a high-energy state compared to vacancy trap sites and that leads to a negligible contribution to diffusion under thermal equilibrium conditions and probably also under irradiation.

Diffusion involving vacancy trap sites occurs by binding a second assisting uranium vacancy (Andersson et al. 2011, 2014; Ball and Grimes 1990; Govers et al. 2008; Thompson and Wolverton 2013, Vathonne et al. 2017) and potentially additional uranium vacancies under irradiation. The binding energies are straightforward to calculate from DFT using supercells, but care must be applied to treat the charge state of each cluster correctly. The corresponding entropies are more challenging for DFT calculations and have so far only been investigated using calculations based on empirical potentials. The second assisting uranium vacancy has a negative binding energy (attractive) for all clusters and in many cases forms a very stable cluster with the original trap site (Andersson et al. 2011, 2014). Diffusion of these clusters involves two steps. In the first Xe moves between the two uranium vacancies constituting the mobile cluster, which is a low-barrier event and in most cases the Xe atom even occupies the central position of the cluster, in-between the two constituent uranium vacancies. The second step is rate-limiting and it involves migration of one of the vacancies to another nearest neighbor position. This step can either occur by a direct mechanism involving a single migration jump to another nearest neighbor site or by first jumping to a second nearest neighbor position and then returning to another nearest neighbor site (Thompson and Wolverton 2013; Andersson et al. 2014). The barriers are obtained from nudged elastic band

calculations. As for diffusion of point defects, the attempt frequencies have not yet been calculated from DFT but rather rely on relating the phonon frequencies at the saddle point and initial position obtained from empirical potentials. The resulting diffusivity can be expressed by an Arrhenius relation with an activation energy given by the sum of the effective vacancy formation energy, the binding energy to the trap site, and the cluster migration barrier. If the binding energy is very strong for a particular cluster and that cluster dominates over the corresponding trap site, the diffusivity will simply be given by the mobility of that cluster. This occurs in UO_2 at low temperatures. The total Xe diffusion rate is the sum of the individual cluster diffusivities weighted by their relative concentration in lattice.

The resulting Xe diffusivities increase going from UO_{2-x} to UO_2 and UO_{2+x} (Andersson et al. 2011, 2014; Miekeley and Felix 1972). This is a consequence of both an increase of the concentration of uranium vacancies and a reduction of the migration barrier of clusters involving fewer oxygen vacancies, though the latter effect is at least partially counteracted by a reduction in the binding energy of vacancies to the trap site.

It is challenging to measure the fission gas diffusion rate experimentally. The reasons are twofold. First, experiments primarily rely on measuring the gas release in postirradiation annealing experiments, which involves more uncertainty than, for example, tracer diffusion experiments, and, second, the diffusivity is very sensitive to non-stoichiometry and thus the experimental conditions and how the oxygen partial pressure is controlled in the experiment (Lawrence 1978). Xe diffusion studies on nominally stoichiometric UO_2 yield activation energies from 2.87 eV to 3.95 eV and pre-exponential factors from 5×10^{-4} m^2/s to 2.90×10^{-12} m^2/s (Andersson et al. 2014). Typically, higher activation energy is compensated by a higher pre-exponential factor, which is expected since the stoichiometry is maintained close to UO_2 for all cases. In order to systematically explore the variation of the diffusivity with non-stoichiometry, experiments have been performed under the same initial irradiation conditions for UO_{2-x}, UO_2, and UO_{2+x} samples (Miekeley and Felix 1972). These experiments established clear trends for the three ranges, which provide data suitable for validation of DFT and atomic-scale simulations. Calculations qualitatively capture the variation in the diffusivity as function of non-stoichiometry (Andersson et al. 2011, 2014), though there are still questions as to the quantitative agreement. The lack of quantitative agreement is at least partially explained by the approximations applied in the DFT methodology, e.g., the choice of exchange-correlation model and the assumption that properties acquired at zero kelvin can be extrapolated to high temperatures where experiments are performed.

In fission gas release models, the intrinsic diffusion coefficient discussed so far is only valid at high temperature (above approximately 1600 K) (Turnbull et al. 1982), where the effects of irradiation are completely annealed. At lower temperatures, the increased concentration of point defects must be accounted for. This gives rise to radiation-enhanced diffusion dominating over intrinsic diffusion in an intermediate temperature range (from about 1300 to 1600 K) (Turnbull et al. 1982). The

activation energy observed in experiments is much lower than in the intrinsic regime, and it has a very low pre-exponential factor. The mechanism responsible for this diffusion rate is presently not known. DFT or atomistic simulations based on empirical potentials are not able to reproduce this behavior, as the mobility predicted by DFT and empirical potential calculations for the Xe defect clusters involving up to two uranium vacancies is too low (Andersson et al. 2014). It is possible that this discrepancy may be a consequence of neglecting the mobility of extended Xe-vacancy clusters, i.e., clusters involving more than one bound uranium vacancy as well as additional oxygen vacancies or possibly Xe interstitials which are known to be highly mobile but also energetically very unfavorable. The hypothesis concerning extended Xe-vacancy clusters implies that, in order to develop accurate models of Xe diffusion under irradiation, it is critical to understand the interaction between Xe and vacancies leading to the formation of extended clusters with potentially increased mobility.

6   Thermal Conductivity of Nuclear Fuels

Thermal conductivity is one of the most important properties for nuclear fuel performance. The most challenging part of the problem is to predict the degradation as a function of damage and chemical evolution in the fuel. For an insulator or semiconductor like UO_2, this implies calculating phonons and phonon scattering parameters. DFT has been used to estimate the conductivity of pure UO_2 (Pang et al. 2013), which leads to a slight overestimation that is highest at low to intermediate temperatures. Phonon-phonon interaction is the main scattering mechanism according to these simulations. There are indications that the discrepancy between simulations and experiments for the thermal conductivity is due to the lack of a mechanism for phonons to be scattered by spins or electronic excitations in the DFT simulations (Gofryk et al. 2014). The importance of spin scattering is emphasized by comparing the experimental thermal conductivity of ThO_2 and UO_2 (Gofryk et al. 2014). Even though the DFT phonon calculations may include spins, capturing this mechanism would require solving the dynamic evolution of spin configurations coupled to the lattice dynamics. This problem has been addressed for a few materials by coupling molecular dynamics simulations based on empirical potentials with a spin Hamiltonian (Ma et al. 2008; Ma and Dudarev 2014), but this treatment has not been applied to UO_2 and extending it involves several additional challenges.

The thermal conductivity of U_3Si_2 and UN fuels is both metallic, and estimating the thermal conductivity from DFT entails solving the transport problem for electrons, rather than for phonons as in UO_2. This has not been attempted in the literature, but the theory for addressing the problem is in place, even though the f electrons in uranium may present some obstacles for the band structure calculations needed to solve the Boltzmann transport problem.

7   Summary and Outlook

We have reviewed applications of DFT calculations to study nuclear fuels and the role of DFT calculations in multi-scale modeling of nuclear fuel performance. UO_2 fuel for light water reactors was emphasized, because of the large body of work performed on this material. Advanced fuels such as U_3Si_2 and UN were also briefly discussed. The uranium 5f electrons are strongly correlated, which leads to regular GGA failing to capture the semiconducting nature of UO_2. This can be rectified by applying exchange-correlation schemes designed to treat this type of material. The Hubbard U model is the most popular choice for defect studies due to its computational efficiency. Other more cutting-edge approaches to treat strong correlations, for example, DMFT, will continue to evolve and be applied to UO_2 and other fuels but may be too expensive for the type of calculations needed to assess nuclear fuel performance properties. Examples of studying point defect thermodynamics and kinetics as well as the behavior of fission gas and thermal conductivity were given. Present DFT calculations of point defect properties only address the energy contribution to the free energies, and the entropies are obtained from empirical potentials. Future work should remove the reliance on empirical potentials for calculating defect entropies. We also expect DFT calculations to be applied or continue to be applied to new advanced fuels, including metal fuel, U_3Si_2 and UN. Further advancements are expected for the multi-scale modeling frameworks, which should increase the use for DFT calculations and thus their value. Experimental validation will be important to assess uncertainty in the calculations. Uncertainty quantification should enable evaluation of how the uncertainty at the electronic and atomistic levels translates through a multi-scale simulation framework to impact fuel performance simulations.

References

Andersson DA, Lezama J, Uberuaga BP, Conradson SD (2009) Cooperativity among defect sites in AO_{2+x} and A_4O_9 (A=U,Np,Pu): density functional calculations. Phys Rev B 79:024110

Andersson DA, Uberuaga BP, Nerikar PV, Unal C, Stanek CR (2011) U and Xe transport in $UO_{2\pm x}$: density functional theory calculations. Phys Rev B 84:054105

Andersson DA, Espinosa-Faller FJ, Uberuaga BP, Conradson SD (2012) Stability and migration of large oxygen clusters in UO_{2+x}: density functional theory calculations. J Chem Phys 136:234702

Andersson DA, Baldinozzi G, Desgranges L, Conradson DR, Conradson SD (2013) Density functional theory calculations of UO_2 oxidation: evolution of UO_{2+x}, U_4O_{9-y}, U_3O_7, and U_3O_8. Inorg Chem 52:2769–2778

Andersson DA et al (2014) Atomistic modeling of intrinsic and radiation-enhanced fission gas (Xe) diffusion in $UO_{2\pm x}$: implications for nuclear fuel performance modelling. J Nucl Mater 451:225–242

Andersson DA, Tonks MR, Casillas L, Vyas S, Nerikar P, Uberuaga BP, Stanek CR (2015) Multiscale simulation of xenon diffusion and grain boundary segregation in UO_2. J Nucl Mater 462:15–25

Anisimov VI, Aryasetiawan F, Lichtenstein AI (1997) First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA+U method. J Phys Condens Matter 9:767–808

Ball RGJ, Grimes RW (1990) Diffusion of Xe in UO_2. Chem Soc Faraday Trans 86:1257–1261

Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 38:3098–3100

Beeler B, Good B, Rashkeev S, Deo C, Baskes M, Okuniewski M (2010) First principles calculations for defects in U. J Phys: Cond Matter 22:505703

Beeler B, Deo C, Baskes M, Okuniewski M (2013) First principles calculations of the structure and elastic constants of α, β and γ uranium. J Nucl Mater 433:143–151

Berna GA, Beyer CE, Davis KL, Lanning DD (1997) FRAPCON–3: a computer code for the calculation of steady-state, thermal-mechanical behavior of oxide fuel rods for high burnup. Technical report NUREG/CR-6534, vol 2, PNNL–11513

Blöchl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953–17979

Brincat NA, Molinari M, Parkera SC, Allen GC, Storr MT (2015) Computer simulation of defect clusters in UO_2 and their dependence on composition. J Nucl Mater 456:329–333

Bruneval F, Varvenne C, Crocombette JP, Clouet E (2015) Pressure, relaxation volume, and elastic interactions in charged simulation cells. Phys Rev B 91:024107

Claisse A, Schuler T, Lopes DA, Olsson P (2016a) Transport properties in dilute UN(X) solid solutions (X=Xe,Kr). Phys Rev B 94:174302

Claisse A, Klipfel M, Lindbom N, Freyss M, Olsson P (2016b) GGA+U study of uranium mononitride: a comparison of the U-ramping and occupation matrix schemes and incorporation energies of fission products. J Nucl Mater 478:119–124

Cooper MWD, Murphy ST, Andersson DA (2018) The defect chemistry of $UO_{2\pm x}$ from atomistic simulations. J Nucl Mater 504:251–260. https://doi.org/10.1016/j.jnucmat.2018.02.034

Crocombette JP (2012) Influence of charge states on energies of point defects and clusters in uranium dioxide. Phys Rev B 85:144101

Desgranges L, Baldinozzi G, Rousseau G, Nièpce JC, Calvarin G (2009) Neutron diffraction study of the in situ oxidation of UO_2. Inorg Chem 48:7585–7592

Desgranges L, Baldinozzi G, Siméone D, Fischer HE (2011) Refinement of the α-U_4O_9 crystalline structure: new insight into the $U_4O_9 \rightarrow U_3O_8$ transformation. Inorg Chem 50:6146–6151

Devanathan R et al (2010) Modeling and simulation of nuclear fuel materials. Energy Environ Sci 3:1406–1426

Dorado B, Amadon B, Freyss M, Bertolus M (2009) DFT+U calculations of the ground state and metastable states of uranium dioxide. Phys Rev B 79:235125

Dorado B et al (2011) First-principles calculation and experimental study of oxygen diffusion in uranium dioxide. Phys Rev B 83:035124

Dorado B, Andersson DA, Stanek CR, Bertolus M, Uberuaga BP, Martin G, Freyss M, Garcia P (2012) First-principles calculations of uranium diffusion in uranium dioxide. Phys Rev B 86:035110

Dorado B, Garcia P (2013) First-principles DFT + U modeling of actinide-based alloys: application to paramagnetic phases of UO_2 and (U,Pu) mixed oxides. Phys Rev B 87:195139

Dudarev SL, Nguyen Manh D, Sutton AP (1997) Effect of Mott-Hubbard correlations on the electronic structure and structural stability of uranium dioxide. Phil Mag B 75:613–628

Freysoldt C, Neugebauer J, Van de Walle CG (2009) Fully ab initio finite-size corrections for charged-defect supercell calculations. Phys Rev Lett 102:016402

Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86:253–305

Geng HY, Chen Y, Kaneta Y, Iwasawa M, Ohnuma T, Kinoshita M (2008) Point defects and clustering in uranium dioxide by LSDA+U calculations. Phys Rev B 77:104120

Georges A, Kotliar G, Krauth W, Rozenberg MJ (1996) Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. Rev Mod Phys 68:13

Gofryk K et al (2014) Anisotropic thermal conductivity in uranium dioxide. Nature Comm 5:4551

Govers K, Lemehov S, Verwerft M (2008) In-pile Xe diffusion coefficient in UO_2 determined from the modeling of intragranular bubble growth and destruction under irradiation. J Nucl Mater 374:461–472

Goyal A, Phillpot SR, Subramanian G, Andersson DA, Stanek CR, Uberuaga BP (2015) Impact of homogeneous strain on uranium vacancy diffusion in uranium dioxide. Phys Rev B 91:094103

Grimes RW, Catlow CRA (1991) The stability of fission products in uranium dioxide. Philos Trans R Soc London Ser A 335:609–634

Guéneau C, Baichi M, Labroche D, Chatillon C, Sundman B (2002) Thermodynamic assessment of the uranium–oxygen system. J Nucl Mater 304:161–175

Hellman O, Abrikosov IA, Simak SI (2011) Lattice dynamics of anharmonic solids from first principles. Phys Rev B 84:180301(R)

Henkelman G, Jónsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. J Chem Phys 113:9978–9985

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113:9901–9904

Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened coulomb potential. J Chem Phys 118:8207–8215

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871

Jones RO (2015) Density functional theory: its origins, rise to prominence, and future. Rev Mod Phys 87:897

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138

Kresse G, Joubert J (1999) From ultrasoft pseudopotentials to the projector augmented wave method. Phys Rev B 59:1758–1775

Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186

Kudin KN, Scuseria GE, Martin RL (2002) Hybrid density-functional theory and the insulating gap of UO_2. Phys Rev Lett 89:266402

Langreth DC, Perdew JP (1980) Theory of nonuniform electronic systems. I. Analysis of the gradient approximation and a generalization that works. Phys Rev B 21:5469–5493

Langreth DC, Mehl MJ (1983) Beyond the local-density approximation in calculations of ground-state electronic properties. Phys Rev B 28:1809–1834

Lany S, Zunger A (2008) Assessment of correction methods for the band-gap problem and for finite-size effects in supercell defect calculations: case studies for ZnO and GaAs. Phys Rev B 78:235104

Laskowski R, Madsen GKH, Blaha P, Schwarz K (2004) Magnetic structure and electric-field gradients of uranium dioxide: an ab initio study. Phys Rev B 69:140408

Lassmann K (1980) The structure of fuel element codes. Nucl Eng Des 57:17–39

Lawrence GT (1978) A review of the diffusion coefficient of fission-product rare gases in uranium dioxide. J Nucl Mater 71:195–218

Liu XY, Uberuaga BP, Andersson DA, Stanek CR, Sickafus KE (2011) Mechanism for transient migration of xenon in UO_2. Appl Phys Lett 98:51902

Ma PW, Woo CH, Dudarev SL (2008) Large-scale simulation of the spin-lattice dynamics in ferromagnetic iron. Phys Rev B 78:024434

Ma PW, Dudarev SL (2014) Dynamic magnetocaloric effect in bcc iron and hcp gadolinium. Phys Rev B 90:024425

Mei ZG, Stan M, Pichler B (2013) First-principles study of structural, elastic, electronic, vibrational and thermodynamic properties of UN. J Nucl Mater 440:63–69

Meredig B, Thompson A, Hansen HA, Wolverton C, van de Walle A (2010) Method for locating low-energy solutions within DFT+U. Phys Rev B 82:195128

Miao Y, Gamble KA, Andersson D, Mei ZG, Yacout AM (2018) Rate theory scenarios study on fission gas behavior of U_3Si_2 under LOCA conditions in LWRs. Nucl Eng Design 326:371–382

Miao Y, Gamble KA, Andersson D, Ye B, Mei ZG, Hofman G, Yacout A (2017) Gaseous swelling of U_3Si_2 during steady-state LWR operation: a rate theory investigation. Nucl Eng Design 322:336–344

Middleburgh SC, Grimes RW, Lahoda EJ, Stanek CR, Andersson DA (2016) Non-stoichiometry in U_3Si_2. J Nucl Mater 482:300–305

Miekeley W, Felix FW (1972) Effect of stoichiometry on diffusion of xenon in UO_2. J Nucl Mater 42:297–306

Mills G, Jónsson H, Schenter GK (1995) Reversible work transition state theory: application to dissociative adsorption of hydrogen. Surf Sci 324:305–337

Nerikar P, Watanabe T, Tulenko JS, Phillpot SR, Sinnott SB (2009a) Energetics of intrinsic point defects in uranium dioxide from electronic-structure calculations. J Nucl Mater 384:61–69

Nerikar PV, Liu XY, Uberuaga BP, Stanek CR, Phillpot SR, Sinnott SB (2009b) Thermodynamics of fission products in $UO_{2\pm x}$. J Phys Condens Matter 21:435602

Noordhoek MJ, Andersson D, Chernatynskiy A, Middleburgh S, Besmann TM (2016) Phase equilibria in the U-Si system from first-principles calculations. J Nucl Mater 479:216–223

Pang JWL, Buyers WJL, Chernatynskiy A, Lumsden MD, Larson BC, Phillpot SR (2013) Phonon lifetime investigation of Anharmonicity and thermal conductivity of UO_2 by neutron scattering and theory. Phys Rev Lett 110:157401

Perdew JP (1985) Accurate density functional for the energy: real-space cutoff of the gradient expansion for the exchange hole. Phys Rev Lett 55:1665–1668

Perdew JP, Wang Y (1986) Accurate and simple density functional for the electronic exchange energy: generalized gradient approximation. Phys Rev B 33:8800–8802

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Rashid Y, Dunham R, Montgomery R (2004) Fuel analysis and licensing code: FALCON MOD01. Technical report EPRI 1011308, Electric Power Research Institute

Rashid JYR, Yagnik SK, Montgomery RO (2011) Light water reactor fuel performance modeling and multi-dimensional simulation. JOM 63:81–88

Sanati M, Albers RC, Lookman T, Saxena A (2011) Elastic constants, phonon density of states, and thermal properties of UO_2. Phys Rev B 84:014116

Souvatzis P, Eriksson O, Katsnelson MI, Rudin SP (2008) Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory. Phys Rev Lett 100:095901

Söderlind P, Grabowski B, Yang L, Landa A, Björkman T, Souvatzis P, Eriksson O (2012) High-temperature phonon stabilization of γ-uranium from relativistic first-principles theory. Phys Rev B 85:060301(R)

Taylor SE, Bruneval F (2011) Understanding and correcting the spurious interactions in charged supercells. Phys Rev B 84:075155

Thompson AE, Wolverton C (2011) First-principles study of noble gas impurities and defects in UO_2. Phys Rev B 84:134111

Thompson AE, Wolverton C (2013) Pathway and energetics of xenon migration in uranium dioxide. Phys Rev B 87:104105

Tonks MR, Andersson D, Phillpot SR, Zhang Y, Williamson R, Stanek CR, Uberuaga B, Hayes SL (2017) Mechanistic materials modeling of nuclear fuel performance. Ann Nucl Energy 105: 11–24

Turnbull JA, Friskney CA, Findlay JR, Johnson FA, Walter AJ (1982) The diffusion coefficients of gaseous and volatile species during the irradiation of uranium dioxide. J Nucl Mater 107: 168–184

Vanderbilt D (1990) Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Phys Rev B 41:7892–7895

Vathonne E, Wiktor J, Freyss M, Jomard G, Bertolus M (2014) DFT+U investigation of charged point defects and clusters in UO$_2$. J Phys Condens Matter 26:325501

Vathonne E, Andersson DA, Freyss M, Perriot M, Cooper MWD, Stanek CR, Bertolus M (2017) Determination of krypton diffusion coefficients in uranium dioxide using atomic scale calculations. Inorg Chem 56:125

Wang BT, Zhang P, Lizarraga R, Marco ID, Eriksson O (2013) Phonon spectrum, thermodynamic properties, and pressure-temperature phase diagram of uranium dioxide. Phys Rev B 88:104107

Williamson RL, Hales JD, Novascone SR, Tonks MR, Gaston DR, Permann CJ, Andrs D, Martineau RC (2012) Multidimensional multiphysics simulation of nuclear fuel behavior. J Nucl Mater 423:149–163

Yin Q, Savrasov SY (2008) Origin of low thermal conductivity in nuclear fuels. Phys Rev Lett 100:225504

Yu J, Devanathan R, Weber WJ (2009) First-principles study of defects and phase transition in UO$_2$. J Phys Condens Matter 21:435401

Yun Y, Kim H, Kim H, Park K (2008) Atomic diffusion mechanism of Xe in UO$_2$. J Nucl Mater 378:40–44

Ram Devanathan

Contents

Abstract

The interatomic potential is a mathematical function that represents the potential energy of a collection of interacting atoms as a function of geometric variables, such as distance and angle. The potential provides the underpinning physical basis for molecular statics, molecular dynamics, and kinetic Monte Carlo simulations of radiation effects in materials. The validity of these simulations is governed by the fidelity of the potential. The functional form of the potential is chosen based on the type of interaction, e.g., metallic, covalent, or ionic. The

R. Devanathan (✉)
Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA
e-mail: ram.devanathan@pnnl.gov

parameters of the function are optimized by fitting to properties of the system of interest obtained from experiment or density functional theory calculations. Caution should be exercised when applying interatomic potentials to conditions that are quite different from the ones used to fit the potential. Charge transfer processes, dissimilar material interfaces (e.g., metal-ceramic interfaces), and multicomponent alloys are challenging to represent using potentials. This chapter introduces potential functions, offers examples of the application of potentials to understand the primary damage state, and presents an assessment of potentials for studying the radiation response of materials.

1   Introduction

Atomistic computer simulation, when used in conjunction with experimental validation, can offer predictive understanding of the performance and degradation of nuclear materials. Simulations can shed light on the early stages of radiation damage, mechanical degradation, species diffusion, gas bubble formation, nucleation and growth of phases, chemical segregation, phase transformation, interface restructuring, corrosion, and fracture of materials under the combined effects of mechanical stress, corrosives, extremes of temperature and irradiation by energetic particles. Often, this information is challenging to obtain by experiment alone. Classical molecular dynamics (MD) and Monte Carlo (MC) simulations require a rigorous description of the potential energy of the system. The widely used MD simulation method involves the numerical integration of Newton's equations of motion for a system of interacting atoms given by Eq. (1).

$$m_i \frac{d^2 \vec{r}_l}{dt^2} = \cdot - \nabla_{\vec{r}_l} V \tag{1}$$

m_i is the mass of atom i, $\vec{r}_i$ is its position vector, t is time, and V is the potential energy, which is a function of geometric variables such as the interatomic distances and bond angles. The potential function consists of a sum of terms that represent different interactions relevant to the material system of interest. These interactions can involve one, two-, three-, or many-body terms. For computational efficiency, the interactions are often limited to a sphere around each atom with the radius being the interaction cutoff distance.

The predictive power of the simulation depends on the reliability of the interatomic potential. The parameters of the potential function are derived by fitting to known experimental properties, such as lattice constants and elastic constants. In addition, point defect properties calculated using density functional theory (DFT) can be used to optimize the potential parameters. The application of DFT to nuclear materials is discussed in detail by Andersson (▶ Chap. 86, "Density Functional Theory Calculations Applied to Nuclear Fuels") in this volume. There is no perfect potential for a material. In selecting functional forms and optimizing parameters to simulate materials phenomena of interest, trade-offs are made by balancing

accuracy, development time, computational cost, ability to represent problem of interest, and transferability to other problems. It is important to understand the limitations imposed by the choice of a function to avoid using potentials outside their range of applicability. In the following sections, we present commonly used functional forms and discuss the emerging trends in potential development.

2   Potential Functions

The functional form of the interatomic potential, also known as the force field, depends on the dominant type of interaction present in the material, e.g., metallic, covalent, or ionic. This function is chosen typically based on physical intuition about the nature of the interaction.

2.1   Intra-molecular Functions

Intra-molecular potentials operate between atoms in a molecule. They are commonly encountered in simulations of polymers and liquids, for instance, simulation of the interaction of water with nuclear waste glass. There are several well-known examples of these potentials, including OPLS (Jorgensen et al. 1996), AMBER (Wang et al. 2004), and DREIDING (Mayo et al. 1990).

Covalent bonds between adjacent atoms are represented by a bond potential that minimizes the energy at an experimentally determined bond distance, r_0. There are many functional forms for this bond potential. The most common one is the harmonic bond potential, where the bond energy V_B between atoms i and j is summed over all bonds in the system. k_{ij} is the force constant and r_{ij} is the bond length between atoms i and j.

$$V_B = k_{ij}(r_{ij} - r_0)^2 \tag{2}$$

Similarly, one can constrain the angle θ_{ijl} between bond vectors $\overrightarrow{r_{ij}}$ and $\overrightarrow{r_{il}}$, around a central atom i between atoms j and l to approach the experimentally determined bond angle θ_0. Here again, the harmonic form of the angle potential V_A with a force constant k_{ijl} is the most common form.

$$V_A = k_{ijl}(\theta_{ijl} - \theta_0)^2 \tag{3}$$

For four consecutive bonded atoms i, j, l, and m, the torsion around the jl bond can be constrained, typically using the cosine of the angle ϕ_{ijlm} between the plane containing the ij and jl bonds and the plane with the jl and lm bonds. This is called a dihedral or torsion potential, and it can also assume the harmonic form similar to Eq. (3). More complex intra-molecular interactions can also be included but will not be discussed here. These intramolecular functions are relevant to the simulation

of radiation effects in polymers (Polvi et al. 2012). The rest of the discussion will focus on intermolecular or interatomic interactions that are especially important for metals, semiconductors, and ceramics.

2.2 Simple Pair Potentials for Model Systems

The Lennard-Jones (LJ) potential (Lennard-Jones 1931) is one of the simplest interaction models. It is a summation of the energies of bonds between pairs of neutral entities (atoms or molecules). It is also known as the 12-6 potential because of its functional form shown in Eq. (4).

$$V_{LJ}\left(r_{ij}\right) = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right]^{-} \tag{4}$$

This energy is summed over all the bonds present in the system to get the potential energy. The potential has two adjustable parameters, σ representing the finite interatomic separation at which V_{LJ} becomes zero and ε representing the depth of the potential minimum that occurs at an interatomic distance r_{ij} of $2^{1/6}\sigma$. The first term in Eq. (4) represents Pauli repulsion and the second term represents van der Waals attraction. The LJ potential has been used to model noble gases and has the advantage of computational efficiency. The potential is too simple by itself to model the behavior of solid systems, especially those relevant to nuclear energy technology.

The repulsive part of the LJ potential was changed to an exponential term by Buckingham (Buckingham 1938) as shown in Eq. (5).

$$V_{Buck}\left(r_{ij}\right) = A_{ij}e^{-B_{ij}r_{ij}} - \frac{C_{ij}}{r_{ij}^{6}} \tag{5}$$

This potential lacks directionality and is spherically symmetric as the LJ potential. It includes 3 adjustable parameters, namely, A, B, and C. The repulsive term converges to A as r_{ij} approaches zero while the magnitude of the attractive term increases leading to unphysical attraction at small values of r_{ij}. This can make MD simulations using the Buckingham potential unstable. The repulsive part must be modified for use in simulations of energetic events, such as displacement cascades (Dong et al. 2016).

2.3 Potentials for Metals and Alloys

The summation of pairwise bond energies over all bonds present in the system is simplistic, because it assumes that a given bond is not influenced by the presence of other bonds in the system. In reality, when the coordination number is increased,

the new bond contributes a smaller value to the cohesive energy than the original bond. It is known that pairwise potentials satisfy the Cauchy relation for elastic constants ($C_{12} = C_{44}$), which is not true for most metals. Moreover, the vacancy formation energy is comparable to the cohesive energy for pair potentials and this does not hold for metals (Ercolessi et al. 1988). To overcome the deficiencies of pair potentials, effective medium potentials have been developed for metals under the Finnis-Sinclair (Finnis and Sinclair 1984) formalism or the embedded atom method (EAM) (Daw and Baskes 1984; Daw et al. 1993). The EAM potentials can be expressed as follows:

$$V_{EAM} = \sum_i F \left(\sum_i \rho_i \left(r_{ij} \right) \right) + \frac{1}{2} \sum_{i,j \ (i \neq j)} V_{rep} \left(r_{ij} \right) \qquad (6)$$

The first term is the attractive embedding energy, which is a functional that varies nonlinearly with the coordination number. It is a function of the electron density, ρ. This electron density is in turn expressed as a function of r_{ij}. The second term is the pairwise repulsive term. In a similar way, the Finnis-Sinclair potential energy is the sum of two terms, an attractive term that is proportional to the square root of the electron density and a pairwise repulsive potential. The square root form was inspired by the second moment approximation to the tight-binding density of states and may not necessarily have a physical significance. The EAM-type potentials have been successfully applied to simulate the behavior of several metals. The EAM method has been extended to covalent materials through the modified EAM (MEAM) approach, where angle-dependence was introduced into the density, ρ (Baskes 1987). MEAM potentials have been developed for C, Al, Si, Ti, V, Cr, Mn, Fe, Ni, Cu, Zr, Nb, Mo, Pd, Ag, Ta, W, Pt, Au, Pb, and various alloys of these elements (Lee et al. 2010).

For nuclear energy applications, we will consider three metals, namely Zr, U, and Fe. We pick these cases as illustrative examples from the large body of work on potentials for metals and alloys. Zr-based alloys are used as fuel cladding in nuclear reactors. Several potentials have been developed for Zr, but they do not reproduce the high stacking fault energy and phase transformations with increasing temperature. Mendelev and Ackland have developed a reliable MEAM potential for hexagonally close packed Zr by using defect formation energies and stacking fault energies from DFT in the fitting procedure (Mendelev and Ackland 2007). This work is interesting for three reasons. First, it links DFT and MD simulations by using DFT results to parameterize potentials. Second, it clearly illustrates the tradeoffs in potential development, by generating two potentials that reproduce different sets of properties. The authors emphasize the need for care in selection of properties to which the potential is fitted. Third, it raises questions about the ability of the EAM functional form to capture the full range of properties of metals. In a similar vein, a MEAM potential has been developed to represent the body centered cubic (bcc) phase of U-Zr alloys (Moore et al. 2015). Alloys of U and

Zr have potential applications as metallic fuel in advanced nuclear reactors. Defect formation energies from DFT calculations were used along with experimental values of the elemental elastic constants in the fitting.

One of the important advances in potential fitting is the Force-Matching method (Ercolessi and Adams 1994). This approach uses DFT to obtain forces for many atomic configurations of interest, including crystals, liquids, surfaces, and atomic clusters, and then fits potentials by matching these forces. An EAM potential for Fe has been developed using force-matching (Mendelev et al. 2003). The authors generated a simple model of liquid Fe at the known density and determined forces for this model using DFT. The EAM potential was fit to these forces, lattice constants of bcc and face centered cubic (fcc) Fe at 0 K, vacancy formation energy, and energy difference between bcc and fcc phases. The potential was found to reproduce properties such as the lattice constant, elastic constants, point defect energies, liquid density, and melting temperature better than existing potentials for Fe. Despite these advances, the empirical potential approach is severely limited for the simulation of a real alloy, such as stainless steel. This task requires potentials that describe the interactions of Fe, Cr, Ni, C, and minor alloying elements that contribute to precipitation hardening and solid solution strengthening. It is also challenging to develop a potential that stabilizes the fcc phase in the composition range of interest. Bonny et al. (2013) have developed an EAM-type potential for FeNiCr that may be a good starting point for modeling radiation damage in austenitic stainless steels. The potential stabilizes the fcc phase. It has been benchmarked against DFT data and experimental observations.

2.4   Potentials for Covalent Materials

The Stillinger-Weber (S-W) potential for Si presents another good example of going beyond the limitations of pair potentials (Stillinger and Weber 1985). In this case, the potential met a need to stabilize the diamond structure and reproduce the densification of Si and Ge on melting. The S-W potential is a sum of two-body and three-body terms as shown in Eq. (7). The latter introduces dependence on the angle between ij and ik bonds, θ_{ijk}.

$$V_{S-W} = \sum_i \sum_{j>i} V_2\left(r_{ij}\right) + \sum_i \sum_{j\neq i} \sum_{k>j} V_3\left(r_{ij}, r_{ik}, \theta_{ijk}\right) \qquad (7)$$

This potential has been applied to the study of Si, Ge, III-V, and II-VI compound semiconductors.

The Tersoff potential brings in the angular term in the form of bond order parameter, b_{ij} (Tersoff 1988). The idea is that the strength of a bond depends on its environment, namely, the number of bonds, bond lengths, and bond angles. This dependence is introduced through b_{ij}. The general form of the Tersoff potential can be written as follows.

$$V_{\text{Ters}} = \sum_i \sum_{j>i} V_{repulsive}\left(r_{ij}\right) + b_{ij} V_{attractive}\left(r_{ij}\right) \tag{8}$$

The Tersoff potential has been widely used to study radiation effects in covalently bonded materials, including Si and SiC. The latter is a component of TRISO nuclear fuel and is also a fusion reactor material. Brenner has developed bond order potentials that are especially tailored for carbon-based materials (Brenner 1990). This work has been developed further in the form of reactive empirical bond order potentials (REBO) that are well suited to model solid carbon-based materials (Brenner et al. 2002). Another key development in this area is the Environment Dependent Interatomic Potential (EDIP) that inverts cohesive energy curves obtained from first principles data to derive many-body potentials (Bazant et al. 1997). The bond order potentials can accommodate multiple bonding states but not changes in the charge state. To understand this issue, we will discuss ionic materials next.

2.5   Potentials for Ionic Materials

We have seen that the interatomic potential can be expressed as a sum of one-, two-, and many-body terms. Terms higher than three-body will be challenging to evaluate computationally. The three body terms are needed for covalent materials as discussed above. For ionic materials, many-body terms are implicitly incorporated into an effective pair potential, which is typically a sum of long-range Coulombic interaction between charges $\bar{q}_i$ and q_j separated by a distance r_{ij}, V_{LR}, and short-range interactions, V_{SR}.

$$V_{LR}\left(r_{ij}\right) = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j e^2}{r_{ij}} \tag{9}$$

ε_0 is the permittivity of vacuum.

$$V_{SR}\left(r_{ij}\right) = A_{ij} e^{-B_{ij} r_{ij}} - \frac{C_{ij}}{r_{ij}^6} + D_{ij}\left[e^{-2m_{ij}\left(r_{ij}-r_{ij}^0\right)} - 2e^{-m_{ij}\left(r_{ij}-r_{ij}^0\right)}\right] \tag{10}$$

The short-range potential in this case is the sum of the Buckingham potential in Eq. (5) and the Morse potential with 3 additional fitting parameters, D, m, and r_0.

The short-range potential can become attractive at close separations less than 0.1 nm. The Ziegler-Biersack-Littmark (ZBL) repulsive potential (Ziegler et al. 2010), V_{ZBL}, is based on a universal screening function and is often smoothly splined to the Buckingham potential for MD simulation of energetic recoils (Dong et al. 2016). The combined potential remains repulsive at very small values of r_{ij}.

$$V_{ZBL}\left(r_{ij}\right) = \frac{1}{4\pi\varepsilon_0} \frac{Z_i Z_j e^2}{r_{ij}} \varphi\left(\frac{r_{ij}}{a}\right) \tag{11}$$

$$\varphi(x) = 0.1818e^{-3.2x} + 0.5099e^{-0.9423x} + 0.2802e^{-0.4029x} + 0.02817\mathrm{e}^{-0.2016x} \tag{12}$$

$$a = \frac{0.8854a_0}{Z_i^{0.23} + Z_j^{0.23}} \tag{13}$$

Here, Z_i and Z_j are the charges of the nuclei of atoms i and j, respectively, and a_0 is the Bohr radius.

The potentials for ionic systems can include rigid-ion models or core-shell models (Dick and Overhauser 1958). The former treat ions as fixed point charges, which ignores polarization. The shell models treat ionic polarizability by considering ions as cores connected to shells (Lewis and Catlow 1985). Each ion is represented by two particles of opposite charge, namely, a core and a shell connected by a harmonic spring. The sum of the core and shell charges is the ion charge. The original core-shell model of Dick and Overhauser (1958) was based on massless shells. At each MD step, one needs to perform an optimization to restore the shells to their zero-force positions. This creates a significant computational overhead. Due to residual forces, there can be small energy drifts even after careful optimization of the shells.

In the dynamical core-shell model, the shells are treated as particles with very small mass compared to the mass of the corresponding core particle. The sum of the core and shell masses is equal to the atomic mass. The more massive core represents the nucleus and its inner electron shells, while the shell represents the valence electrons. The electrostatic interaction between ions i and j is made of four terms instead of one – core i-core j, core i-shell j, core j-shell i, and shell i-shell j terms. There is no electrostatic interaction between the core and shell of the same ion. The short-range potential acts only between the shells.

The shells are propagated concurrently with the cores during each MD time step. The movement of the lighter shell in response to changes in the electric field acting on the ion produces a dipole, which represents the polarization of the electron cloud by the external electric field. The extent of the core-shell motion limits the size of the MD time step. The core-shell model requires the use of a smaller MD time step compared to the rigid-ion model to ensure energy conservation and stability of the simulation.

MD simulations of energetic processes, such as recoil damage events, can become unstable if a shell model potential is used. In such simulations, dynamic processes can drive two ions to approach each other closely leading to a polarization catastrophe. At short separations, the strong Coulomb attraction between the core of one ion and the shell with opposite charge of a nearby ion can cause the simulated structure to collapse. This instability limits core-shell potentials to low-temperature simulations, and these models are not used in simulations of displacement cascades

and thermal spikes. For such simulations, rigid-ion potentials offer the advantages of lower computational intensity and greater stability. In rigid-ion models, the ions can be assigned formal charges ($+4$ for U and -2 for O) or partial charges based on fitting to experimental data. However, the shell models give a better representation of the equilibrium properties including the elastic constants in static calculations. The choice of rigid ion or shell model will be based on the phenomena of interest. These ionic potentials have the limitation that the charge of the ion is held fixed. The charge cannot be changed in response to changes in the local environment. This fixed charge model reduces the fidelity of the simulation and imposes restrictions on the types of materials or processes that can be simulated.

2.6   Charge Optimized Potentials and Reactive Force Fields

Reactive force fields are needed to model dissimilar material interfaces, for instance, to simulate corrosion of metals and alloys. In addition, materials used in nuclear technologies are composed of multiple phases, for instance, carbide precipitates in steel or oxide phases in oxide dispersion strengthened steel. In the last two decades, new approaches have been developed to treat these types of interfacial problems. There are two key concepts underlying these approaches. One of these concepts, bond order, has been presented in the previous section in the context of potentials developed by Tersoff and Brenner. The bond order can be used to self-consistently identify the type of interaction present in the simulation. The other concept is the self-consistent equilibration of charge (Phillpot and Sinnott 2009).

A series of charge optimized many-body (COMB) potentials have been developed by Phillpot, Sinnott, and coworkers (Liang et al. 2013) to simulate multiphase materials. The charge state of an ion is dynamically determined based on the local environment using the principle of electronegativity equalization. The charges are iteratively optimized, which imposes an order of magnitude higher computational cost than using fixed charge potentials. The fidelity and cost of simulations with such potentials lie somewhere between those of molecular dynamics with fixed charges and density functional theory. The potential energy in the COMB formalism is the sum of electrostatic, short-range, van der Waals, and correction terms as shown in Eq. (14).

$$V_{COMB}(q,r) = V_{electrostatic}(q,r) + V_{SR}(q,r) + V_{VDW}(r) + V_{correction}(r) \tag{14}$$

The first two terms depend on ionic charges q and positions r, while the last two depend only on positions. The electrostatic energy is a sum of the self-energy to form a charge, the charge-charge (qq) interaction energy, the charge-nucleus (qZ) interaction energy, and an energy related to polarizability.

$$V_{electrostatic}(q,r) = V_{self}(q,r) + V_{qq}(q,r) + V_{qZ}(q,r) + V_{polar}(q,r) \tag{15}$$

The short range term is a bond order potential that is charge dependent. The Van der Waals term has been discussed previously. The correction term is included to eliminate unphysical interactions and add new functionality. In addition, this term can be modified to reproduce experimentally determined lattice constants. COMB potentials have been used to simulate metal-oxide and semiconductor-oxide interfaces and the oxidation process at the surface of Cu.

The bond order formalism has been used by van Duin and coworkers (Senftle et al. 2016) to develop reactive force fields, ReaxFF, that allow for dynamic bond breaking and bond formation during the simulation. The potentials have a good degree of transferability, because a given element is treated the same way regardless of whether it is in the gas phase, liquid phase, or a dissimilar material interface. The ReaxFF potential energy is the sum of the following terms: bond energy, bond angle, torsion, over-coordination penalty, van der Waals, Coulomb interaction, and a specific energy term.

$$V_{Reax} = V_B + V_A + V_T + V_{over} + V_{VDW} + V_{Coulomb} + V_{specific} \qquad (16)$$

The first four terms are bond order dependent, and the next two terms are charge dependent.

ReaxFF started as a hydrocarbon potential but has been adapted to simulate many different classes of materials, including metals and ceramics. Over the years, different parameter sets have been developed for different materials families, and it is important to choose the set that is relevant to the problem at hand. Despite the previously mentioned transferability of ReaxFF, parameters developed to study O for a combustion chemistry problem may not be appropriate for O at a mineral-water interface. ReaxFF has been developed by chemists interested in reactive processes and may be a good choice to model catalysts. Elastic properties of interest to metallurgists have not been usually included in the force field fitting. However, there is nothing that precludes the use of elastic properties to reparameterize ReaxFF to study thermo-mechanical degradation of alloys for energy technology applications. Shin et al. (2012) have developed a ReaxFF parameterization for binary alloys in the Fe-Ni-Al system using data from first principles calculations. The fitting procedure considered elastic properties of various phases, surface energies, adatom binding energies, and diffusivities. Simulations performed with this force field have provided insights into segregation in Fe-Ni-Al alloys.

3 Application of Potentials to Simulate the Primary Damage State

There is a long history of using interatomic potentials to study the dynamics of radiation damage in materials (Gibson et al. 1960; Robinson and Torrens 1974; Guinan and Kinney 1981). This section presents recent examples of the use of interatomic potentials to study the primary damage state in a variety of materials mainly using MD simulations. Djurabekova and Nordlund

(► Chap. 88, "Molecular Dynamics Simulations of Non-equilibrium Systems") provide a detailed discussion of MD simulations of collision cascades and swift heavy ion damage elsewhere in this volume. Given that the radiation damage resistance of a crystalline material is dictated by the competing effect of short-range covalent and long-range ionic forces (Trachenko et al. 2005), it is evident that MD simulations of radiation damage will strongly depend on the choice of interatomic potential. In the last decade, there has been considerable interest in evaluating the fidelity of potentials for simulating radiation damage in a material, usually by comparing defect properties to DFT calculations or experimental data.

3.1   Defects in Metals and Alloys

There is a large database of DFT defect energies in α-Fe, and many potentials have been developed to simulate radiation damage in α-Fe (Malerba et al. 2010). This combination makes this material a good test bed for evaluating the effect of the features of a given interatomic potential on the ability to match defect energy data from DFT studies. Malerba et al. (2010) compared five semi-empirical potentials by studying point defect properties and defect cluster distributions. These included Mendelev-type potentials (Mendelev et al. 2003) as well as magnetic potentials that modify the embedding term in EAM using a combination of Stoner and Ginzburg-Landau models (Dudarev and Derlet 2005). They found that Mendelev-type potentials match DFT results most closely and thus are ideal for extending the scale of DFT calculations to simulate radiation damage. This finding is not surprising given that the fitting of the Mendelev potential was informed by DFT calculations. Gao et al. (2017) have simulated displacement cascades with energy up to 20 keV in α-Fe using three magnetic potentials and a Mendelev-type potential. For the same recoil energy, they found differences in defect production and cascade volume between these potentials. The magnetic potentials gave a larger interstitial formation energy than the Mendelev-type potential, which resulted in smaller self-interstitial clusters in the former. If the cluster size distribution is used as input in coarser grained simulations, the evolution of the radiation damage will be quite different for these potentials. Malerba et al. (2010) have concluded that none of the potentials they studied were satisfactory for describing self-interstitial clusters and dislocation interactions in Fe. There is a need to improve the fidelity of these potentials to make them more reliable for the study of radiation damage evolution. This conclusion is consistent with a recent finding of Klaver et al. (2016) based on a study of the FeCr system.

FeCr alloys are promising candidates for structural applications in nuclear reactors due to their radiation tolerance and corrosion resistance. The addition of Cr to Fe changes the radiation response. The combined effects of radiation, stress, and elevated temperature can accelerate the precipitation of a Cr-rich α' phase in FeCr alloys. Such precipitates have the potential to cause embrittlement of the material. The sign of the mixing enthalpy in this system changes with increasing Cr concentration (Malerba et al. 2008), which presents a challenge for EAM potentials.

Efforts have been made to refine the FeCr interaction and reproduce the change in the mixing enthalpy. These modifications include multiplying the repulsive term in q. (6) by a local concentration-dependent term and the introduction of two embedding terms – one for the *s* band and another for the *d* band. These potentials do not explicitly account for an important property of this system, namely, magnetism. The development of a spin-polarized potential will require a new functional form and a new algorithm to handle spin dynamics in MD simulations (Malerba et al. 2008).

Klaver et al. (2016) have evaluated the ability of four FeCr potentials to model interstitial formation and binding energies in FeCr alloys containing 8 at.% and 16 at.% Cr. The energies were compared to corresponding values from DFT calculations. Getting the defect energies correct is key to reliably simulating the primary damage state produced by energetic recoils. These recoils initiate atomic displacement cascades that displace many atoms within 1 ps followed by dynamic vacancy-interstitial recombination and defect clustering on the scale of tens of ps (Devanathan 2009). The size distributions of the resulting vacancy clusters and interstitial clusters are different, and their thermal evolution is also different due to differences in migration energies. These differences in cluster fractions, defect stability, segregation at interfaces, and cluster lifetime can lead to large differences in the numbers of freely migrating vacancies and interstitials known as defect production bias. This bias can influence the long-term evolution of radiation damage and observed properties such as swelling. The work of Klaver et al. (2016) shows inconsistencies between defect energies calculated using empirical potentials and DFT. This discrepancy raises questions about the ability to reliably model radiation damage in FeCr alloys with high Cr content using existing empirical potentials. While it is possible to refit potentials to improve the agreement with defect properties from DFT, there is practically no wiggle room for improving the fitting using current functional forms. One must look beyond the EAM formalism to model radiation effects in binary and multicomponent alloys.

3.2   Radiation Damage in a Covalent Material

SiC is a covalent material that has potential applications as a structural material in fusion reactors and a fuel cladding in advanced fission reactors. Crystalline SiC is known to be amorphized under energetic particle irradiation. Gao, Weber, and Devanathan (2002) have reviewed fundamental mechanisms of radiation damage accumulation in SiC based on the findings of MD simulations using bond-order potentials. In this material, the minimum energy needed to displace an atom from its lattice site, known as the displacement threshold energy (E_d), is quite different for Si and C regardless of the potential used. The threshold energy surface is also highly anisotropic. The E_d values serve as important input parameters for coarser-grained simulations of radiation damage. MD simulations with empirical potentials show that the defect production in SiC is dominated by C interstitials and vacancies. The damage produced by energetic recoils consists of a central vacancy-rich region and

an interstitial-rich periphery. The primary damage state consists of isolated interstitials and vacancies and small defect clusters. In contrast to the primary damage state in alloys with defect clusters containing tens of defects, in SiC defect clusters containing more than three defects were rarely seen. The amorphization of SiC proceeded by the gradual accumulation of clusters or the overlap of displacement cascades instead of direct amorphization within the cascade. The damage dose in displacements per atom for complete amorphization and the swelling behavior of the material showed good agreement with experiment, which provides validation for this bond-order potential. High-resolution transmission electron microscopy (HRTEM) images calculated from the MD simulation cells also showed agreement with corresponding experimental images of irradiated SiC and provided insights into atomic-level mechanisms behind the experimental observations.

Subsequent MD simulations showed that chemical short-range disorder (site exchange between Si and C) measured by the fraction of homonuclear (C-C) bonds may play an important role in the amorphization of SiC (Devanathan 2009). The increases in enthalpy and volume of the system with the introduction of atomic displacements or Si-C exchanges closely followed the increase in the C-C homonuclear bond ratio regardless of the sublattice in which displacement was introduced. Interestingly, the simulated amorphous states produced by radiation damage accumulation and melt quenching had slightly different densities and homonuclear bond ratios. This points to the existence of different amorphous or highly damaged states in SiC. Debelle et al. (2014) have used MD simulations with the Gao-Weber bond-order potential (2002) to shed light on recrystallization by swift heavy ion energy deposition in SiC with different levels of preexisting damage or amorphization. The simulated recrystallization behavior was quite different depending on the extent of amorphization and the findings were in agreement with experiment.

Recently, Samolyuk et al. (2015) have studied the migration barrier for carbon interstitials in SiC using three commonly used potentials and compared the results to DFT calculations. Two of the potentials yielded a migration barrier that was much higher than DFT results, while the third potential by Gao and Weber (2002) yielded a migration barrier closer to the DFT result. When 10 keV and 50 keV displacement cascades were simulated with these potentials, the Gao-Weber potential produced fewer defects in the primary damage state due to more effective damage annealing by the rapid migration of carbon interstitials. This study has identified the source of differences in the damage state produced in SiC with different potentials, by benchmarking defect energies from these potentials against DFT results.

3.3 Radiation Effects in Ionic Materials

Sickafus et al. (2000, 2007) have performed molecular statics calculations with a set of shell model potentials (Minervini et al. 2000) to shed light on the radiation tolerance of pyrochlore oxides. These potentials are transferable across a wide range of pyrochlore oxides, which makes it possible to evaluate trends in the

radiation response as A and B cations are varied in $A_2B_2O_7$ pyrochlore. Sickafus et al. (2007) concluded based on their calculations integrated with experiments that better radiation tolerance is expected in materials with inherently disordered phases with wide ranges of stoichiometry. Devanathan et al. (2010a) have used a rigid ion potential based on the Minervini potential to study radiation-induced defect accumulation in gadolinium zirconate and titanate pyrochlores. The study showed that cation Frenkel pair accumulation played a key role in the amorphization of $Gd_2Ti_2O_7$. The simulations also quantified the swelling accompanying the damage accumulation and amorphization. In contrast, when defects were introduced in $Gd_2Zr_2O_7$, there was efficient dynamic annealing and no amorphization. Zhang et al. (2010) integrated experimental swift heavy ion irradiation and HRTEM characterization of pyrochlores with MD simulation of thermal spike damage using these potentials. The morphology of individual ion tracks produced by 2.2 GeV Au ions in $Gd_2Ti_2O_7$, Gd_2TiZrO_7, and $Gd_2Zr_2O_7$ as revealed by HRTEM was reproduced by the thermal spike simulations in these materials. The simulations provided insights into the time evolution of ion tracks and revealed the phase transition pathways leading to the concentric track structures observed experimentally. In $Gd_2Ti_2O_7$, the simulations showed an amorphous core for the ion track core, which was surrounded by a concentric shell of defect fluorite, and an outermost shell of defective pyrochlore, consistent with the nanoscale core-shell structure seen in HRTEM images. In $Gd_2Zr_2O_7$, rapid recrystallization of the ion track prevented amorphization and instead led to a defect fluorite core consistent with experimental observation. The reproduction of the experimentally observed morphology by the simulation for three different compositions provides validation for the potential. At the same time, the simulation has revealed atomic-level mechanisms behind the different responses of these pyrochlore to the passage of swift heavy ions. These pyrochlores have potential applications as waste forms for immobilization of high-level nuclear waste.

Devanathan (2009) has used MD with a partial charge model and Buckingham-type potentials to simulate energetic recoils in zircon ($ZrSiO_4$), which is also a nuclear waste form. The simulations revealed that about 5000 atoms were displaced by 30 keV U recoils consistent with the findings of a nuclear magnetic resonance experiment (Farnan et al. 2007). The simulations showed that direct amorphization occurred in the cascade. The amorphous core was relatively enriched in Si, while the periphery of the cascade was enriched in Zr. The experimental study by Farnan et al. (2007) assumed that Si displacements were about one-sixth of total displacements according to stoichiometry. The simulations confirmed that this is indeed the case. The simulations also revealed the occurrence of Si-O-Si polymerization and under-coordination of Zr in the displacement cascade.

In contrast to the amorphization by direct impact in zircon and by damage accumulation in SiC, UO_2 and its fluorite structured surrogate CeO_2 are not amorphized by swift heavy ions. MD simulations of swift heavy ion damage in CeO_2, simulated using thermal spikes, showed that the primary damage state consisted of point defect clusters with no amorphization (Yablinsky et al. 2015). The results were consistent with experiments carried out by these authors. CeO_2

remained crystalline following irradiation with 940 keV Au gold ions. Experimental results suggested a small reduction in density at the center of the ion track. Simulation results showed a lower density track core and indicated the formation of interstitial loops at the periphery of the track consistent with experimental findings.

MD simulations of U recoils with energies in the range of 0.5–80 keV in UO_2 have been performed with rigid ion potentials (Devanathan et al. 2010b). Regardless of the recoil energy, there was no amorphous zone in the primary damage state of UO_2. Vacancies were preferentially created at the center of the cascade and interstitials at the periphery. The cluster size distribution was quite different for vacancies (large clusters) and interstitials (smaller clusters). In addition, swift heavy ion simulations in UO_2 (Devanathan et al. 2010b) have shown that the radiation response depends on the linear energy transfer (LET). Defects were not produced for LET less than 4 keV/nm. From 4 to 16 keV/nm, isolated point defects were produced. At much higher LET (66 keV/nm), dislocation loops were produced with Burger's vector along the <110> direction. When the simulations were extended to nanograined material, defect production was dominant near the grain boundaries. The boundaries also acted as a sink for point defects produced. The ability of MD simulations to reproduce the experimentally observed radiation response of different materials shows that the potentials are sensitive to the characteristic behavior of the material even though radiation response was not used to fit the potentials.

Govers et al. (2007) performed a comprehensive evaluation of 19 potentials for UO_2 by carrying out static calculations of elastic constants, lattice constants, and defect formation, migration, and binding energies. These values were benchmarked against experimental data or DFT results. The study considered core-shell potentials and rigid ion models. The potentials reproduced lattice constants and elastic constants well as they were fitted to these properties. The agreement with DFT data on defect energies and configurations was dependent on the type of defect considered. No single potential emerged as the best choice for simulating defects and radiation damage in UO_2 based on static calculations. The class of potential, the functional form, or the charges of the ions did not have a strong influence on the ability of the potential to reproduce defect properties. In a subsequent study, Govers et al. (2008) also studied the evolution of the lattice parameter, specific heat, bulk modulus, and Gruneisen parameter with temperature and calculated the bulk melting temperature for 16 UO_2 potentials. The study concluded that rigid ion potentials with nonformal charges provide the best agreement with experimental data but may be unsuitable for the study of charged defects.

Devanathan et al. (2009) performed MD simulations of the premelting transition, melting temperature, and defect production by 1 keV U recoils in UO_2 using five different rigid ion potentials. For all the potentials studied, the experimentally observed premelting transition was reproduced, dynamic defect annealing in the displacement cascade was highly effective, and anion sublattice replacements were dominant. The primary damage state included isolated Frenkel pairs and interstitial and vacancy clusters of various sizes. The average displacement energy varied by a

factor of three and the number of stable Frenkel pairs in the primary damage state was also different by a factor of three depending on the choice of potential. The vacancy and interstitial cluster sizes and their spatial distribution were significantly different for the potentials studied.

Interatomic potential development is an ongoing activity with tremendous intellectual effort devoted to developing new functional forms and parameter optimization methods. As computing power moves from the Petascale to the Exascale (10^{18} floating point operations per second), large scale DFT calculations will become more and more accessible. One can develop a large database of defect properties to be used in potential fitting or to evaluate the fidelity of the potentials. It will also be possible to use complex functional forms, such as ReaxFF, in large scale MD simulations of interfacial processes in multiphase materials. Recently, low-energy displacement cascades in graphite have been simulated using ReaxFF and the resulting defects have been compared to the findings of DFT calculations (Smith et al. 2017). Such reactive MD simulations are in the early stages. There is considerable room for improving potentials even in well-studied materials to make them suitable for simulation of defect processes initiated by radiation interactions with matter.

4   Summary and Outlook

Despite the tremendous progress that has been made in the last two decades in developing bond order and charge-optimized potentials, considerable work remains to be done in developing potentials for materials used in nuclear technologies. This wish list includes the UO_2/Zr interface, stainless steel, oxide dispersion strengthened alloys, Ni-based superalloys, and glasses for nuclear waste disposal. The challenges associated with potential development include generating a large database of relevant properties from experiments and DFT calculations, assigning a weight to each property of interest based on its significance (almost an art form), and painstakingly minimizing the weighted square of the difference between the simulated value and the target value (Martinez et al. 2013). Since the properties fitted are often static, dynamic simulations (e.g., heating to different temperatures and going beyond the melting point) are needed to ensure that the simulated system is stable and the phase transitions and defect properties from DFT calculations are reasonably reproduced. There is invariably a tradeoff due to balancing accuracy, transferability, computational intensity, and the effort needed to develop the potential. Increasingly the power of machine learning from massive DFT data sets is being put to use to identify new functional forms and parameters for interatomic potentials (Behler 2016). With such an approach, the force acting on an atom can be predicted by the machine-learning algorithm based on the training set or computed on-the-fly for unusual environments using an *ad hoc* DFT calculation. When the results of this calculation are added to the machine learning database, the need to call upon DFT decreases progressively as the simulation proceeds (Li et al. 2015). While this approach holds promise, there are questions about the physical

basis of functional forms chosen through machine learning and the transferability of potentials developed from a small training set. With ongoing advances in data analytics, machine learning, and computing power, there is cause for optimism that reliable and transferable potentials for studying radiation effects multiphase materials will be developed by combining the best of physical intuition-based approaches and machine learning methods.

References

Baskes MI (1987) Application of the embedded-atom method to covalent materials: a semiempirical potential for silicon. Phys Rev Lett 59(23):2666–2669

Bazant MZ, Kaxiras E, Justo JF (1997) Environment-dependent interatomic potential for bulk silicon. Phys Rev B 56(14):8542

Behler J (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145(17):170901

Bonny G, Castin N, Terentyev D (2013) Interatomic potential for studying ageing under irradiation in stainless steels: the FeNiCr model alloy. Model Simul Mater Sci Eng 21(8):085004

Brenner DW (1990) Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. Phys Rev B 42(15):9458–9471

Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys Condens Matter 14:783–802

Buckingham RA (1938) The classical equation of state of gaseous helium, neon and argon. Proc Roy Soc Lond A 168(933):264–283

Daw MS, Baskes MI (1984) Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B 29(12):6443–6453

Daw MS, Foiles SM, Baskes MI (1993) The embedded-atom method: a review of theory and applications. Mater Sci Rep 9(7–8):251–310

Debelle A, Backman M, Thome L, Nordlund K, Djurabekova F, Weber WJ, Monnet I, Pakarinen OH, Garrido F, Paumier F (2014) Swift heavy ion induced recrystallization in cubic silicon carbide: new insights from designed experiments and MD simulations. Nucl Instrum Methods Phys Res B 326:326–331

Devanathan R (2009) Radiation damage evolution in ceramics. Nucl Instrum Methods Phys Res B 267:3017–3021

Devanathan R, Yu J, Weber WJ (2009) Energetic recoils in UO_2 simulated using five different potentials. J Chem Phys 130(17):174502

Devanathan R, Weber WJ, Gale JD (2010a) Radiation tolerance of ceramics – insights from atomistic simulation of damage accumulation in pyrochlores. Energ Environ Sci 3(10): 1551–1559

Devanathan R, Van Brutzel L, Chartier A, Guéneau C, Mattsson AE, Tikare V, Bartel T, Besmann T, Stan M, Van Uffelen P (2010b) Modeling and simulation of nuclear fuel materials. Energ Environ Sci 3(10):1406–1426

Dick BG Jr, Overhauser AW (1958) Theory of the dielectric constants of alkali halide crystals. Phys Rev 112(1):90

Dong L, Li Y, Devanathan R, Gao F (2016) Molecular dynamics simulation of the structural, elastic, and thermal properties of pyrochlores. RSC Adv 6(47):41410–41419

Dudarev SL, Derlet PM (2005) A 'magnetic'interatomic potential for molecular dynamics simulations. J Phys Condens Matter 17(44):7097

Ercolessi F, Parrinello M, Tosatti E (1988) Simulation of gold in the glue model. Phil. Mag. A 58:213–226

Ercolessi F, Adams JB (1994) Interatomic potentials from first-principles calculations: the force-matching method. Europhys Lett 26(8):583–588

Farnan I, Cho H, Weber WJ (2007) Quantification of actinide α-radiation damage in minerals and ceramics. Nature 445(7124):190

Finnis MW, Sinclair JE (1984) A simple empirical N-body potential for transition metals. Philos Mag A 50(1):45–55

Gao C, Tian D, Li M, Qian D (2017) Comparative study of displacement cascades simulated with 'magnetic'potentials and Mendelev-type potential in α-Fe. J Nucl Mater 487:167–173

Gao F, Weber WJ (2002) Empirical potential approach for defect properties in 3C-SiC. Nucl Instrum Methods Phys Res, Sect B 191(1-4):504–508

Gao F, Weber WJ, Devanathan R (2002) Defect production, multiple ion-solid interactions and amorphization in SiC. Nucl Instrum Methods Phys Res B 191:487–496

Gibson JB, Goland AN, Milgram M, Vineyard GH (1960) Dynamics of radiation damage. Phys Rev 120:1229

Govers K, Lemehov S, Hou M, Verwerft M (2007) Comparison of interatomic potentials for UO2. Part I: static calculations. J Nucl Mater 366(1-2):161–177

Govers K, Lemehov S, Hou M, Verwerft M (2008) Comparison of interatomic potentials for UO2: part II: molecular dynamics simulations. J Nucl Mater 376(1):66–77

Guinan MW, Kinney JH (1981) Molecular dynamic calculations of energetic displacement cascades. J Nucl Mater 104:1319–1323

Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236

Klaver TPC, del Rio E, Bonny G, Eich SM, Caro A (2016) Inconsistencies in modelling interstitials in FeCr with empirical potentials. Comput Mater Sci 121:204–208

Lee BJ, Ko WS, Kim HK, Kim EH (2010) The modified embedded-atom method interatomic potentials and recent progress in atomistic simulations. Calphad 34(4):510–522

Lennard-Jones JE (1931) Cohesion. Proc Phys Soc 43(5):461

Lewis GV, Catlow CRA (1985) Potential models for ionic oxides. J Phys C 18(6):1149

Li Z, Kermode JR, De Vita A (2015) Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. Phys Rev Lett 114(9):096405

Liang T, Shan TR, Cheng YT, Devine BD, Noordhoek M et al (2013) Classical atomistic simulations of surfaces and heterogeneous interfaces with the charge-optimized many body (COMB) potentials. Mater Sci Engg R 74(9):255–279

Malerba L, Caro A, Wallenius J (2008) Multiscale modelling of radiation damage and phase transformations: the challenge of FeCr alloys. J Nucl Mater 382(2–3):112–125

Malerba L, Marinica MC, Anento N, Björkas C, Nguyen H, Domain C et al (2010) Comparison of empirical interatomic potentials for iron applied to radiation damage studies. J Nucl Mater 406(1):19–38

Martinez JA, Yilmaz DE, Liang T, Sinnott SB, Phillpot SR (2013) Fitting empirical potentials: challenges and methodologies. Curr Opin Solid State Mater Sci 17(6):263–270

Mayo SL, Olafson BD, Goddard WA (1990) DREIDING: a generic force field for molecular simulations. J Phys Chem 94(26):8897–8909

Mendelev MI, Ackland GJ (2007) Development of an interatomic potential for the simulation of phase transformations in zirconium. Phil Mag Lett 87(5):349–359

Mendelev MI, Han S, Srolovitz DJ, Ackland GJ, Sun DY, Asta M (2003) Development of new interatomic potentials appropriate for crystalline and liquid iron. Phil Mag 83(35):3977–3994

Minervini L, Grimes RW, Sickafus KE (2000) Disorder in pyrochlore oxides. J Am Ceram Soc 83(8):1873–1878

Moore AP, Beeler B, Deo C, Baskes MI, Okuniewski MA (2015) Atomistic modeling of high temperature uranium–zirconium alloy structure and thermodynamics. J Nucl Mater 467: 802–819

Phillpot SR, Sinnott SB (2009) Simulating multifunctional structures. Science 325(5948): 1634–1635

Polvi J, Luukkonen P, Nordlund K, Järvi TT, Kemper TW, Sinnott SB (2012) Primary radiation defect production in polyethylene and cellulose. J Phys Chem B 116(47):13932–13938

Robinson MT, Torrens IM (1974) Computer simulation of atomic-displacement cascades in solids in the binary-collision approximation. Phys Rev B 9(12):5008

Samolyuk GD, Osetsky YN, Stoller RE (2015) Molecular dynamics modeling of atomic displacement cascades in 3C–SiC: comparison of interatomic potentials. J Nucl Mater 465:83–88

Senftle TP, Hong S, Islam MM, Kylasa SB, Zheng Y, Shin YK et al (2016) The ReaxFF reactive force-field: development, applications and future directions. NPJ Comput Mater 2:15011

Shin YK, Kwak H, Zou C, Vasenkov AV, van Duin AC (2012) Development and validation of a ReaxFF reactive force field for Fe/Al/Ni alloys: molecular dynamics study of elastic constants, diffusion, and segregation. J Phys Chem A 116(49):12163–12174

Sickafus KE, Minervini L, Grimes RW, Valdez JA, Ishimaru M, Li F, McClellan KJ, Hartmann T (2000) Radiation tolerance of complex oxides. Science 289(5480):748–751

Sickafus KE, Grimes RW, Valdez JA, Cleave A, Tang M, Ishimaru M, Corish SM, Stanek CR, Uberuaga BP (2007) Radiation-induced amorphization resistance and radiation tolerance in structurally related oxides. Nat Mater 6(3):217

Smith R, Jolley K, Latham C, Heggie M, van Duin A, van Duin D, Wu H (2017) A ReaXFF carbon potential for radiation damage studies. Nucl Instrum Methods Phys Res, Sect B 393:49–53

Stillinger FH, Weber TA (1985) Computer simulation of local order in condensed phases of silicon. Phys Rev 31:5262–5271

Tersoff J (1988) New empirical approach for the structure and energy of covalent systems. Phys Rev B 37(12):6991–7000

Trachenko K, Pruneda JM, Artacho E, Dove MT (2005) How the nature of the chemical bond governs resistance to amorphization by radiation damage. Phys Rev B 71(18):184104

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174

Yablinsky CA, Devanathan R, Pakarinen J, Gan J, Severin D, Trautmann C, Allen TR (2015) Characterization of swift heavy ion irradiation damage in ceria. J Mater Res 30(9):1473–1484

Zhang J, Lang M, Ewing RC, Devanathan R, Weber WJ, Toulemonde M (2010) Nanoscale phase transitions under extreme conditions within an ion track. J Mater Res 25(7):1344–1351

Ziegler JF, Ziegler MD, Biersack JP (2010) SRIM–the stopping and range of ions in matter. Nucl Instrum Meth B 268(11–12):1818–1823

Molecular Dynamics Simulations of Non-equilibrium Systems

88

Flyura Djurabekova and Kai Nordlund

Contents

Abstract

The materials in nuclear reactors are subject to several radiation damage mechanisms. High-energy neutrons can give high (keV) recoil energies to nuclei via elastic collisions, many different nuclear reactions can also lead to lower recoil energies to nuclei, and nuclear fission fragments can get MeV kinetic energies from the fission process. The nuclei with high kinetic energies can damage the material both via nuclear and electronic excitations. In this chapter, we first overview briefly the nuclear damage mechanisms and then discuss how

F. Djurabekova (✉) · K. Nordlund
Department of Physics and Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland
e-mail: flyura.djurabekova@helsinki.fi; kai.nordlund@helsinki.fi

molecular dynamics methods can be used to model the ensuing damage, both in the nuclear and electronic collision regimes. Some examples of recent results from both the nuclear and electronic regime are provided in the end.

1   Introduction

Nuclear reactor systems involve a large number of non-equilibrium effects that can affect the materials in the reactors. The neutrons coming from a fission or fusion reactor can damage materials in many ways. When they have high (MeV) energies, they will occasionally collide strongly with nuclei in the materials and transfer a high amount of energy to these. The maximum energy can easily be estimated from the classical kinematics equation of energy transfer between particles of masses m_1 and m_2 in a head-on collision:

$$E_{transferred} = \frac{4m_1 m_2}{(m_1 + m_2)^2} E_{initial} \tag{1}$$

For a neutron energy of 2 MeV moving in Cu, this amounts to a maximum energy transfer of about 120 keV. More accurate calculations, considering that not all collisions are head-on, show that the typical transferred energies are lower but still tens of keV's (Averback and Diaz de la Rubia 1998). These energies are about four orders of magnitude higher than the cohesive energy of the material, making it clear that they can lead to considerable damage in the materials. The atomic collisions induced by such high-energy ions are known as collision cascades and the slowing down by these collisions as *nuclear stopping*.

Also neutrons at lower energies can lead to damage in the materials. Damage production by the direct collision mechanism is possible down to energies of the minimum threshold displacement energy (Nordlund et al. 2005) of the material, i.e., neutron energies of a few hundred eV. When the neutron is thermalized to even lower energies, it will eventually undergo a nuclear reaction. These reactions almost always lead to another recoil energy. For instance, in (n, γ) reactions, the emitted γ particle has enough momentum to give the nucleus a recoil energy of a few hundred eV, enough for damage production (Raman et al. 1994). Naturally, if the emitted particle has a mass, such as in (n, α) reactions, the energy transfer is even higher, and also the emitted massive particle can cause damage.

Finally, if the neutron causes a fission reaction, the fission fragments can have energies of hundreds of MeV's (note that this energy release is the origin of energy production in normal nuclear power plants). At these very high energies, the fission fragments slow down mainly by collisions with electrons, known as *electronic stopping* (Ziegler et al. 1985). On the other hand, the nuclear collision energy transfer is much reduced at these high energies, and hence the electronic stopping dominates the energy loss. Heavy ions that slow down mainly by the electronic stopping are commonly called swift heavy ions. This interaction excites a large number of electrons into the conduction band. Since the energy loss to electrons

does not alter the path of the ion appreciably, the swift heavy ions move in straight paths, and the damage they produced is in straight cylindrical regions called ion tracks (Lang et al. 2008). De-excitation of the excited electrons in the insulating materials takes place within the limited region around the ion track. The released energy can then be transferred to the atoms in the lattice, known as "electron-phonon" coupling, leading to damage production. This damage can be produced via different atomic-level mechanisms.

Naturally, both mechanisms of damage production either in collision cascades or by swift heavy ion (via nuclear or electronic stopping powers, respectively) are relevant also in other systems. Ion implantation is a routinely used approach to dope silicon in the semiconductor industry, and the ions induce collision cascades that produce damage with mechanisms very similar to the neutron-induced ones (Chason et al. 1997). Ion tracks, on the other hand, are also naturally present in many minerals, produced in these by fission fragments from radioactive decay. The density of these is used routinely as a dating technique (Gallagher et al. 1998). They can also be produced deliberately by high-energy particle accelerators and are a basis of many promising nanoscience applications (Apel 2003; Toimil-Molares 2012). Such very high-energy particles are also present in space and can be a serious issue for reliability of electronics in space (Fleetwood et al. 1994).

Experimentally, the radiation damage can be detected in many different ways, and we mention here only a few key methods. As collision cascades only exist for a few ps, the atom motion in them cannot be directly detected. However, there is a special nuclear approach that has been used to determine the lifetime of a cascade in Fe to be 6 ps (Stuchbery and Bezakova 1999). Transmission electron microscopy (TEM) observation of amorphous pockets in semiconductors (Ruault et al. 1984; Bench et al. 2000) and local phase transitions in ceramics (Meldrum et al. 1998) also provide compelling evidence of dense regions of collision associated with the nuclear stopping. Moreover, field ion microscopy experiments have shown that 20–30 keV heavy ion irradiation can lead to dense regions of vacancies in W (Current et al. 1983), and more recent transmission electron microscopy has shown that 150 keV W cascades can directly produce dislocation loops in this material (Yi et al. 2015).

Ion tracks can also be detected experimentally in many different ways. In initially crystalline materials, chemical etching can reveal the location and areal density of tracks, since the damage track core etches faster than perfect crystal (Skupinski et al. 2005). However, this method does not tell anything about the atomic structure of the tracks. Cross-sectional TEM can be used to image tracks on the atomic level, either from the sides or as a cross section. TEM experiments have shown that the core of a track is amorphous (Meftah et al. 1994; Toulemonde et al. 2006). Rather recently, small-angle x-ray scattering experiments have revealed that the core of an ion track is underdense (Kluth et al. 2008).

In this chapter, we focus on computational methods for collision cascades and ion tracks, with some emphasis on new developments during the last 10 years. We review the methods of atomistic simulations of such high-energy irradiation effects relevant in nuclear systems, both in the nuclear collision and swift heavy ion

regimes. In both cases, standard molecular dynamics simulation approaches will not work directly, since the high kinetic energies involved may lead to serious energy nonconservation issues.

2   Modeling Approaches

The computer simulation approaches, developed to simulate materials in out-of-equilibrium conditions, are similar to those used for thermodynamically equilibrated systems. The latter, however, must be adjusted to capture the physics of processes taking place under irradiation with energetic ions. Every energetic ion passing through the surface transfers the energy and the momentum to the atoms of the irradiated material in a cascade until all the energy, which was brought in by the ion, dissipates inside the material. A prolonged ion irradiation with a high fluence (number of ions arriving at a unit surface during the entire irradiation time) may cause appreciable damage; however, the experimental rates – ion fluxes given by the number of ions per unit surface per unit time – are usually very low. The actual time interval between two subsequent cascades is much longer than the duration of a single cascade, and hence, any temporal (and/or spatial) overlap of the cascades is routinely neglected. If the fluxes are very high, and the subsequent ion arrives at the surface while the previous cascade is still not extinguished, the temporal and spatial overlap of simultaneous multiple cascades may lead to strong nonlinear effects. This situation is common during arc plasma-surface interaction (Timkó et al. 2010; Djurabekova et al. 2012) and needs to be taken a special care by adjusting the size of the simulation box. However, in this chapter, we will not consider the dense ion fluxes, but rather focus on individual cascades, which can be simulated one at a time. Bearing in mind that the size of a single cascade depends on the energy of an ion, the size of the simulation box can also be selected by using a simple empirical expression:

$$N_{at} = 20 \cdot \frac{E_{kin}}{1 \, eV} \tag{2}$$

This expression guarantees that the energy can be absorbed by the simulation box and dissipated gradually in a quasi-microcanonical ensemble. The latter requires additional clarification.

If a system of atoms is in a thermodynamic equilibrium, the properties of such a system can be described as averages of different thermodynamic ensembles (Allen and Tildesley 1989), depending on the property of interest. By fixing different external conditions, one can consider the micro-canonical (*NVE*) ensemble of an isolated system with the fixed number of atoms, N, within the fixed volume, V; the canonical (*NVT*) ensemble, connected to a heat bath of temperature T; and the isobaric-isothermal ensemble (*NPT*), which is connected to the heat bath and the volume of the system is flexible to keep the pressure P at a given value.

The control of temperature or pressure affects the velocities of the atoms. If baro- and thermostats are used during irradiation event simulations, the atom trajectories are perturbed from the true ones. For a cascade to be simulated properly, one has to allow a natural evolution of energy exchange in the system. In reality, the heat generated in the system during the development of the cascade will dissipate through the heat conduction to the bulk outside the actual cascade region. However, as it was mentioned in the previous paragraph, the system to simulate a single cascade is not large enough to dissipate all the energy, and if applying a fully *NVE* ensemble in the simulations, the system will inevitably heat up. This is an artifact, which may affect the results of the cascade. To avoid it, an emulation of thermal conduction can be used, i.e., the side atoms outside of a cascade region can be connected to a heat bath, providing a channel for energy dissipation in a reasonable manner, removing the energy from the system, but not interfering with the atoms, which were directly involved in the cascade. We call such an ensemble quasi-microcanonical, since the energy is not fully conserved, although no scaling of the atom velocities is performed in the actual cascade simulation zone.

Thus far, two basic atomistic approaches have been used to simulate irradiation effects in large-scale systems with the considerable number of atoms, binary collision approximation (BCA), and molecular dynamics (MD). Both approaches employ the Born-Oppenheimer approximation, which allows separating nuclear and electronic subsystems and taking into consideration the large difference in the time and size domains of both subsystems. In this approximation, the equations of motion are applied only to the massive and heavy nuclei, while the electrons follow the former immediately. This simplifies the simulation task, and the atoms in this approximation can be considered as spherical objects interacting between themselves according to the forces. The latter are derived from potential functions, which are described in greater detail in Sect. 3.

Both BCA and MD were developed initially to simulate the materials in two different limits. The BCA approach describes the radiation effects in materials, i.e., it follows the trajectory of energetic projectile (either the ion, which entered the material, or the recoil atom, which received the energy in a cascade) and calculates collision of this projectile and the encountered atom of the material at each BCA step (see Fig. 1). BCA is designed to describe the short-range interactions, i.e., at the distances between the colliding particles which are too short to be affected by the other atoms in the material, so the collision is strictly binary and the interaction is only repulsive. Since the BCA calculates the collision between two particles at each simulation step, the simulations are very fast and do not require large computational capacity. There exist many codes developed by different groups, which either use a so-called magic formula replacing the exact solution of the collisional integral for the universal ZBL potential at every BCA step (Ziegler 2013; Pugacheva et al. 1998) or precalculate all the scattering angles for each projectile-atom pair in the system in advance and tabulate them as the function of impact parameter (Miyagawa et al. 2002). However, the BCA is not valid at low energies; already at the energies around 500 eV, the multi-body interactions between the atoms become important and determine the dynamics of defect formation. Also the BCA approach is working
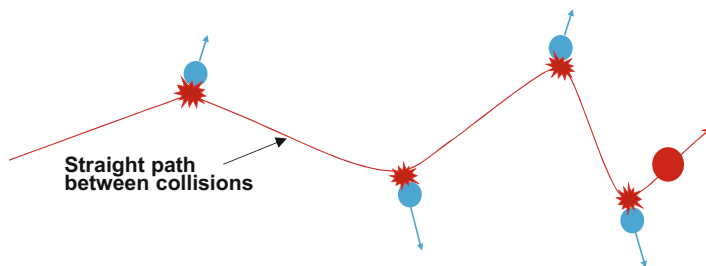
Fig. 1 Illustration of collisions in the binary collision approximation. Note that the other atoms, which are located near the knocked-on atoms, are not taken into account

well for linear cascades: the interaction between a projectile (energetic particle) and the particle in rest. Strictly speaking, the particle in rest may be involved in thermal motion, which is much less intense compared to the motion of the projectile particle. In this case, it can be neglected at the moment of collision. In the dense cascades, when many atoms are involved in the energy dissipation process (dense materials, such as metals of heavy elements), the interaction may happen between both moving atoms which leads to nonlinear cascades or thermal spikes, where the application of BCA becomes questionable.

The MD method, on the other hand, was originally developed to describe the materials in thermodynamic ensembles. Generally speaking, the MD method solves the system of Newton equations, with the driving forces derived from the potential. In other words, the interatomic potential for the material of interest is the key function, which defines the accuracy of the MD simulations. Very careful attention is usually paid to develop a potential, which is able to reproduce certain properties, which the material exhibits in thermodynamic equilibrium conditions. If the potential is able to reproduce all or some properties of materials, such as crystal structure(s), lattice parameter, cohesive energy, melting and boiling temperatures, elastic moduli, Young's and shear moduli, etc., the MD simulation of this material in a specific thermodynamic ensemble gives very reasonable results. Unfortunately, under ion irradiation, the thermodynamic equilibrium condition does not hold. In the process of momentum and energy transfer, the colliding atoms can approach each other much closer, and the potential, which was fitted to reproduce the thermodynamically equilibrated parameters, has no information about the nature of interactions at these short distances. The best solution that can be used to circumvent this problem is to spline the interatomic potential and the ZBL potential to introduce strong repulsive interactions within MD.

The other feature of MD method, which also requires special attention, is the time step, which is used in the MD iteration loop. Usually it is selected to be rather short that the atom cannot advance too far during a single MD step. In the equilibrium simulations, when the energies of atoms are varying only insignificantly, using a fixed time step is reasonable. However, in the cascade simulations, a single particle may have very high energy, which would allow it to move too far within a fixed time

step, such that it would cause the nonconservation of energy and instability in MD calculations. If the time step is fixed to a very small value, then after the ballistic phase, the simulations will proceed very slowly, and the simulation time will be unreasonably long. This is why the adaptive time step is necessary to advance the time depending on the highest value of the energy in the system (Nordlund 1995).

3   Interatomic Potentials for Simulation of Non-equilibrium Effects

Both the BCA and MD simulations rely on interatomic potentials that are discussed in detail in the chapter on Interatomic Potentials in this volume. The solution of the scattering integral in BCA relies on a repulsive interatomic potential:

$$V_{highE,rep,ij}(r_{ij}),$$

and this part also dominates strong collisions (kinetic energy roughly $>10\,\text{eV}$) in MD. The procedures to obtain these are well established. Ziegler, Biersack, and Littmark have from an averaging procedure formulated a simple analytical "universal repulsive potential" for all atom pairs Z_1–Z_2, which is typically valid to within $\sim 5\%$ in the high-energy region of interest (Ziegler et al. 1985). An alternative universal analytical potential has more recently been developed by Zinoviev (2015). Quantum chemical calculations specific to a given Z_1–Z_2 pair can be used to calculate repulsive potential accurate to within $\sim 1\%$ (Nordlund et al. 1997) which is comparable to the accuracy of the best experiments (Zinoviev and Nordlund 2017).

A typical curve illustrating the ZBL potential is shown in Fig. 2, which also illustrates one of the quantum chemical potentials that also have an attractive part (corresponding to the bound state of a dimer molecule).

The equilibrium potential required for MD simulations is far less accurately known. For simulations of ion irradiation, it is crucial to use equilibrium potentials where chemical bonds can break. Hence in this chapter, we only consider so-called "reactive" potentials, i.e., potentials where all bonds between atoms can break and reform (contrary to the case of almost all molecular mechanics force fields (Leach 2001). In general, these can be written in the form

$$V_i = \sum_j V_{rep,ij}(r_{ij}) + \sum_{jk} V_{attr,ijk}(r_{ij}, r_{ik}, \Theta_{ijk}) \tag{3}$$

Here V_i denotes the energy of atom i and j and k indices of neighboring atoms. Typically, the equilibrium potentials are constructed for near-equilibrium properties and hence are not consistent with the high-energy repulsive part $V_{highE,rep,ij}(r_{ij})$. In MD simulations that involve also high-energy collision cascades, it is thus customary to join the $V_{highE,rep,ij}(r_{ij})$ and $V_{rep,ij}(r_{ij})$ smoothly, e.g., with fifth-order spline interpolation function or by summation with a function that goes smoothly from 1 to 0 in a narrow distance interval (Nordlund et al. 1996). Usually
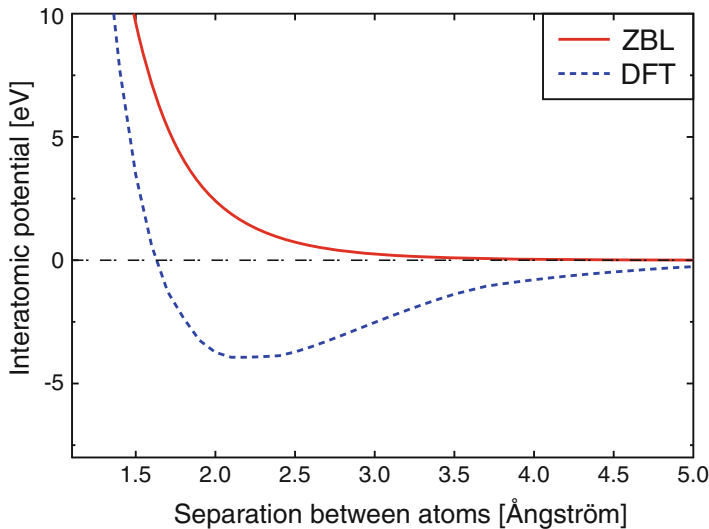
Fig. 2 Interatomic potentials for a Si-Si pair. In the legend, ZBL is used to show the shape of the universal potential developed by Ziegler et al. (1985), and DFT shows the density functional theory potential from Nordlund et al. (1997). While ZBl potential predicts only repulsive interactions at all distances, the DFT calculations show that beyond a certain distance, the interactions become attractive

this joining is done just by finding a functional form that gives a smooth connection between the two parts. However, recently the effect of the joining on damage production has been examined systematically (Sand et al. 2016), and sometimes it is carefully adjusted via DFT calculations (Belko et al. 2002; Stoller et al. 2016).

The swift heavy ion track formation induced by the electronic excitations leads to kinetic energies of atoms of the order of 10 eV (although, as evident from the discussion above, the exact mechanism of energy transfer to the atoms remains unclear, the observations of destruction of even strong covalent crystals imply that at least some atoms do get energies clearly above the cohesive energies, which are of the order of 5 eV). Hence for swift heavy ion effects above the synergy regime, it is in principle sufficient to have a good equilibrium potential. In the remainder of this section, we discuss briefly some of the classes of interatomic potentials that have been used in collision cascade and track simulations, as a comprehensive review of the vast field of interatomic potential development is beyond the aims of this chapter. Metals can be well described with so-called embedded-atom method (EAM) potentials (Daw et al. 1993). These can, via effective-medium theory, be motivated from fully quantum mechanical density functional theory (Puska et al. 1981; Manninen 1986). Originally (Foiles et al. 1986), they were formulated in the form

$$V_i = \sum_j V_{rep,ij}(r_{ij}) + F_i(\sum_j \rho_j(r_{ij})) \tag{4}$$

Here ρ_j is the electron density from atom j, and F_i is the energy that the solid gains when the ion core (metal atom without free electrons) is embedded in this electron density. Originally the EAM potentials were constructed with actual electron densities; however most recent implementations use ρ_j as fitting function of arbitrary form. In many potentials, the functional form of F_i is simply the square root function, which can be well motivated from the second-moment approximation of tight-binding electron structure models (Finnis and Sinclair 1984; Brenner 1989; Cleri and Rosato 1993; Albe et al. 2002). Regardless of exact physical motivations, all potentials that share the basic functional form can be called *EAM-like potentials*.

Regarding ionic materials, it is natural to seek a description of them based on the Coulomb potential between two different atomic charges q_i and q_j. These give directly the V_{attr} attractive part as pair potentials, while the repulsive part V_{rep} can, e.g., be a simple exponential, similar to the EAM-like potentials. For highly ionic materials near equilibrium, such a simple pair potential description is well motivated physically. However, one has to bear in mind that under irradiation, atoms can come far away from their initial positions and, e.g., can start forming N gas bubbles in some nitrides (Kucheyev et al. 2001). For describing such processes, fixed-charge ionic potentials are completely inappropriate, as they predict that the N-N interaction is completely repulsive. To circumvent this, one can (at least in principle) use charge-transfer interatomic potentials, where the charge of atoms depends on the local neighborhood, e.g., via electronegativity descriptions (Alavi et al. 1992; Albe et al. 2009; Yu et al. 2007).

In materials where the bonding is a mixture of ionic and covalent bonding, a pure pair potential is not likely to give a good description of the material. A prototypical example of this is SiO_2, which, while having a fairly high degree of ionicity, has a complex low-symmetry crystal structure with tetrahedrally arranged O atoms around the Si atoms. Such a structure cannot be described with pure pair potentials. Instead, a combination of an angular three-body term and an ionic one has proven to be suitable for describing at least the basic tetrahedral fourfold bonding of the Si atoms (Feuston and Garofalini 1988).

Materials with predominantly covalent bonding such as C, Si, Ge, and organic structures like hydrocarbons have a strong directional character of the bonds, due to the quantum mechanical hybridization of the valence electrons into sp^2 and sp^3 configurations. This hybridization leads to a preferred angle between bonds ($120°$ for sp^2 and about $109.47°$ for sp^3), and hence it is natural to formulate potentials that depend on the angle between bonds Θ_{ijk} and give an energy minimum at the preferred angle.

One widely used potential that utilizes this approach is the Stillinger-Weber potential originally made for Si (Stillinger and Weber 1985). It has in principle the form

$$V_i = \sum_j V_{2,rep,ij}(r_{ij}) + \sum_{jk} V_{3,ijk}(r_{ij}, r_{ik}, \Theta_{ijk}) \qquad (5)$$

i.e., has separate two- and three-body terms, where the two-body term has both repulsive and attractive parts and the three-body term has an explicit minimum at

$\Theta_{ijk} = 109.47°$. Although relatively simple in form and formulated more than 30 years ago, it is still widely used and been found to describe fairly well properties far removed from its initial fitting database (Balamane et al. 1992; Sastry and Angell 2003; Holmström et al. 2010). It has also been extended for a wide range of other materials (Wang and Stroud 1988; Pailthorpe and Mahon 1990; Oligschleger et al. 1996; Ichimura 1996; Aichoune et al. 2000), including giving a nonionic but rather successful description of SiO2 (Ohta and Hamaguchi 2001; Watanabe et al. 2004; Samela et al. 2008; Kluth et al. 2008). While rather successful for Si, Ge, and tetrahedral compound semiconductors, the Stillinger-Weber potential clearly is not suitable for carbon, which has the peculiar feature that the sp^2 (the bonding type in graphite and graphene) and sp^3 (diamond) bonding environments are almost identical in energy, i.e., two different bonding angles should have almost equal energy. To be able to describe several different bonding environments, Abell (1985) and Tersoff (1988) developed a formalism based on Linus Pauling's theory of the chemical bond (Pauling 1939). This formalism takes into account the observation by Pauling that the energy/bond decreases with increasing coordination number and allows having local energy minima at several different bonding configurations – including the possibility to have the exact same energy for two of them. The potential is written to formally appear like a pair potential:

$$V_i = \sum_j V_{rep,ij}(r_{ij}) + \sum_j g_{ijk} V_{attr,ij}(r_{ij}). \tag{6}$$

However, the g_{ijk} term is constructed to also depend on the angle between bonds Θ_{ijk} and the coordination number of atom i. In effect, g_{ijk} modulates the strength of the attractive part in a manner consistent with the Pauling theory. This "Tersoff-like" general form, sometimes with slight variations in the functional form of the three-body term g_{ijk}, has been parameterized for a very wide range of materials (e.g., Ashu et al. 1995; de Brito Mota et al. 1998; Matsunaga et al. 2000; Humbird and Graves 2004; Juslin et al. 2005; Billeter et al. 2007; Powell et al. 2007; Munetoh et al. 2007 and interestingly has been shown to be with certain parameter choices consistent with the Finnis-Sinclair and tight-binding-motivated EAM-like potentials (Brenner 1989; Albe et al. 2002).

The Tersoff formalism has also been extended for hydrocarbons (Brenner 1990; Brenner et al. 2002), giving an interatomic potential that is fully reactive and can describe a wide range of carbon allotropes and hydrocarbon molecules. These potentials, variably called Brenner, REBO, or AIREBO potentials, have been later implemented to include long-range forces (Stuart et al. 2000) and also include oxygen, to enable simulation of carbohydrates (Ni et al. 2004). They have been used to model several different kinds of irradiation effects, such as plasma-wall interactions (Salonen et al. 2001; Träskelin et al. 2004; Marian et al. 2007) and low-energy recoil damage in polyethylene and cellulose (Polvi et al. 2012). The Brenner carbon potential, without the complex bond-conjugation terms, has also been found to be reasonably good in describing radiation effects in carbon nanotubes and graphene (for a review of the extensive works in this area, see Krasheninnikov and Nordlund 2010).

Studying swift heavy ion effects in organic materials is challenging, in part because it is likely that much of the damage is produced by electronic effects causing bond breaking and in part because the atomic potentials for organic materials (like the Brenner one) are rather slow. However, in many cases, organic materials can be described also well by so-called coarse-grained models, in which the basic object is not single atoms, but parts of, or a full molecule. The simplest variety of such models is describing compact molecules as spheres interacting with Lennard-Jones potentials. Such models have been used to get good insights into, e.g., laser ablation (Zhigilei et al. 1998; Schäfer et al. 2002) and swift heavy ion irradiation of organic materials (Bringa et al. 2002; Bringa 2003).

4   MD Simulation of Interaction of Energetic Ions with Materials

As mentioned earlier, incident ions interact with irradiated materials in two distinct ways, either via direct collisions with atoms of the material or via excitation of electronic subsystem, which are described by nuclear, S_n, and electronic, S_e, stopping powers, respectively. Typical S_n and S_e curves are shown in Fig. 3. As one can see, both regimes are well separated to large extent; however, the energy regions where both nuclear and electronic stopping powers are significant also exist. Atomic collisions in the regime of nuclear stopping power can be simulated directly; the interaction of ions with electronic subsystem in the electronic stopping power regime (>100 keV/amu) requires development of additional multiscale models to enable these effects within a classical MD. With exception of a few key details
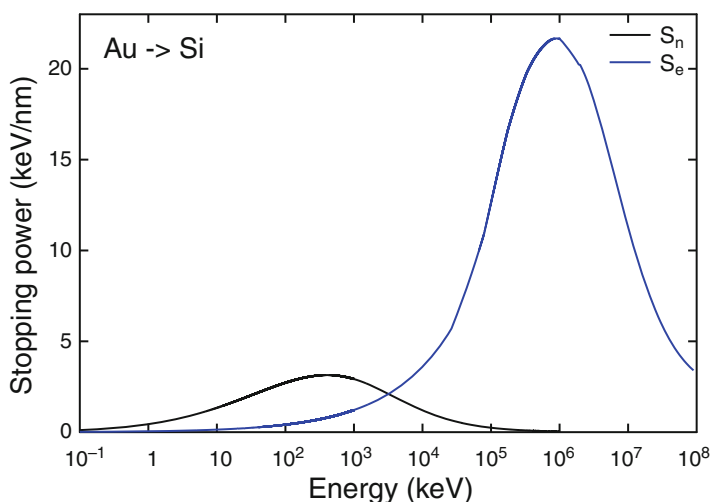
Fig. 3 Nuclear (S_n) and electronic (S_e) stopping powers in Si as a function of Au ion energy as calculated by TRIM95 software (Ziegler 1995)

which need to be paid thorough attention in the simulations of energetic atomic cascades in MD, in this chapter, we will mostly focus on the electronic effects in materials, as this is a new area of application of atomistic simulations, where the methodology is still under development.

4.1   Collision Cascades: Nuclear Stopping Power

Unlike the BCA approach, the MD simulation cell must be constructed according to the crystal or amorphous structure of the material and thoroughly equilibrated in the suitable potential that all atoms assume the sites, which correspond to the lattice site in the chosen thermodynamic ensemble. The size of the cell must be sufficiently large to ensure that the core of the cascade is in the middle of the cell and sufficiently far from the borders of the simulation cell. Introducing an energetic particle into the cell can be done either from the outside of the cell (simulating the near-surface radiation effects) or by randomly selecting an atom of the cell and giving it a recoil energy in a random direction. The latter approach will emulate the evolution of a cascade inside the bulk. Bearing in mind a small size of the MD simulation box (a couple of hundred nanometers at most), these simulations can be thought as representative of a small box cut out from the real, very large bulk system.

4.2   Electronic Stopping Power

Electronic effects are not explicitly taken into account within the classical molecular dynamics simulations. Solving the time-dependent (TD) Schrödinger equations for the atomic system with a highly energetic particle moving through became possible recently (most commonly using a TD-DFT approach) (Pruneda et al. 2007; Correa et al. 2012; Zeb et al. 2012; Ojanperä et al. 2014). Hence solving time-dependent Schrödinger equation at every step between the collisions, as well as during the collision by MD, may provide an accurate answer on energy loss to the excitations in electronic subsystem; however, to date this approach is computationally very expensive.

Instead, simplified approximations can be used to take into account the energy loss due to electronic stopping power, which agree fairly well with experiments. Following the basic idea of BCA simulations, the electronic stopping power can be implemented also in MD as a so-called frictional force. The implementation of the friction force in MD is similar to the algorithm used in SRIM, namely, the moving atom loses the energy between the collisions proportional to its velocity. The function $-(\frac{dE}{dx})_e(v)$ is obtained, e.g., from the broad SRIM database, where the experimental measurements for different materials as well as theoretical estimations, if such measurements are not available, are collected (Ziegler 2013). This has been implemented in MD by several groups (Pronnecke et al. 1991;

Nordlund 1995; Nordlund et al. 1998; Cai et al. 1998; Hou et al. 2000; Duffy and Rutherford 2007; Duvenbeck et al. 2007; Sandoval and Urbassek 2009) and demonstrated to work well by comparisons of experimental and simulated ion range distributions (Sillanpää et al. 2001; Peltola et al. 2006).

This approach, however, does not allow taking into account the energy, which may be returned back to the lattice atoms via electron-phonon coupling mechanism (Flynn and Averback 1988; Koponen 1992). The first attempt to introduce an analytical expression for calculation of cooling of the heat released in a cascade in metals by electrons was made in Finnis et al. (1991). This approach was developed later by Hou et al. (2000), deriving the parameters for the electron-phonon coupling from the Sommerfeld model, giving a simple way to include the electron-phonon coupling on atoms as a frictional force. These approaches can give insight into the role of electron-phonon coupling on collision cascade development in the nuclear stopping regime (Pronnecke et al. 1991; Björkas and Nordlund 2009). However, they are not suitable for modeling swift heavy ions, where it is crucial to know how heat is transferred from an evolving hot electron system to atoms. For this purpose, models have been developed where the electronic heat conduction is modeled concurrently with the atomic dynamics by solving the heat conduction equations as a function of space and coupling the electronic and atomic heat locally (Duffy and Rutherford 2007; Vazquez et al. 2017).

4.3   Thermostats for Simulations of Radiation Effects in Materials

Nuclear and electronic stopping powers imply that the incident energy is deposited to the irradiated material partly as potential (multiple defects remaining in the structure after the irradiation) and partly as kinetic energy. While the former form of the deposited energy is of interest as a result of the simulation, the latter may result in artificial heating of the lattice. The heating of irradiated samples is also observed in experiments if the irradiation fluxes are high and the sample is not well coupled to the surrounding, and special care is taken to keep the irradiated sample under the same temperature condition. Analogously, in molecular dynamics simulations of collision cascades or swift heavy ions, it is important to ensure that the extra heat is removed from the finite simulation cell, not to introduce an artificial heating of the system.

The standard approach to achieve this is to remove heat from the system at the boundaries, far from the high-energy recoils or ion track. This can be done in several different ways: by linear scaling (Nordlund et al. 1994) of velocities or using the Langevin (Caturla et al. 1996) or Berendsen (Nordlund et al. 1998) thermostats. Although none of these approaches can perfectly damp the pressure wave emanating from cascades, system size and parametric scaling studies can be easily used to determine whether the actual damage production results are affected by the boundary scaling (Samela et al. 2005).

5   Modeling Radiation Damage Produced by Swift Heavy Ions

In the electronic stopping power-dominated swift heavy ion (energy $\gg 10$ keV/amu) regime, the probability of individual collisions with atoms is rather negligible. Instead, the intense electronic excitation produces a narrow trail of permanent damage along the ion path – the ion track (Fleischer et al. 1975). The complex nature of electronic stopping power is discussed in detail in several publications (Itoh and Stoneham 1998; Itoh et al. 2009; Duffy et al. 2012; Schiwietz and Grande 2011; Rethfeld et al. 2014); however, none of them suggests an accurate model free of fitting parameters, which can be directly implemented in atomistic simulations.

Different mechanisms of how electronic excitations may result in permanent damage in materials, which have been proposed thus far, are the following: (i) heating of a track core through electron-phonon coupling (a two-temperature model) (Toulemonde et al. 2006; Duffy et al. 2012); (ii) Coulomb explosion, which may be caused by many electrons being rapidly excited out of the core because of the passing ion (Bringa 2003); (iii) "cold melting," i.e., electrons being excited into antibonding states causing bond breaking (Gorbunov et al. 2014; Medvedev et al. 2015a, b); and, finally, (iv) plasma formation in the track core (Cherednikov et al. 2013a, b). All models are dramatically different, in that the model (i) explains everything by rapid heating of the track core into temperatures of the order of 100,000 K, while model (iii) states that the material can be damaged even if the temperature never exceeds the melting point of about 2000 K. The models (ii) and (iv) are more specific to certain materials, where the charge relaxation may delay by a few ps.

Historically, the Coulomb explosion mechanism was the first mechanism, which was suggested to explain formation of nuclear particle tracks in minerals to analyze the records of energetic radiation in lunar rocks (Fleischer et al. 1965, 1967). Already then, the authors admitted that this model describes best the track formation in polymers, which was clearly confirmed later by Schiwietz et al. in (2004). Bringa and Johnson (2002) showed that Coulomb explosion on the early stage in combination with thermal spike on a later stage is able to jointly explain high electronic sputtering yields in polymeric materials. In other materials, the Coulomb explosion mechanism has not been confirmed; however, we mention here the attempt to explain track formation in Pd_80Si_20 metallic glass, where the authors observe plastic flow in the wake of the projectile, which may have been caused by Coulomb explosion inducing metal polarization (Klaumünzer et al. 1986). However, it is generally accepted that in many solid materials, the charge neutralization time is too short to cause any significant displacement of atoms (Duffy et al. 2012).

The model (iii), known mostly as "cold melting" or also as "lattice destabilization," assumes the modification of interatomic interactions due to electronic excitations. The equilibrium state of bonds in non-excited system will be different from that in the system, where the atoms forming the bonds are excited. If the

excitation lasts sufficiently long time, it may cause strong displacements in vicinity of the excitation, leading to fast melting (at temperatures much below the melting point) known as athermal (or sometimes also nonthermal) melting. Although very interesting, this mechanism has inherent problem of implementation in atomistic simulations as it requires a modification of interatomic potential depending on the excited state of the atoms. Some attempts to take this process into account were performed by Rymzhanov et al. in (2016) using density functional theory calculations; however, the validity of such an approach is not fully justified.

Finally, there are also several attempts to explain the formation of ion tracks by using so-called thermal spike model. This model was proposed to quantify track formation mostly in insulators as it implies thermal heat exchange between electronic and atomic subsystems. Strictly speaking, a thermal spike does not explain how the energy was deposited to electrons but rather describes the system after an electronic temperature has reached locally a thermal equilibrium. The thermal spike could result from an initial Coulomb explosion (Bringa and Johnson 2002) or changes in the interatomic forces (Itoh et al. 2009). The thermal energy of electrons is transferred then to the atoms via electron-phonon coupling, heating the lattice to the temperatures, which may cause atomic disorder (e.g., melting or boiling). Knowing the electronic stopping power and assuming a Gaussian distribution of the energy deposition around the track core and that the temperature within the track must rise, at least, to the melting point, it was possible to deduce the radius of the disordered track (Szenes 2005). Toulemonde et al. proposed to deduce the melt radius by solving analytically a coupled system of heat equations written separately for electronic (T_e) and ionic (T_i) temperatures (Toulemonde et al. 2000):

$$\begin{cases} C_e \frac{\delta T_e}{\delta t} = \nabla(K_e \nabla T_e) - G(T_e - T_i) + B(r, t) \\ \rho C_i(T_i) \frac{\delta T_e}{\delta t} = \nabla(K_i(T_i) \nabla T_i) + G(T_e - T_i) \end{cases} \qquad (7)$$

Here, $T_{e,i}$, $C_{e,i}$, $K_{e,i}$ are local temperature, specific heat capacity, and heat conduction of electronic and ionic subsystems and G is the electron-phonon coupling parameter. The function $B(r, t)$ describes the energy density supplied to electronic system during the SHI path. Integrating this function over the time and space must result in the measured value of $\frac{dE}{dx}_e$. This model is mainly known as the inelastic thermal spike (*i*-TS) model (to distinguish it from a thermal spike developed in dense collision cascades). Since the model follows the evolution of two temperatures T_e and T_i, the model is also called the two-temperature model. This two-temperature approach is based on the model which was developed by Kaganov et al. (1957) and Lifshits et al. (1960) for metals, and then, it was applied to study laser (Fujimoto et al. 1984) and SHI-induced effects also in metals (Wang et al. 1994). Application of this model to insulators is not rigorously proven; however, using fitting parameters in the two-temperature equations, good agreement with experiments was shown also for different insulators (Toulemonde et al. 2006).

6   Modeling of SHI Impacts by Using MD Methods

Since electronic structure is not explicitly described in MD algorithm, it is not immediately clear and straightforward how a SHI impact can be modeled by using this method. One of the first attempt to simulate the SHI impact in MD was made by Urbassek et al. (1994). In these simulations, a cylindrical track of a given radius was instantaneously energized by giving a constant energy, E_0, to all atoms within this track. Similar approach was used also in later papers by Bringa et al. (2002), Bringa and Johnson (2002), and Bringa (2003). In the latter works, these and other authors used the value of the stopping power $\frac{dE}{dx}_e$ integrated over the length of the simulation cell to estimate the amount of energy, which can be given to the atoms in the track to simulate more realistic impacts (Lima et al. 2016; Lan et al. 2013; Jiang et al. 2013; Lang et al. 2015; Rivera et al. 2017). The profile in these simulations can be either flat or distributed as Gaussian with the standard deviation about 1–2 nm. Although rather primitive, this model of instantaneous energy deposition can already provide valuable insights on how the lattice can accommodate the energy in a restricted volume and what is the structure of the tracks inside. For instance, it was shown that track formation in different oxides is driven by competition of heat and mass transfer, fast quenching of the molten phase, and eventual recrystallization. However, what remains unclear is the size of the restricted region where this energy has to be deposited. Usually the radius of the track is left as the free parameter, and the simulations are made for a number of combinations of E_0 and r_{tr} with results compared to the experiment (Urbassek et al. 1994; Bringa and Johnson 2002).

It is also possible to benefit from existing i-TS model and develop a multiscale approach, where the Eqs. 7 are solved numerically until the heat is maximally transferred from electronic to ionic subsystem at time t_{max} (usually around 100 fs) and, then, translate the temperature profile $T_i(r, t_{max})$ into kinetic energies of atoms in the simulation cell (Kluth et al. 2008; Pakarinen et al. 2009; Leino et al. 2015). Since the i-TS model uses the parameters of the system obtained in thermal equilibrium, it is necessary to adjust the profile $T_i(r, t_{max})$ scaling it to the melting point given by the potential in use. As it was shown, the choice of the model how the energy is deposited to the simulation cell may affect significantly the simulation results (Mookerjee et al. 2008), where significant differences in the sputtering rates as predicted by the inelastic thermal spike deposition and the cylinder model were reported.

In all aforementioned methods, it is important to set correctly the boundary conditions. For instance, periodic boundary conditions can be used to mimic the bulk case (all directions periodic), and the surface can be simulated by leaving one direction to be open (nonperiodic). Despite the choice of the boundary, the limited size of the simulation cell (finite number of atoms) causes two simulation artifacts. First of all, a strong shock wave, which is generated in the cell after the energy was introduced in the center of the track only, will return back to the track passing through periodic boundary condition. Moreover, the heat that is introduced in the cylindrical center of the cell eventually will spread in the cell and thermalize at the

temperature, which is much higher than expected if heat conduction is assumed. Hence, a special attention must be paid to mimic heat conduction effect to the bulk and to dampen the pressure waves from the track. To achieve this goal, the velocities of the atoms near the boundaries of the simulation cell must be scaled by using, e.g., Berendsen thermostat. A more advanced method to cancel the pressure wave was introduced in Zhigilei and Garrison (1999).

6.1   Two-Temperature Molecular Dynamics Model

The two-temperature MD (2TMD) model follows the temperature evolution in electronic and ionic subsystems directly within MD simulations. It can be done by using a Langevin thermostat with the friction force added to the MD equations of motion of atoms:

$$m_i \mathbf{a}_i = \mathbf{F}_i(t) + \xi m_i \mathbf{v}_i \tag{8}$$

Here, m_i, $\mathbf{a}_i$, $\mathbf{F}_i$, and $\mathbf{v}_i$ are mass, acceleration, force, and velocity of a given atom i. The value ξ is estimated as

$$\xi = \frac{V G(T_r)(T_e - T_i)}{\Sigma_i m_i v_i^2} \tag{9}$$

In this model, the simulation cell is divided into smaller cells of equal volume V, where electronic and ionic temperatures are averaged separately (see Fig. 4). Within this geometry, the boundary condition can be set to the desired temperature to imitate the heat conduction to the bulk. Moreover, the size of the ionic and electronic structure does not need to be the same. By imposing a larger mesh for electronic subsystem, the electronic temperature can be converged to the desired bulk temperature in more gradual manner. This model was initially used to simulate
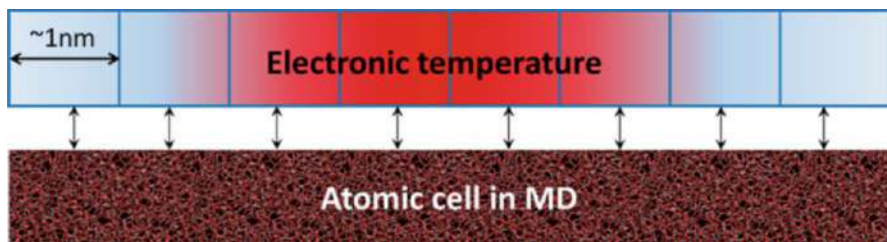
Fig. 4 Illustration of the 2TMD simulation cell. For clarity the grids of electronic and ionic subsystem are shown separately. In each cell of the grid, a local electronic and ionic temperature is defined. Dirichlet boundary condition is applied at the sides of the electronic temperature grid (light blue color). Size of the grid point should be sufficiently large to enable statistically meaningful values of local temperature

the short-pulse laser melting (Ivanov and Zhigilei 2003) but later on was extended to be used to simulate the electronic effects in radiation damage simulations (Duffy and Rutherford 2007).

6.2   Parameterization of the Inelastic Thermal Spike Model

Advantage of one implementation of the inelastic thermal spike model over the other, such as an instantaneously deposited energy or by a natural evolution within the 2TMD model, depends on the specific problem and the studied material. However, both implementations rely heavily on the parameters used in the model.

First of all, what remains ambiguous is how the energy is deposited to the electrons by a SHI. In the original model by Toulemonde et al., the initial distribution of energy is described by the expression suggested by Waligórski et al. (1986). This distribution is then scaled to give stopping power predicted by the SRIM software (Ziegler 2013).

There are more theoretical attempts to reproduce initial electronic cascades, especially in the short-pulse laser irradiation community (Rethfeld 2004), later on adopted for simulations of electronic effects during a SHI impact on materials by using Monte Carlo approach (Medvedev et al. 2010, 2015a, b). These types of simulations may provide a more realistic distribution of initial energy deposition, giving deeper insight on the energy distribution taken away from the track by fast δ-electrons.

Furthermore, it is not clear how to estimate parameters of electronic subsystem in such a far-from-equilibrium condition. Originally, in the Toulemonde's model (Toulemonde et al. 2000), the free-electron gas model is used to estimate the electronic heat capacity ($C_e = \frac{3}{2} N_e k_B$, where N_e is the electron density and $k_B b$ is the Boltzmann constant) and conductivity ($K_e = D_e C_e = (\frac{1}{3} l v_f C_e)$ with D_e being the electron diffusivity and l and v_f the electron mean free path and Fermi velocity, respectively.) The electron-phonon coupling constant remained in the Toulemonde's model as a free parameter adjusted to result in the track radius observed in experiments.

The free-electron gas model leads to severe underestimation of the C_e parameter, since the electronic temperature may rise during the SHI impact by several orders of magnitude. In the limit when the electronic temperature homogenized, but still $T_e \gg T_i$, the Eqs. 7 can be simply rewritten as

$$C_e \frac{\delta T_e}{\delta t} = G \cdot (T_e - T_i) \tag{10}$$

Solving this equation gives a clear relationship between G and C_e:

$$\tau = \frac{C_e}{G} \tag{11}$$
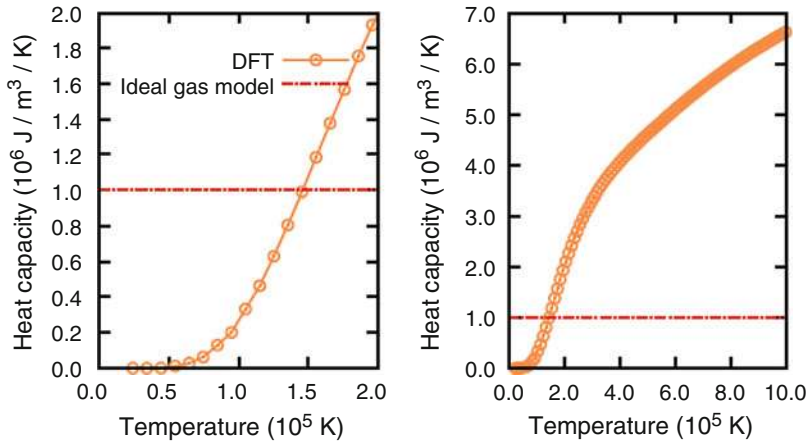
Fig. 5 Electronic temperature dependence of the specific heat capacity $C_e(T_e)$ as calculated with DFT. The dash-dotted line shows the commonly used C_e from the free-electron gas model assuming two excited electrons (Toulemonde et al. 2000). The left image shows the same data zoomed in at lower temperatures. (Reprinted with permission from Leino et al. (2015). Copyright (2018) by the EDP sciences)

where τ is the relaxation time, which can be obtained from pump-probe experiments by direct measurement of the electron temperature after laser pulse using angle-resolved photoemission spectroscopy (Johannsen et al. 2013).

Daraszewicz et al. (2014) suggested the recipe to calculate the electron-phonon coupling from the first principles for W. However, this is less straightforward, if the studied material is an insulator. In this case, the specific heat capacity still can be calculated from $C_e = \frac{\delta E}{\delta T_e}$, where modification of the internal energy E is found from the analysis of the electronic density of states (DOS) obtained in finite-temperature generalization of the density functional theory (DFT) using Quantum Espresso (Giannozzi et al. 2009). If the value of the relaxation time is known experimentally, the strength of electron-phonon coupling can be estimated from Eq. 11. Figure 5 shows the difference between the electronic specific heat capacity calculated by using Quantum Espresso and the free-electron gas model (temperature-independent), where it is clear that C_e at low temperature is strongly overestimated, while at high electronic temperature, it must be multifold higher.

7   Simulations of Track Formation in Bulk Materials

As described in Sect. 5, simulation methods, which exist to date to simulate SHI impacts on materials, are still under development. However, already initial attempts to simulate this complex phenomenon with the available tools have resulted in a few very valuable insights, and we review here briefly a few significant examples.

The first experimental measurements of tracks in amorphous materials, e.g., a-SiO$_2$, by using small-angle X-ray scattering spectroscopy (SAXS), indicated the existence of a core-shell structure. By then, it was clear that the nature of ion track damage ranges from small differences between track and the rest of material in amorphous structures (e.g., SiO$_2$ Klaumünzer 2004), formation of point defect and defect clusters in ionic crystals (Trautmann et al. 2000; Khalfaoui et al. 2005), to amorphization of crystalline materials (e.g., InP Kamarou et al. 2008; Klaumünzer 2004; Meftah et al. 1994). However, the new finding clearly showed that the subtle difference in densities appears even within a track region itself. Although it was clear that in the core of the track the density was different from that of a periphery, it was not obvious whether the ratio was toward the overdense shell and underdense core or the other way around. The insight on the track formation was possible through MD simulations. A few profiles obtained for different energy depositions with profiles calculated by using Toulemonde's *i-TS* model (7) are shown in comparison to the experimental profile reconstructed from the SAXS signals. It is clear that the ion track leaves a clear signature of an underdense core surrounded by the overdense shell (see Fig. 6). In the same figure, we demonstrate an excellent agreement between the size of the track obtained in crystalline quartz by MD simulations using the energy loss of 15.9 keV/nm and the TEM image of the same track, but using energy loss of 14 keV/nm (Meftah et al. 1994).

The mechanism revealed in these simulations can be summarized as shown in Fig. 7. The energy deposited to the electronic subsystem in a SHI impact is
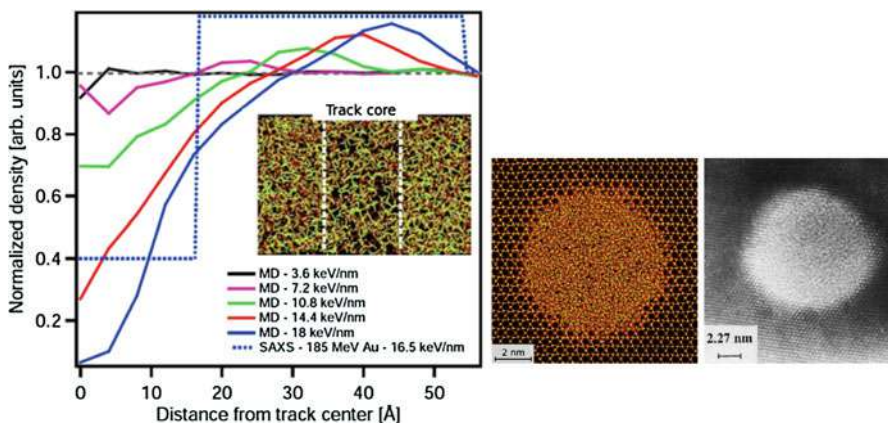
Fig. 6 Left: Radial density profiles obtained in MD simulations for a-SiO$_2$ shown in arbitrary units. Unmodified density of the cell is shown with a dashed line at 1.0. The dotted line shows for comparison a density profile extracted from the SAXS measurements of the 185 MeV Au (Kluth et al. 2008). The inset shows a snapshot of MD simulations at the moment when the track has already cooled down to its final state. Right: Comparison of track sizes in crystalline quartz obtained in MD simulations with energy loss $S_e = 15.9$ eV/nm and in experimental TEM image with $S_e = 14$ keV/nm from Meftah et al. (1994). (Reprinted from Pakarinen et al. (2009), Copyright (2018), with permission from Elsevier)
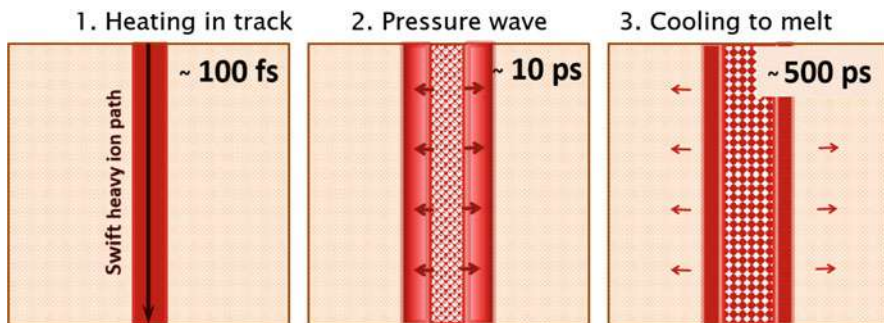
Fig. 7 Mechanism of track formation in amorphous materials. The track is heated in the narrow region around the swift heavy ion path. In the following few picoseconds, the pressure waves formed due to extremely high-temperature gradient that transport the material from the track core to the periphery of the track. The intense heat exchange with the infinite thermal bath of a surrounding material leads to the temperature quench, and the structure freezes in the density-alternating state

subsequently transferred to the lattice via the electron-phonon coupling. Large amount of energy deposited nearly instantaneously (within a few hundreds of femtoseconds) in a limited volume of an ion track leads to very fast phase transition within the track to a high-pressure vapor phase of the material, generating strong pressure waves, which transport the material from the track center. The cold material surrounding the track resists the wave propagation pressing the material to a higher density state. At the same time, the efficient heat exchange between the small volume of a heated track and cold, but infinitely large, surrounding bulk material dissipates the heat from the track freezing in the inhomogeneous density distribution around the track. Since a-SiO$_2$ is known to be highly viscous, the density variation remains as a permanent track after the impact region cooled down to the room temperature.

The combination of the *i*-TS model by Toulemonde (Eqs. 7) and MD simulations would lead to a monotonic increase of the track radius approximately as the square root function of the deposited energy, which can be intuitively expected. However, this was not confirmed in the experiments (Afra et al. 2013), where it was found that the track radius saturates and does not exceed a certain value for the given material. Intriguingly, in the other experiments for the same material, a-SiO$_2$, the saturation of the track radius with increase of the stopping power was not observed while using a different technique, the Rutherford backscattering (RBS) (Meftah et al. 1994). Having in mind that specific heat capacity $C_e = \frac{\delta E}{\delta T_e}$ depends on the electron density of states, which are different at different electronic temperatures, one can calculate $C_e(T_e)$ as shown in Fig. 5. Plugging in this function to the full 2TMD model, the results show that, indeed, the track radius saturates with increase of the stopping power. This is explained by rapid growth of the C_e at high electronic temperatures and thus the slower rise of the temperature in the track. Now, if we turn our sight to the defects, which are created out of the track radius, we see that the distance from the track center to the individual defects keeps increasing with increase of

the stopping power, although the track radius has already been saturated. Since the higher stopping power would produce higher-energy δ-electrons, the individual defects formed in the electronic cascades initiated by the δ-electrons may locate further away from the track. This can explain the discrepancy between the SAXS (Afra et al. 2013) and RBS (Meftah et al. 1994) experiments, since the former technique is sensitive to change of densities in the track but the latter measures the defects and disordered regions. In Fig. 8, the MD track radii measured by analyzing the density variations and the furthest defects found in the MD cell away from the center of the track are compared to the SAXS and RBS results. Although the MD track radii obtained by measuring the furthest defects are still somewhat smaller than the RBS signal, the growing trend is clear when it is compared to the saturated trend of the track radii obtained by measuring the density variation in the track.

Effect of the parameterization of Eqs. 7 was also clearly observed in SrTiO$_3$ in Weber et al. (2014), where the reduction of the thermal conductivity of the electronic and ionic subsystems was associated with defects in a predamaged SrTiO$_3$ layer. By reducing these parameters by the order of magnitude, the authors found that track in the predamaged SrTiO$_3$ layer (with up to 1.5% of Frenkel pairs in the structure)
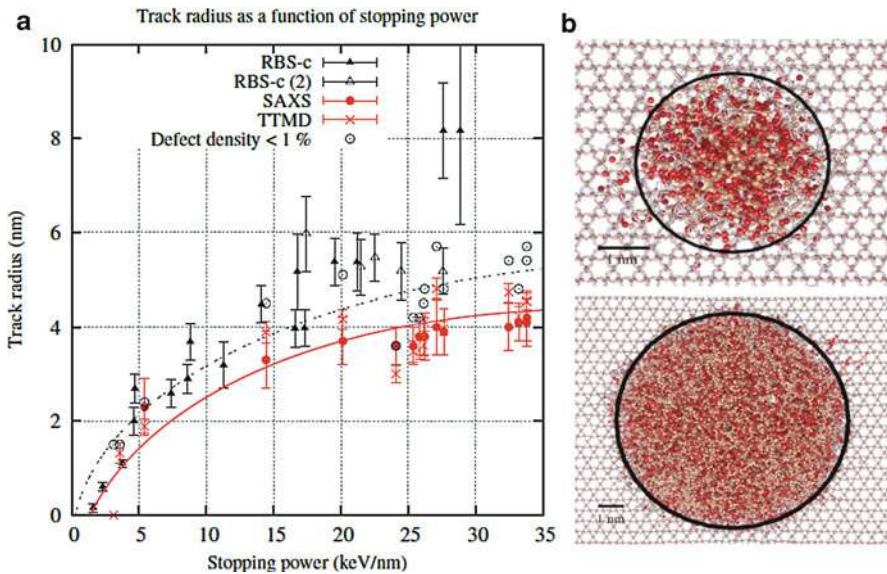
Fig. 8 (**a**) Track radius as a function of stopping power. The experimental data are from Meftah et al. (1994), Tombrello (1984), and Afra et al. (2013). The lines are polynomial fits to the simulation data points to guide the eye. The empty circles indicate a radius at which the defect concentration falls below 1% in the simulations; (**b**) Snapshots of tracks obtained with lower-energy ions (27 MeV Au, top) and at the saturation region (1.4 GeV Au, bottom). The circles are the track radii that are obtained by fitting density variation. Defected atoms are drawn as large spheres. (Reprinted with permission from Leino et al. (2015). Copyright (2018) by the EDP sciences)

was larger and compared much better to the tracks seen in the experiment. On the other hand, the SHI tracks were found to cause a healing effect in the predamaged SiC via the so-called SHIBIEC effect, which stands for swift heavy ion-induced epitaxial recrystallization (Debelle et al. 2012, 2014). In these experiments, the ideal zinc blende structure of 3C-SiC was predamaged by 100 keV Fe ions at room temperature with the fluences of 2×10^{14} cm^2 and 4×10^{14} cm^2 to achieve partial and full amorphization of a layer of the 3C-SiC sample. After introducing SHI tracks according to Eqs. 7 generated by 0.87 GeV Pb ions, the damage existed in 3C-SiC partially recovered due to temperature-induced recrystallization at the interface between crystal-amorphized phases of 3C-SiC. The lower the degree of amorphization, the stronger recovery was achieved in both experiments and simulations, as shown in Fig. 9.

One more interesting phenomenon, which was explained by using MD simulations of ion tracks, was the formation of "bowtie"-shaped voids in amorphous germanium. In this material, the heating induced by a SHI track has led to a series of phase transitions within the track region (Ridgway et al. 2013; Hooda et al. 2017). Simulation of track evolution in amorphous germanium revealed the stages schematically illustrated in Fig. 10. First two stages are similar to track dynamics in other insulating materials: during the first 100 fs, the track core rapidly heats up, and the strong pressure waves transport the material away from the center of the track during the following 3 ps. While cooling down from a vapor to the liquid phase, a phase transition occurs: the hot low-density liquid germanium turns into much denser cooler liquid germanium phase. This transition creates empty space, and a multitude of small voids form along the track during another 10 ps. For more than 30 ps, the temperature in the track is sufficient to enable the coalescence of
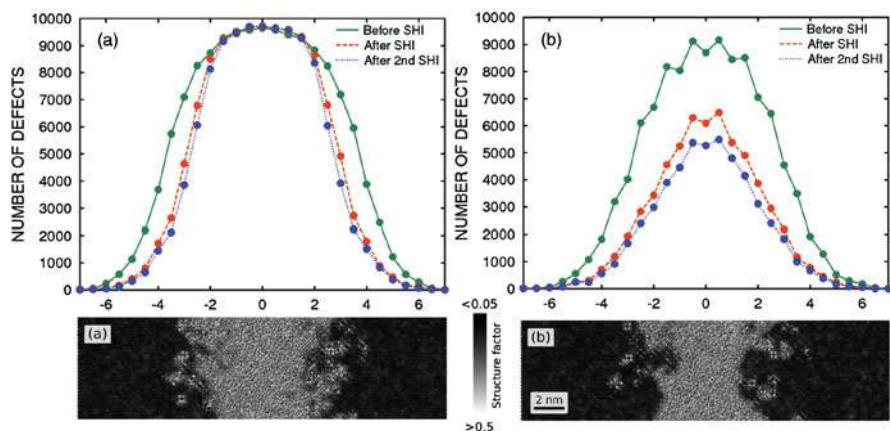
Fig. 9 Simulated depth distributions of the damage in the 3C-SiC cells damaged by low-energy recoils and subsequently heat treated by a thermal spike caused by 0.87 GeV Pb ions Case of (**a**) fully amorphous and (**b**) partially amorphous. The corresponding images of the analysis of angular structure factors show atomistic image of the recovered buried layer of the predamaged 3C-SiC
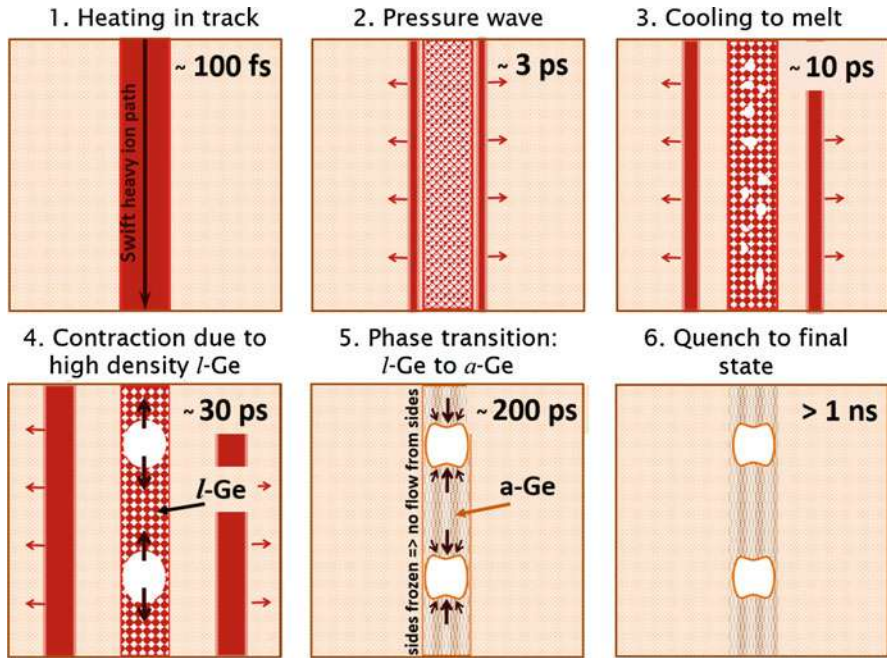
Fig. 10 Illustration of the mechanism of formation of "bowtie" voids in amorphous germanium. Hot track vaporizing Ge atoms in a limited volume (stage 1), pressure waves are transporting material away from the track center (stage 2). In the stage 3, the hot low-density liquid germanium changes to high-density cold germanium melt, which leads to formation of small voids. The voids agglomerate in larger voids during the stage 4. Continuous cooling leads to another phase transition from high-density liquid to low-density amorphous phase. The frozen high-density shell prevents expansion of material outward, and the voids contract only along the track axis, where the temperature still allows the atom movement during the stage 5. At the last stage 6, the formed bowtie shape freezes in

small voids into larger ones, which gradually contract along the track axis due to expansion of the amorphous Ge after a reverse phase transition from the liquid to amorphous state, whose density is known to be lower. During this last transition, the void assumes the final "bowtie"-shape that remains unchanged longer than 1 ns. The comparison of the simulated void in amorphous Ge and the TEM image of the void obtained with the same irradiation condition are shown in Fig. 11.

8   Conclusions

In this chapter, we reviewed briefly the basics of physics of far-from-equilibrium processes, which take place during modification of materials by energetic ions or recoils in the wide range of ion energies. We also discuss how molecular dynamics methods can be modified to model realistically the ensuing damage, created by
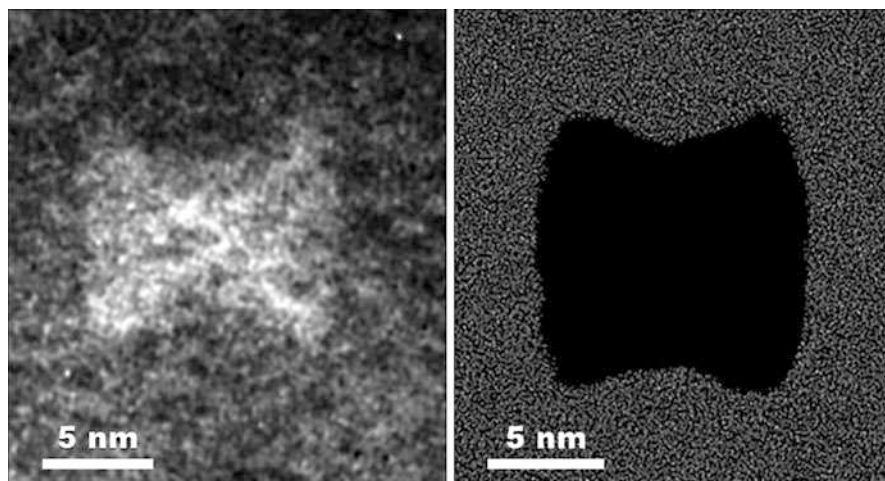
Fig. 11 On the left, a TEM image of a bowtie-shaped void formed in amorphous Ge by swift heavy ion irradiation. On the right, the void simulated by using *i*-TS model in combination with classical MD simulations during the same radiation conditions. (Reprinted with permission from Ridgway et al. (2013). Copyright (2018) by the American Physical Society)

ions via nuclear collisions and the electronic excitation regime. Examples of recent insights, which became available owing to newly developed multiscale models based on classical MD methods, are also given at the end of the chapter. These examples are focused on damage generated by swift heavy ions via interactions with electronic subsystem of the irradiated materials.

References

Abell GC (1985) Empirical chemical pseudopotential theory of molecular and metallic bonding. Phys Rev B 31:6184

Afra B, Rodriguez MD, Trautmann C, Pakarinen OH, Djurabekova F, Nordlund K, Bierschenk T, Giulian R, Ridgway MC, Rizza G, Kirby N, Toulemonde M, Kluth P (2013) Saxs investigations of the morphology of swift heavy ion tracks in α-quartz. J Phys Condens Matter 25:0455006

Aichoune N, Potin V, Ruterana P, Hairie A, Nouet G, Paumier E (2000) An empirical potential for the calculation of the atomic structure of extended defects in wurtzite GaN. Comput Mater Sci 17:380

Alavi A, Alvarez LJ, Elliott R, MacDonald IR (1992) Charge-transfer molecular dynamics. Philos Mag B 65:489

Albe K, Nordlund K, Averback RS (2002) Modeling metal-semiconductor interaction: analytical bond-order potential for platinum-carbon. Phys Rev B 65:195124

Albe K, Nord J, Nordlund K (2009) Dynamic charge-transfer bond-order potential for gallium nitride. Philos Mag A 89:3477–3497

Allen MP, Tildesley DJ (1989) Computer simulation of liquids. Oxford University Press, Oxford

Apel P (2003) Swift ion effects in polymers: industrial applications. Nucl Instr Methods Phys Res B 208:11

Ashu PA, Jefferson JH, Cullis AG, Hagston WE, Whitehouse CR (1995) Molecular dynamics simulation of (100) InGaAs/GaAs strained-layer relaxation processes. J Cryst Growth 150: 176–179

Averback RS, Diaz de la Rubia T (1998) Displacement damage in irradiated metals and semiconductors. In: Ehrenfest H, Spaepen F (eds) Solid state physics, vol 51. Academic, New York, pp 281–402

Balamane H, Halicioglu T, Tiller WA (1992) Comparative study of silicon empirical interatomic potentials. Phys Rev B 46(4):2250

Belko V, Posselt M, Chagarov E (2002) Improvement of the repulsive part of the classical interatomic potential for SiC. Nucl Instrum Methods Phys Res B 202:18–23

Bench MW, Robertson IM, Kirk MA, Jenčič I (2000) Production of amorphous zones in GaAs by the direct impact of energetic heavy ions. J Appl Phys 87(1):49–56

Billeter SR, Curioni A, Fischer D, Andreoni W (2007) Ab initio derived augmented Tersoff potential for silicon oxynitride compounds and their interfaces with silicon. Phys Rev B 73:155329

Björkas C, Nordlund K (2009) Assessment of the relation between ion beam mixing, electron-phonon coupling, and damage production in Fe. Nucl Instrum Methods Phys Res B 267: 1830–1836

Brenner D (1989) Relationship between the embedded-atom method and Tersoff potentials. Phys Rev Lett 63:1022

Brenner DW (1990) Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. Phys Rev B 42(15):9458

Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Sinnott SB (2002) A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys Condens Matter 14:783

Bringa EM (2003) Molecular dynamics simulations of Coulomb explosion. Nucl Instrum Methods Phys Res B 209:1–9

Bringa EM, Johnson RE (2002) Coulomb explosion and thermal spikes. Phys Rev Lett 88(16):165501

Bringa EM, Johnson RE, Papaleo RM (2002) Crater formation by single ions in the electronic stopping regime: comparison of molecular dynamics simulations and experiments on organic films. Phys Rev B 65:094113

Cai D, Snell CM, Beardmore KM, Grønbech-Jensen N (1998) Simulation of phosphorus implantation into silicon with a single parameter electronic stopping power model. Int J Mod Phys C 9(3):459

Caturla MJ, Diaz de la Rubia T, Marques LA, Gilmer GH (1996) Ion-beam processing of silicon at keV energies: a molecular-dynamics study. Phys Rev B 54(24):16683

Chason E, Picraux ST, Poate M, Borland JO, Current MI, Diaz de la Rubia T, Eaglesham DJ, Holland OW, Law ME, Magee CW, Mayer JW, Melngailis J, Tasch AF (1997) Ion beams in silicon processing and characterization. J Appl Phys 81(10):6513–6561

Cherednikov Y, Sun S, Urbassek H (2013a) Hybrid paricle-in-cell/molecular dynamics simulation of swift heavy ion tracks in LiF. Phys Rev B 87(24):245424

Cherednikov Y, Sun S, Urbassek H (2013b) Sputtering from swift-ion trails in lif: a hybrid pic/md simulation. Nucl Instrum Methods Phys Res B 315:313–317

Cleri F, Rosato V (1993) Tight-binding potentials for transition metals and alloys. Phys Rev B 48(1):22

Correa AA, Kohanoff J, Artacho E, Sanchez-Portal D, Caro A (2012) Nonadiabatic forces in ion-solid interactions: the initial stages of radiation damage. Phys Rev Lett 108(21):213201

Current MI, Wei CY, Seidman DN (1983) Direct observation of the primary state of damage of ion-irradiated tungsten II. Definitions, analyses and results. Philos Mag A 47(3):407

Daraszewicz S, Giret Y, Tanimura H, Duffy D, Shluger A, Tanimura K (2014) Determination of the electron-phonon coupling constant in tungsten. Appl Phys Lett 105:023112. http://scitation. aip.org/content/aip/journal/apl/105/2/10.1063/1.4890413

Daw MS, Foiles SM, Baskes MI (1993) The embedded-atom method: a review of theory and applications. Mat Sci Rep 9:251

Debelle A, Backman M, Thomé L, Weber WJ, Toulemonde M, Mylonas S, Boulle A, Pakarinen OH, Juslin N, Djurabekova F, Nordlund K, Garrido F (2012) Combined experimental and computational study of the recrystallization process induced by electronic interactions of swift heavy ions with silicon carbide crystals. Phys Rev B 86:100102(R). https://doi.org/10.1103/PhysRevB.86.100102

Debelle A, Backman M, Thome L, Nordlund K, Djurabekova F, Weber WJ, Monnet I, Pakarinen OH, Garrido F, Paumier F (2014) Swift heavy ion induced recrystallization in cubic silicon carbide: new insights from designed experiments and md simulations. Nucl Instrum Methods Phys Res B 326:326–331. https://doi.org/10.1016/j.nimb.2013.10.080

de Brito Mota F, Justo JF, Fazzio A (1998) Structural properties of amorphous silicon nitride. Phys Rev B 58:8323

Djurabekova F, Samela J, Timko H, Nordlund K, Calatroni S, Taborelli M, Wuensch W (2012) Crater formation by single ions, cluster ions and ion "showers". Nucl Instrum Methods Phys Res B 272:374–376. https://doi.org/10.1016/j.nimb.2011.01.104

Duffy DM, Rutherford AM (2007) Including the effects of electronic stopping and electron-ion interactions in radiation damage simulations. J Phys Condens Matter 19:016207

Duffy DM, Daraszewicz S, Mulroue J (2012) Modelling the effects of electronic excitations in ionic-covalent materials. Nucl Instrum Methods Phys Res B 277:21–27

Duvenbeck A, Weingart O, Buss V, Wucher A (2007) Electron promotion and electronic friction in atomic collision cascades. New J Phys 9:38

Feuston BP, Garofalini SH (1988) Empirical three-body potential for vitreous silica. J Chem Phys 89(9):5818

Finnis MW, Sinclair JE (1984) A simple empirical N-body potential for transition metals. Philos Mag A 50(1):45. See also Erratum, ibid. 53:161 (1986)

Finnis MW, Agnew P, Foreman AJE (1991) Thermal excitation of electrons in energetic displacement cascades. Phys Rev B 44(2):44

Fleetwood D, Kosier S, Nowlin R, Schrimpf R, Reber R, Delaus M, Winokur P, Wei A, Combs W, Pease R (1994) Physical-mechanisms contributing to enhanced bipolar gain degradation at low-dose rates. IEEE Trans Nucl Sci 41(6):1871–1883. 1994 IEEE annual conference on nuclear and space radiation effects (NSREC 94), Tucson, 18–22 July 1994

Fleischer R, Price P, Walker RM (1965) Ion explosion spike mechanism for formation of charged particle tracks in solids. J Appl Phys 36:3645

Fleischer RL, Walker RM, Price PB, Hubbard EL (1967) Criterion for registration in dielectric track detectors. Phys Rev 156:353

Fleischer RL, Price PB, Walker RM (1975) Nuclear tracks in solids. University of California, Berkeley

Flynn CP, Averback RS (1988) Electron-phonon interactions in energetic displacement cascades. Phys Rev B 38:7118

Foiles SM, Baskes MI, Daw MS (1986) Embedded-atom-method functions for the FCC metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. Phys Rev B 33(12):7983. Erratum: ibid, Phys Rev B 37:10378 (1988)

Fujimoto JG, Liu JM, Ippen EP, Bloembergen N (1984) Femtosecond laser interaction with metallic tungsten and nonequilibrium electron and lattice temperatures. Phys Rev Lett 53:1837–1840

Gallagher K, Brown R, Johnson C (1998) Fission track analysis and its applications to geological problems. Annu Rev Earth Planet Sci 26:519–572

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Corso AD, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Seitsonen GSAP, Smogunov A, Umari P, Wentzcovitch RM (2009) Quantum espresso: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502-1–395502-19

Gorbunov SA, Medvedev NA, Rymzhanov R, Terekhind PN, Volkov AE (2014) Excitation and relaxation of olivine after swift heavy ion impact. Nucl Instrum Methods Phys Res B 326: 163–168

Holmström E, Krasheninnikov AV, Nordlund K (2010) Quantum and classical molecular dynamics studies of the threshold displacement energy in Si bulk and nanowire. In: Ila D, Lindner JKN, Chu PK, Baglin J, Kishimoto N (eds) Ion beams and nano-engineering, MRS symposium proceedings, vol 1181, MRS, Warrendale, pp 111–122

Hooda S, Avkhachev K, Khan S, Djurabekova F, Nordlund K, Satpati B, Bernstorff S, Ahlawat S, Kanjilal D, Kabiraj D (2017) Mechanistic details of the formation and growth of nanoscale voids in Ge under extreme conditions within an ion track. J Phys D: Appl Phys 50:225302

Hou Q, Hou M, Bardotti L, Prevel B, Melinon P, Perez A (2000) Deposition of AuN clusters on Au(111) surfaces. I. Atomic-scale modeling. Phys Rev B 62(4):2825–2834

Humbird D, Graves DB (2004) Molecular dynamics simulations of Si-F surface chemistry with improved interatomic potentials. Plasma Sources Sci Technol 13(3):548

Ichimura M (1996) Stillinger-Weber potentials for III–V compound semiconductors and their application to the critical thickness calculation for InAs/GaAs. Physica Status Solidi A 153(2):431–437

Itoh N, Stoneham A (1998) Excitonic model of track registration of energetic heavy ions in insulators. Nucl Instrum Methods Phys Res B 146:362–366

Itoh N, Duffy D, Khakshouri S, Stoneham A (2009) Making tracks: electronic excitation roles in forming swift heavy ion tracks. J Phys: Condens Matter 21:474205. https://doi.org/10.1088/0953-8984/21/47/474205

Ivanov DS, Zhigilei LV (2003) Combined atomistic-continuum modeling of short pulse laser melting and disintegration of metal films. Phys Rev B 68:064114

Jiang W, Devanathan R, Sundgren C, Ishimaru M, Sato K, Varga T, Manandhar S, Benyagoub A (2013) Ion tracks and microstructures in barium titanate irradiated with swift heavy ions: a combined experimental and computational study. Acta Materialia 61:7904–7916

Johannsen JC, Ulstrup S, Cilento F, Crepaldi A, Zacchigna M, Cacho C, Turcu ICE, Springate E, Fromm F, Raidel C, Seyller T, Parmigiani F, Grioni M, Hofmann P (2013) Direct view of hot carrier dynamics in graphene. Phys Rev Lett 111:027403

Juslin N, Erhart P, Träskelin P, Nord J, Henriksson KOE, Nordlund K, Salonen E, Albe K (2005) Analytical interatomic potential for modelling non-equilibrium processes in the W-C-H system. J Appl Phys 98:123520

Kaganov MI, Lifshitz IM, Tanatarov LV (1957) Relaxation between electrons and crystalline lattice. Sov Phys JETP 4(2):173

Kamarou A, Wesch W, Wendler E, Undisz A, Rettenmayr M (2008) Radiation damage formation in inp, insb, gaas, gap, ge, and si due to fast ions. Phys Rev B 78:054111

Khalfaoui N, Rotaru CC, Bouffard S, Toulemonde M, Stoquert JP, Haas F, Trautmann C, Jensen J, Dunlop A (2005) Characterization of swift heavy ion tracks in CaF$_2$ by scanning force and transmission electron microscopy. Nucl Instrum Methods Phys Res B 240(4):819–828. https://doi.org/10.1016/j.nimb.2005.06.220

Klaumünzer S (2004) Ion tracks in quartz and vitreous silica. Nucl Instrum Methods Phys Res B 225:136

Klaumünzer S, Hou M-d, Schumacher G (1986) Coulomb explosions in a metallic glass due to the passage of fast heavy ions? Phys Rev Lett 57:850–853

Kluth P, Schnohr CS, Pakarinen OH, Djurabekova F, Sprouster DJ, Giulian R, Ridgway MC, Byrne AP, Trautmann C, Cookson DJ, Nordlund K, Toulemonde M (2008) Fine structure in swift heavy ion tracks in amorphous SiO$_2$. Phys Rev Lett 101:175503

Koponen I (1992) Atomic mixing in ion-bombardment-induced temperature spikes in metals. J Appl Phys 72(3):1194

Krasheninnikov AV, Nordlund K (2010) Ion and electron irradiation-induced effects in nanostructured materials. J Appl Phys 107:071301

Kucheyev SO, Williams JS, Jagadish C, Zou J, Li G, Titov AI (2001) Effect of ion species on the accumulation of ion-beam damage in GaN. Phys Rev B 64:035202

Pauling L (1939) The nature of the chemical bond. Cornell University Press, Ithaca

Lan CH, Xue JM, Wang YG, Zhang YW (2013) Molecular dynamics simulation of latent track formation in α-quartz. Chin Phys C 37:783

Lang M, Lian J, Zhang F, Hendriks BW, Trautmann C, Neumann R, Ewing RC (2008) Fission tracks simulated by swift heavy ions at crustal pressures and temperatures. Earth Planet Sci Lett 274:355

Lang M, Devanathan R, Toulemonde M, Trautmann C (2015) Advances in understanding of swift heavy-ion tracks in complex ceramics. Curr Opin Solid State Mater Sci 19:39–48

Leach AR (2001) Molecular modelling: principles and applications, 2nd edn. Pearson Education, Harlow

Leino AA, Daraszewicz SL, Pakarinen OH, Nordlund K, Djurabekova F (2015) Atomistic two-temperature modelling of ion track formation in silicon dioxide. EPL 110(1):16004. http://stacks.iop.org/0295-5075/110/i=1/a=16004

Lifshits IM, Kaganov MI, Tanatarov LV (1960) On the theory of radiation-induced changes in metals. J Nucl Energy Part A: React Sci 12:69

Manninen M (1986) Interatomic interactions in solids: an effective-medium approach. Phys Rev B 34(12):8486

Marian J, Zepeda-Ruiz LA, Couto N, Bringa EM, Gilmer GH, Stangeby PC, Rognlien TD (2007) J Appl Phys 101:044506

Matsunaga K, Fisher C, Matsubara H (2000) Tersoff potential parameters for simulating cubic boron carbonitrides. Jpn J Appl Phys 39:L48–L51

Medvedev N, Volkov A, Shcheblanov N, Rethfeld B (2010) Early stage of the electron kinetics in swift heavy ion tracks in dielectrics. Phys Rev B 82:125425

Medvedev N, Rymzhanov R, Volkov A (2015a) Time resolved electron kinetics in swift heavy ion irradiated solids. J Appl Phys D 48:355303

Medvedev N, Volkov A, Ziaja B (2015b) Electronic and atomic kinetics in solids irradiated with free-electron lasers or swift-heavy ions. Nucl Instrum Methods Phys Res B 365:437–446

Meftah A, Brisard F, Costantini JM, Dooryhee E, Hage-Ali M, Hervieu M, Stoquert JP, Studer F, Toulemonde M (1994) Track formation in SiO_2 quartz and the thermal-spike mechanism. Phys Rev B 49:12457–12463

Meldrum A, Zinkle SJ, Boatner LA, Ewing RC (1998) A transient liquid-like phase in the displacement cascades in zircon, hafnon and thorite. Nature 395:56–58

Miyagawa Y, Nakadate H, Djurabekova F, Nakao S, Miyagawa S (2002) Dynamic-sasamal: simulation software for high dose ion implantation. Surf&Coat Tech 158–159:87

Mookerjee S, Beuve M, Khan SA, Toulemonde M, Roy A (2008) Sensitivity of ion-induced sputtering to the radial distribution of energy transfers: a molecular dynamics study. Phys Rev B 78:045435

Munetoh S, Motooka T, Moriguchi K, Shintani A (2007) Interatomic potential for Si-O systems using Tersoff parameterization. Comput Mater Sci 39(2):334–339

Lima NW, Gutierres LI, Gonzalez RI, Müller S, Thomaz RS, Bringa EM, Papaléo RM (2016) Molecular dynamics simulation of polymerlike thin films irradiated by fast ions: a comparison between FENE and Lennard-Jones potentials. Phys Rev B 94:195417

Ni B, Lee KH, Sinnott SB (2004) A reactive empirical bond order (REBO) potential for hydrocarbon oxygen interactions. J Phys: Condens Matter 16:7261–7275

Nordlund K (1995) Molecular dynamics simulation of ion ranges in the 1–100 keV energy range. Comput Mater Sci 3:448

Nordlund K, Keinonen J, Kuronen A (1994) Effect of the interatomic Si-Si potential on vacancy production during ion implantation of Si. Physica Scripta T54:34

Nordlund K, Keinonen J, Mattila T (1996) Formation of ion irradiation-induced small-scale defects on graphite surfaces. Phys Rev Lett 77(4):699

Nordlund K, Runeberg N, Sundholm D (1997) Repulsive interatomic potentials calculated using Hartree-Fock and density-functional theory methods. Nucl Instrum Methods Phys Res B 132:45–54

Nordlund K, Ghaly M, Averback RS, Caturla M, Diaz de la Rubia T, Tarus J (1998) Defect production in collision cascades in elemental semiconductors and FCC metals. Phys Rev B 57(13):7556–7570

Nordlund K, Wallenius J, Malerba L (2005) Molecular dynamics simulations of threshold energies in Fe. Nucl Instrum Methods Phys Res B 246(2):322–332

Ohta H, Hamaguchi S (2001) Classical interatomic potentials for Si-O-F and Si-O-Cl systems. J Chem Phys 115(14):6679–90

Ojanperä A, Krasheninnikov AV, Puska M (2014) Electronic stopping power from first-principles calculations with account for core electron excitations and projectile ionization. Phys Rev B 89:035120

Oligschleger C, Jones RO, Reimann SM, Schober HR (1996) Model interatomic potential for simulations in selenium. Phys Rev B 53:6165

Pailthorpe B, Mahon P (1990) Molecular dynamics simulation of thin film diamond. Thin Solid Films 193/194:34

Pakarinen OH, Djurabekova F, Nordlund K, Kluth P, Ridgway M (2009) Molecular dynamics simulations of the structure of latent tracks in quartz and amorphous SiO_2. Nucl Instrum Methods Phys Res B 267:1456–1459

Pakarinen OH, Djurabekova F, Nordlund K (2010) Density evolution in formation of swift heavy ion tracks in insulators. Nucl Instrum Methods Phys Res B 268:3163

Peltola J, Nordlund K, Keinonen J (2006) Electronic stopping power calculation method for molecular dynamics simulations using local Firsov and free electron-gas models. Radiat Eff Defects Solids 161(9):511–521

Polvi J, Luukkonen P, Nordlund K, Järvi TT, Kemper TW, Sinnott SB (2012) Primary radiation defect production in polyethylene and cellulose. J Phys Chem B 116(47):13932

Powell D, Migliorato MA, Cullis AG (2007) Optimized Tersoff potential parameters for tetrahedrally bonded III–V semiconductors. Phys Rev B 75(11):115202

Pronnecke S, Caro A, Victoria M, Diaz de la Rubia T, Guinan MW (1991) The effect of electronic energy loss on the dynamics of thermal spikes in Cu . J Mater Res 6(3):483–91

Pruneda JM, Sánchez-Portal D, Arnau A, Juaristi JI, Artacho E (2007) Electronic stopping power in LiF from first principles. Phys Rev Lett 99:235501. https://doi.org/10.1103/PhysRevLett.99.235501

Pugacheva TS, Djurabekova FG, Valiev SK (1998) Effects of cascade mixing, sputtering and diffusion by high dose light ion irradiation of boron nitride. Nucl Instrum Methods Phys Res B 141:99–104

Puska MJ, Nieminen RM, Manninen M (1981) Atoms embedded in an electron gas: immersion energies. Phys Rev B 24(6):3037

Raman S, Jurney ET, Warner JW, Kuronen A, Keinonen J, Nordlund K, Millener DJ (1994) Lifetimes in ^{15}N from gamma-ray lineshapes produced in the ^{2}H(^{14}N,$p\gamma$) and ^{14}N(thermal n, γ) reactions. Phys Rev C 50(2):682

Rethfeld B (2004) Unified model for the free-electron avalanche in laser-irradiated dielectrics. Phys Rev Lett 92(18):187401

Rethfeld B, Rämer A, Brouwer N, Medvedev N, Osmani O (2014) Electron dynamics and energy dissipation in highly excited dielectrics. Nucl Instrum Methods Phys Res B 327:78–88

Ridgway M, Bierschenk T, Giulian R, Afra B, Rodriguez MD, Araujo L, Byrne AP, Kirby N, Pakarinen OH, Djurabekova F, Nordlund K, Schleberger M, Osmani O, Medvedev N, Rethfeld B, Wesch W, Kluth P (2013) Track and voids in amorphous Ge induced by swift heavy-ion irradiation. Phys Rev Lett 110:245502

Rivera A, Olivares J, Prada A, Crespillo ML, Caturla MJ, Bringa EM, Perlado JM, Pena-Rodriquez O (2017) Permanent modifications in silica produced by ion-induced high electronic excitation: experiments and atomistic simulations. Sci Rep 7:10641

Ruault MO, Chaumont J, Penisson JM, Bourret A (1984) High resolution and *in situ* investigation of defects in Bi-irradiated Si. Philos Mag A 50(5):667

Rymzhanov RA, Medvedev NA, Volkov AE (2016) Effects of model approximations of electron, hole, and photon transport in swift heavy ion tracks. Nucl Instrum Methods Phys Res 388:41–52

Salonen E, Nordlund K, Keinonen J, Wu CH (2001) Swift chemical sputtering of amorphous hydrogenated carbon. Phys Rev B 63:195415

Samela J, Kotakoski J, Nordlund K, Keinonen J (2005) A quantitative and comparative study of sputtering yields in Au. Nucl Instrum Methods Phys Res B 239(4):331–346

Samela J, Nordlund K, Popok VN, Campbell EEB (2008) Origin of complex impact craters on native oxide coated silicon surfaces. Phys Rev B 77:075309

Sand AE, Dequeker J, Becquart CS, Domain C, Nordlund K (2016) Non-equilibrium properties of interatomic potentials in cascade simulations in tungsten. J Nucl Mater 470:119–127

Sandoval L, Urbassek HM (2009) Influence of electronic stopping on sputtering induced by cluster impact on metallic targets. Phys Rev B 79(14):144115

Sastry S, Angell CA (2003) Liquid-liquid phase transition in supercooled silicon. Nat Mater 2:739

Schäfer C, Urbassek HM, Zhigilei LV (2002) Phys Rev B 66:115404

Schiwietz G, Grande PL (2011) Introducing electron capture into the unitary-convolution-approximation energy-loss theory at low velocities. Phys Rev A 84:052703. http://link.aps.org/doi/10.1103/PhysRevA.84.052703

Schiwietz G, Czerski K, Roth M, Staufenbiel F, Grande PL (2004) Femtosecond dynamics – snapshots of the early ion-track evolution. Nucl Instrum Methods Phys Res Sect B 226(4):683–704

Sillanpää J, Peltola J, Nordlund K, Keinonen J, Puska MJ (2001) Electronic stopping calculated using explicit phase shift factors. Phys Rev B 63:134113

Skupinski M, Toulemonde M, Lindeberg M, Hjort K (2005) Ion tracks developed in polyimide resist on Si wafers as template for nanowires. Nucl Instrum Methods Phys Res B 240(3):681–689. https://doi.org/10.1016/j.nimb.2005.04.128

Stillinger FH, Weber TA (1985) Computer simulation of local order in condensed phases of silicon. Phys Rev B 31:5262

Stoller RE, Tamm A, Beland LK, Samolyuk GD, Stocks GM, Caro A, Slipchenko LV, Osetsky YN, Aabloo A, Klintenberg M, Wang Y (2016) Impact of short-range forces on defect production from high energy collisions. J Chem Theory Comput 12(6):2871–2879

Stuart SJ, Tutein AB, Harrison JA (2000) A reactive potential for hydrocarbons with intermolecular interactions. J Chem Phys 112:6472

Stuchbery AE, Bezakova E (1999) Thermal-spike lifetime from picosecond-duration preequilibrium effects in hyperfine magnetic fields following ion implantation. Phys Rev Lett 82(18):3637

Szenes G (2005) Ion-induced amorphization in ceramic materials. J Nucl Mater 336(1):8–89

Tersoff J (1988) New empirical approach for the structure and energy of covalent systems. Phys Rev B 37:6991

Toimil-Molares ME (2012) Characterization and properties of micro- and nanowires of controlled size, composition, and geometry fabricated by electrodeposition and ion-track technology. Beilstein J Nanotechnol 3(1):860

Tombrello TA (1984) Track damage and erosion of insulators by ion-induced electronic processes. Nucl Instrum Methods Phys Res B https://doi.org/10.1016/0168-583X(84)90265-9

Timkó H, Djurabekova F, Costelle L, Nordlund K, Matyash K, Schneider R, Toerklep A, Arnau-Izquierdo G, Descoeudres A, Calatroni S, Taborelli M, Wuensch W (2010) Mechanism of surface modification from the arc plasma-surface interaction in Cu. Phys Rev B 81:184109

Toulemonde M, Dufour C, Meftah A, Paumier E (2000) Transient thermal processes in heavy ion irradiation of crystalline inorganic insulators. Nucl Instrum Methods Phys Res B 166–167: 903–912

Toulemonde M, Assmann W, Dufour C, Meftah A, Studer F, Trautmann C (2006) Experimental phenomena and thermal spike model description of ion tracks in amorphisable inorganic insulators. Mat Fys Medd Kong Dan Vid Selsk 52:263

Träskelin P, Salonen E, Nordlund K, Keinonen J, Wu CH (2004) Molecular dynamics simulations of CH_3 sticking on carbon surfaces, angular and energy dependence. J Nucl Mater 334(1):65

Trautmann C, Klaumünzer S, Trinkaus H (2000) Effect of stress on track formation in amorphous iron boron alloy: ion tracks as elastic inclusion. Phys Rev Lett 85(17):3648

Urbassek HM, Kafemann H, Johnson R (1994) Atom ejection from a fast-ion track: a molecular-dynamics study. Phys Rev B 49(2):786

Vazquez H, Ahlgren EH, Ochedowski O, Leino AA, Mirzayev R, Kozubek R, Lebius H, Karlusic M, Jaksic M, Krasheninnikov AV, Kotakoski J, Schleberger M, Nordlund K, Djurabekova F (2017) Creating nanoporous graphene with swift heavy ions. Carbon 114:511

Waligórski RM, Hamm RN, Katz R (1986) The radial distribution of dose around the path of a heavy ion in liquid water. Nucl Tracks Meas 11:309

Wang ZG, Dufour C, Paumier E, Toulemonde M (1994) The s_e sensitivity of metals under swift-heavy-ion irradiation: a transient thermal process. J Phys: Condens Matter 53(34):6733

Wang ZQ, Stroud D (1988) Monte Carlo studies of liquid semiconductor surfaces: Si and Ge. Phys Rev B 38(2):1384

Watanabe T, Yamasaki D, Tatsumura K, Ohdomari I (2004) Appl Surf Sci 234:207

Weber WJ, Zarkadoula E, Pakarinen OH, Sachan R, Chisholm MF, Liu P, Xue H, Jin K, Zhang Y (2014) Synergy of elastic and inelastic energy loss on ion track formation in srtio3. Sci Rep 5:7726

Yi X, Sand AE, Mason DR, Kirk MA, Roberts SG, Nordlund K, Dudarev SL (2015) Direct observation of size scaling and elastic interaction between nano-scale defects in collision cascades. EPL 110:36001

Yu JG, Sinnott SB, Phillpot SR (2007) Charge optimized many-body potential for the Si/SiO_2 system. Phys Rev B 75(8):085311

Zeb MA, Kohanoff J, Sanchez-Portal D, Arnau A, Juaristi JI, Artacho E (2012) Electronic stopping power in gold: the role of d electrons and the H/He anomaly. Phys Rev Lett 108:225504

Zhigilei LV, Garrison BJ (1999) Pressure waves in microscopic simulations of laser ablation. Mat Res Soc Symp Proc 538:491

Zhigilei L, Kodali PBS, Garrison BJ (1998) A microscopic view of laser ablation. J Chem Phys B 102:2845

Ziegler JF (1995) TRIM-95 computer code, private communication

Ziegler JF (2013) SRIM-2013 software package, available online at http://www.srim.org

Ziegler JF, Biersack JP, Littmark U (1985) The stopping and range of ions in matter. Pergamon, New York

Zinoviev AN (2015) Electron screening of the Coulomb potential at small internuclear distances. Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 354:308–312

Zinoviev AN, Nordlund K (2017) Comparison of repulsive interatomic potentials calculated with an all-electron DFT approach with experimental data. Nucl Instrum Methods Phys Res B 406:511–517

Kinetic Monte Carlo Algorithms for Nuclear Materials Applications

89

J. P. Balbuena, María José Caturla, and Enrique Martínez

Contents

Abstract

Microstructure evolution of irradiated materials is a complex phenomenon that involves time and length scales that can expand several orders of magnitude. Defects produced in the irradiation can interact with the existing microstructure, sometimes inducing changes in the mechanical, electrical, or even magnetic properties. The selection of the most adequate material for nuclear applications requires an understanding at a fundamental level of the evolution of these defects during the lifetime of the reactors. Therefore, very efficient simulation tools, with physical and accurate parameters must be used. In this review,

J. P. Balbuena · M. J. Caturla (✉)
Departamento de Física Aplicada, Facultad de Ciencias, Fase II, Universidad de Alicante, Alicante, Spain
e-mail: mj.caturla@ua.es; maria.caturla@gmail.com

E. Martínez
Material Science and Technology Division, MST-8, Los Alamos National Laboratory, Los Alamos, NM, USA

one of the computational methods that is commonly employed to study defect evolution, kinetic Monte Carlo, is described. The differences and similarities between three algorithms are explained: atomistic (or lattice), object, and event kinetic Monte Carlo. In order to reveal the applicability of these methods in the nuclear field, examples are given for the case of one of the candidates for first wall materials for both fusion and IV generation fission reactors: FeCr alloys. Finally, the limitations of the models to describe these systems and the current efforts to improve the predictive capabilities of this approach, as well as other developments in the field are discussed.

1   Introduction

The search for materials for nuclear applications, in both fusion and next generation fission reactors, requires reliable models capable of describing the production and evolution of defects produced during the irradiation and the interaction of these defects with the microstructure (Zinkle et al. 2009; Dudarev et al. 2009; Knaster et al. 2016). Due to the difficulties in performing neutron irradiation experiments, ion irradiation has become during the last years a commonly used method to obtain information about how defects are produced and how they interact. However, an extrapolation from ion irradiation conditions to neutron damage is needed, and in this task, simulations are a key component. With the use of accurate computer models, the large range of parameters in these experiments can be analyzed to understand their influence in microstructural changes. These parameters are temperature, dose rate, dose, irradiation energy, surface effects, dislocation density, or impurity content, among others. However, no single model is capable of describing the effects of irradiation from the time scale of defect production (a few picoseconds) to the lifetime of a reactor (years) or from the size of the collision cascade (a few nanometers) to sizes that can be compared to experimental characterization results.

Recently, multiscale modeling (Phillips 2001) has been quite successful in building validated models starting from first principles calculations (Heinisch 1990; Caturla et al. 2000; Domain et al. 2004; Fu et al. 2005; Malerba et al. 2008; Martínez et al. 2012, among others). Binary collision approximation models together with molecular dynamics can provide information about the initial defect distribution produced by the irradiation up to several picoseconds. The defect distribution, their character, type, and structure can be obtained from these models although, to this day, there is no experimental validation of these results at this short time scales and localized dimensions. However, this information can be used as input for other models such as rate theory (RT) or kinetic Monte Carlo (KMC). In fact, already in the early 1990s, the importance of the defect distribution at the picosecond time scale was revealed to explain phenomena such as void swelling using results obtained with molecular dynamics and rate theory models (Woo et al. 1991). Later, the relevance of the initial defect distribution on microstructure evolution has been

shown in different metals and irradiation conditions using kinetic Monte Carlo models (Heinisch 1990; Caturla et al. 2000; Domain et al. 2004).

One important point in the development of multiscale methods is the validation of the models with experiments. And in this validation, KMC plays a crucial role. As it was mentioned above, results obtained from molecular dynamics about damage produced from a collision cascade cannot be, to this day, directly compared to experiments. Similarly, results of defect migration that can be obtained with classical molecular dynamics or density functional theory (DFT) can only be compared with experiments with the aid of KMC or RT models. Irradiated materials characterization can be done using different techniques. Some of the most widely used are transmission electron microscopy (TEM), atom probe tomography (APT), small angle neutron scattering (SANS), and positron annihilation spectroscopy (PAS). Transmission electron microscopy, for example, can provide information about defect concentrations, sizes, character (vacancy or self-interstitial) (Jenkins et al. 1976), structure (e.g., orientation of dislocation loops), and, under some conditions, chemical composition. The resolution of this method has improved significantly during the last few years (Zinkle et al. 2009), but it is still limited to diameters of the order of 1 nm in most cases. Also, *in situ* TEM measurements during irradiation reveal the dynamic nature of defect production and evolution and can reveal processes that would be lost in a "postmortem" analysis of the irradiated sample (Yao et al. 2008). SANS or PAS can provide information about smaller defects than TEM, but they also have their limitations. PAS, for example, can only detect open defects such as vacancies or voids of all sizes (Kuriplach 2002); however, the interpretation of the results is not straightforward. APT (Miller 2000) is able to provide information about local chemical composition and therefore help in understanding phenomena such as irradiation-induced segregation of solutes and precipitation (Hornbogen and Glenn 1960; Miller et al. 2003).

Besides characterization techniques, there are also other experimental methods to obtain information about defects in irradiated metals. Resistivity recovery experiments, which consist of irradiating the sample at low temperatures and then annealing at increasing temperatures to observe the different stages of defect recombination, have proven to be particularly useful in identifying migration energies of point defects (Takaki et al. 1983; Matsui et al. 1988; Ullmaier 1991; Fu et al. 2005). In magnetic materials such as Fe and FeCr alloys, other methods can also be used to understand changes in magnetic properties after irradiation, such as polarized neutron reflectivity (Papamilhail et al. 2016a, 2016b). As mentioned above, comparing fundamental calculations of defect properties and defect distributions to these experimental results can only be done with the use of a kinetic model.

In this review, the kinetic Monte Carlo algorithms that are more widely used in nuclear applications are described, in particular, lattice (or atomistic) kinetic Monte Carlo (AKMC), object kinetic Monte Carlo (OKMC), and event kinetic Monte Carlo (EKMC). After the description of the methods in general terms and the parameters required, examples are provided for AKMC and OKMC. In nuclear applications, ferritic/martensitic steels are considered as good candidates for structural materials

both for fusion and generation IV fission reactors (Andreani et al. 2006; van der Schaaf et al. 2006) with Cr concentrations between 9 and 12% (Murty and Charit 2008; Mansur et al. 2004). These materials have been selected due to their resistance to void swelling as well as creep; however, they present problems of embrittlement as well as some concerns about phase stability under irradiation (Yamamoto et al. 2006). Therefore, the examples that will be given here are of FeCr alloys under irradiation for low and high Cr concentrations. One of the main limitations of the KMC and rate theory methods is that the relevant events to be included are selected a priori, and there is no unambiguous method for this selection; it is to the discretion of the modeler, who normally chooses these processes through an iterative comparison with selected experiments. In recent years, new methods have emerged to avoid this limitation. At the end of this chapter, the limitations of these techniques, the needs for nuclear materials applications, and the developments toward reaching these goals are discussed.

2 Methodology

Given the complexity of systems with many constituents and possible evolution pathways, predicting the dynamic evolution of such a system, which might be out of equilibrium, will require numerical methods. Hereby a stochastic approach to solve this problem is described, which consists in solving the single master equation of all possible reactions without dealing with it directly, since the latter equation is very difficult to solve even numerically. Despite the fact that this method was originally aimed to solve coupled chemical reaction systems, it is being also widely used to model radiation damage microstructure time evolution in nuclear materials.

2.1 The Kinetic Monte Carlo Algorithm

In a similar way to the Bortz-Kalos-Lebowitz (BKL) algorithm (Bortz et al. 1975), this time-dependent method for solving the master equation relies in random selection of possible physical events attending to their rate of occurrence (Gillespie 1976). The method assumes that each of the events experienced by the system are *memoryless* or Markovian, which means that the transition rate for event j, r_j, apart from being independent of time, only depends on the initial and final states. A common assumption in these methods from transition state theory (TST) is expressing the transition rates r_j in the Arrhenius form for simplicity. This way, the list of required parameters for such a rate is limited to the prefactor r_j^0 and the activation energy of the process, E_j:

$$r_j = r_j^0 \, exp \, \left(-E_j / k_B T \right) \tag{1}$$

A cumulative rate function R_N is defined as the sum of all the possible transitions that can occur during the many-body system dynamics. According to its definition,
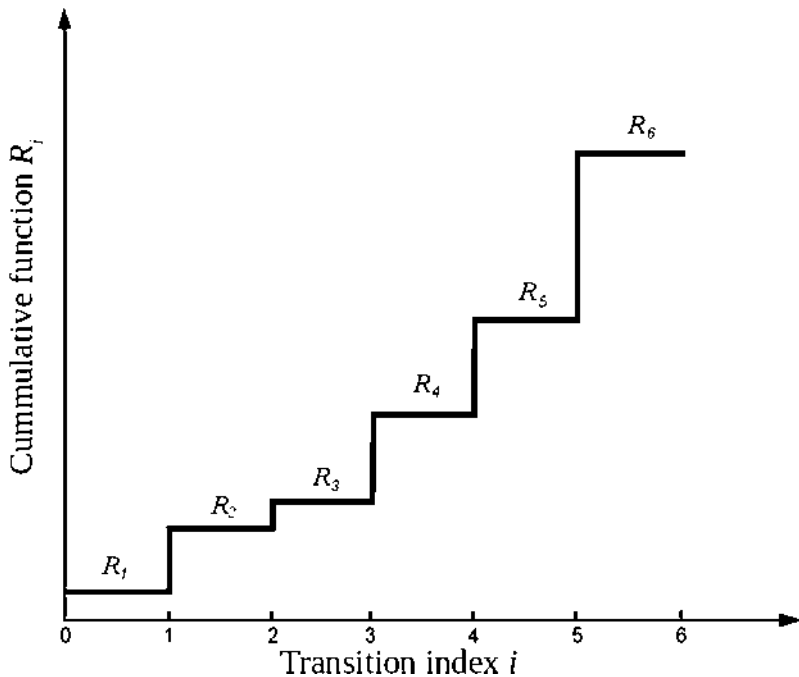
Fig. 1 Representation of the cumulative function R_i as a step function

R_N is a monotonously growing step function, as illustrated in Fig. 1, whose value R_i is constant in the interval {i-1, i}:

$$R_N = \Sigma_{i=1}{}^N R_i \qquad (2)$$

where N is the total number of possible events in the system, thus identifying both the transition and the specific element in the system which undergoes such transition.

The selection of which event takes place during the dynamic evolution of the system using this algorithm is described as follows:

1. Calculation of the cumulative function R_N, as the sum of all the possible transition rates R_i for every element in the multibody system:

$$R_N = \Sigma_{i=1}{}^N \Sigma_{j=1}{}^i r_j \qquad (3)$$

2. Sample of a first random number $rand_1$ using a uniform random number generator in the unit interval.
3. Perform the transition i that corresponds to the cumulative rate R_i, considering $R_{i-1} < rand_1 R_N \leq R_i$ and as shown in Fig. 2.
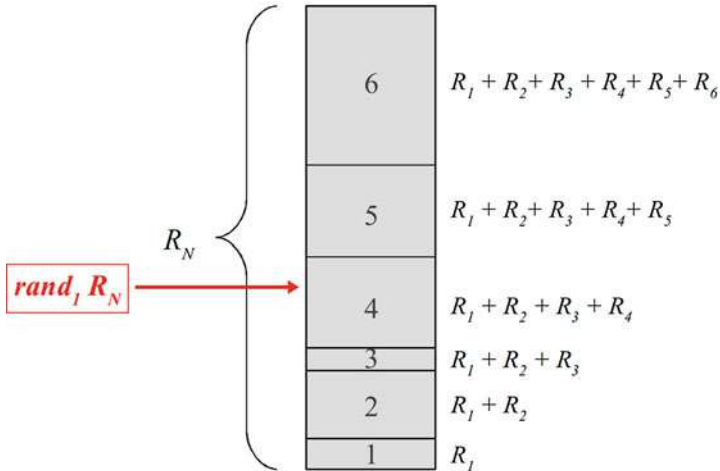
Fig. 2 Selection of the transition i after a time step dt in the kinetic Monte Carlo algorithm

4. Sample of a second random number $rand_2$ as in step (2).
5. Increase of the "real" elapsed time by a time step dt calculated as:

$$dt = \ln(1/rand_2)/R_N \qquad (4)$$

6. Recalculation of the rates R_i affected by the executed transition and updating the value of the cumulative function R_N.
7. Return to step (2) until a simulation ending condition is reached (elapsed time, number of executed events, etc.).

This algorithm, as depicted above, is only taking care of the time evolution of the multibody system. Spatial dependencies can be integrated to the particular event (transition undergone by a particular element) and can be updated as desired. Several alternative applications of this method using different space managements are illustrated in the following sections.

2.2 AKMC: Atomistic Kinetic Monte Carlo

When the KMC method is applied to elements belonging to a rigid lattice structure, it is commonly referred as lattice KMC (LKMC) or atomistic KMC (AKMC). Elements in this system which experience the state transitions selected from the KMC algorithm are bonded to specific fixed positions given by the crystallographic structure of the material. This technique is widely used to simulate phase separation (or intermixing) of alloys assisted by the vacancy diffusion mechanism (Young and Elock 1966).

In general, the AKMC method can be used to simulate the migration of mobile point defects (I and V). This fact is of particular importance since materials studied applying this method are aimed to be used in nuclear power plants. The main consequence of the radiation damage produced in these materials is the formation of Frenkel Pairs; thus, concentrations of I and V will be present at all times in these materials. The migration of point defects affects the local composition after every migration jump, as concentration of I and V, i.e., C_I and C_V, respectively, is modified locally but also because the migration of V involves a bulk atom to fill the previous location of the vacancy. In case of alloys, the latter process can change the atomic distribution of solute and solvent, as it was mentioned earlier.

2.3   OKMC: Object Kinetic Monte Carlo

An alternative application of the KMC algorithm to study the microstructure evolution of defects is the object KMC (OKMC) or off-lattice KMC method. OKMC codes do not explicitly make use of a predefined lattice structure, but it can be taken into account in the migration steps of point defects, as the code is tracking the coordinates of every single defect in the simulations. The simulation domain in these algorithms is divided into a homogeneous mesh, as shown in Fig. 3. This feature allows the calculation of either point defects or solute concentrations to obtain information such as tracer diffusivities or energy of mixing in alloy materials.

Among other features, OKMC codes are not only able to handle all point defects and impurities but also point defect clusters, mixed clusters of different shapes and Burger vector directions, interaction with interfaces, and more. This is of great interest as these methods can be used to reproduce, and thus explain, for example, the formation of dislocation loops found in transmission electron microscopy characterization of irradiated materials.

Fig. 3 Schematic representation of the resulting mesh containing point defects, impurities, clusters, and interfaces. In this case each point represents a point defect; colors indicate different defect types
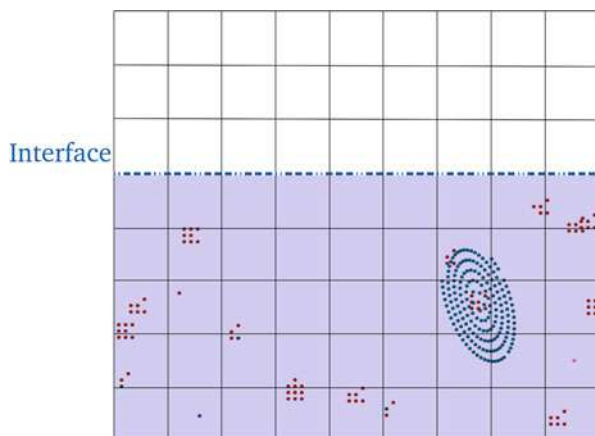
2.4   EKMC: Event Kinetic Monte Carlo

The two methods described above (OKMC and AKMC) use the BKL algorithm (Bortz et al. 1975) to evolve the time. The so-called event kinetic Monte Carlo (EKMC) and the first-passage kinetic Monte Carlo (FPKMC) methods use a different algorithm for advancing the simulation time. As in OKMC a list of events is required, but here the time delay for all the possible events is calculated first, and the event with the shortest time delay is the one picked. This event is then performed and all the new time delays calculated again for the next simulation step. The time advances by the delay time of the event picked. In the EKMC method, some approximations are made in order to calculate the delay time of events (Lanore 1974; Dalla Torre et al. 2005; Fu et al. 2005; Jourdan and Crocombette 2012; Becquart et al. 2010). A more general method was developed by Opplestrup et al. (2006, 2009), the FPKMC. Here a "protective domain" is defined to calculate the event rate, and if that event is selected, it is moved to the edge of its domain. This approximation makes the algorithm very efficient for low defect concentrations compared to OKMC; however, the efficiency decreases at high concentrations since the "protective domains" have to be calculated at each time scale (Donev et al. 2010).

3   AKMC Applied to Modeling Fe and FeCr Alloys

High chromium ferritic and martensitic steels are considered as essential materials for several concepts for next-generation fission reactors (Klueh and Nelson 2007) as well as for future fusion reactors (Zinkle and Busby 2009). The addition of Cr has a beneficial effect on mechanical, corrosion, and radiation resistance properties of the matrix: the study of the stability of Fe-Cr solid solutions versus the α-α' decomposition, which occurs below the domain of stability of the σ phase, is therefore of great technological importance. The thermodynamic and kinetic modeling of this demixing process is a challenging problem, even without irradiation, due to the specific properties of Fe-Cr alloys. This system exhibits a unique inversion between ordering and segregation tendencies at the Cr contents of technological interest (9–12%) (Mirebeau et al. 1984). This behavior is related to the magnetic properties of the Fe-Cr alloys, as it has been shown by recent ab initio studies (Klaver et al. 2006; Olsson et al. 2006). Fe-Cr alloys also present significant vibrational entropies (Swan-Wood et al. 2005). Magnetic and vibrational contributions to the entropy affect both the thermodynamic and diffusion properties of the system. The details of the thermodynamic and kinetic model used in the KMC calculations are described in Levesque et al. (2011), Martinez et al. (2011), and Senninger et al. (2014, 2016).

  LKMC has been used to study the precipitation kinetics of a Fe-20%Cr sample at 773 K, which are the same conditions used by Novy et al. (2009) in their 3D atom probe experiments. Transport has been considered through vacancy diffusion. The rate for vacancy hopping has been calculated using a harmonic transition state theory approximation. The migration barrier is obtained via a broken bond model
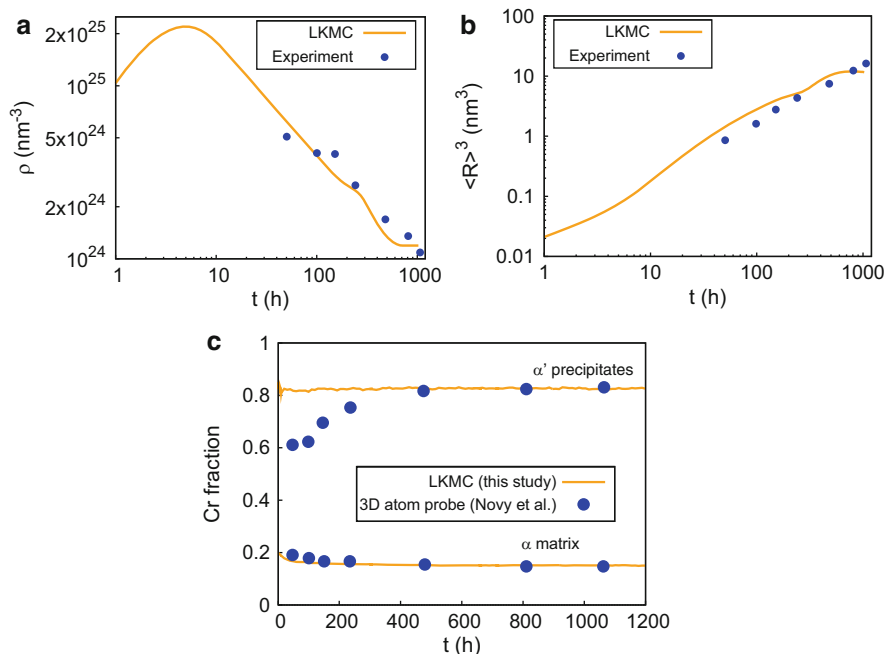
Fig. 4 Comparison between the LKMC simulations and 3DAP experiments (Novy et al. 2009) for a Fe-20%Cr sample (**a**) evolution of the density of α' precipitates during annealing, (**b**) evolution of the average radius of α' precipitates, and (**c**) evolution of the Cr fraction. Reprinted with permission from (Martinez et al. 2012) Copyright (2012) by the American Physical Society

with pair interactions fitted to ab initio data. Figure 4a and b show the evolution of the density and average radius, respectively, of the α' precipitates as given by both experiments and LKMC. We observe that simulations predict extremely well the experimental values for both the density and average radius although the average precipitate size seems slightly larger in the simulations.

Comparison of the evolution of the composition in the different phases, α and α', has also been performed. Results are shown in Fig. 4c, where it might be observed that the equilibrium compositions, after 500 h, predicted by the simulations are also in good agreement with 3DAP measurements. For shorter aging times however, the evolution of the composition appears much more rapid in the simulation than in the experiments of Novy et al., especially for the α' phase. It should be noticed that in the LKMC simulations, this evolution is found to depend on the criterion chosen to define the precipitates, but the previous conclusion remains qualitatively valid: the Cr concentration in the α' precipitates reaches its equilibrium value much faster in the simulations than in the experiments. It is well known that aberrations of ion trajectories and local magnification effects limit the spatial resolution of 3DAP. The composition of very small precipitates (below 2 nm) is thus subjected to significant statistical error (Vurpillot et al. 2000). The choice of parameters used for the identification of solute clusters may also induce additional uncertainties

(Hyde et al. 2011). It is worth noting that the size and density of precipitates, as well as the solute concentration of the matrix, are much less sensitive to such uncertainties. A typical case is copper precipitation in iron, where LKMC simulations (Soisson et al. 2010) usually predict almost pure Cu for precipitate radii as small as 1 nm, while 3DAP measurements overestimate the iron content (Morley et al. 2009).

Irradiation effects can also be taken into account in these models (Senninger et al. 2016; Martínez et al. 2018). For that end, self-interstitials have been modeled in the same way as vacancies, characterizing the formation and migration energies through DFT data. LKMC can be used to model radiation-induced segregation (RIS) near point defect sinks, going beyond the Wiedersich model. The LKMC simulations take into account the dependence of the transport coefficients on the local concentration and can treat precipitation. Figure 5 shows the steady-state concentration of defects and the Cr profile for a 10%Cr alloy at 650 K and 950 K with dose rate 10–6 dpa/s and 10–3 dpa/s. As predicted by the Wiedersich model, one observes Cr enrichment at 650 K and depletion at 950 K. The concentration of Cr at the ideal sink evolves very rapidly and reaches steady-state at low doses (a few 0.01 dpa). On the other hand, the total Cr segregated, defined as:

$$S^{Cr} = \int \left(X_{Cr}(z) - X_{Cr}(0) \right) dz \tag{5}$$

evolves much more gradually.
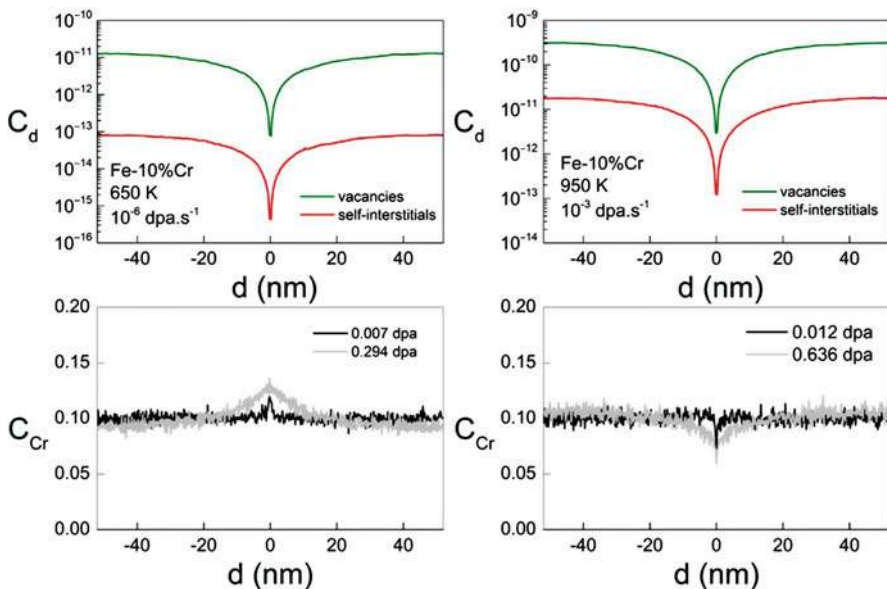
Fig. 5 Steady-state point defect and Cr concentration profiles in the vicinity of an ideal sink for an alloy at 10% Cr. Left, irradiated at 650 K and 10–6 dpa/s and, right, irradiated at 950 K and 10–3 dpa/s. Reprinted from (Senninger et al. 2016) Copyright (2016), with permission from Elsevier

4 OKMC Applied to Modeling Fe and FeCr Alloys

An advantage of working without a rigid lattice is the savings in memory use that allows for simulations with larger sizes. OKMC codes are specific for all of those objects of interest that can be found inside irradiated steels, such as point defects and impurities, and study their time evolution. Diffusion rate of mobile particles is a thermally activated process whose frequency can be described as:

$$\nu = \nu_0 \exp\left(-E_m/k_B T\right) \tag{6}$$

where ν_0 and E_m are the migration prefactor and activation energy, respectively. These parameters, among many others, are introduced in the simulator beforehand so it is important to find accurate values either from experimental data available in the literature or by numerical calculations methods such as density functional theory (DFT) and molecular dynamics (MD). In a three-dimensional movement, the diffusivity of impurities (C, etc.), point defects (I or V), and defect-driven lattice atoms (Fe or Cr) can be calculated as:

$$D = \nu \, \lambda^2/2d \tag{7}$$

where λ is the jump length and d is the dimension of the diffusion, often considered as three-dimensional although in metals, some types of defects move in one dimension. The frequency of migration, ν, introduced in this equation is straightforward for impurities and point defects from Eq. (15.1). Diffusivities of defect-driven lattice atoms in FeCr alloys can be also obtained through the simulation execution.

Migration of point defects and impurities eventually leads to their interaction with other objects of the same type but also with boundaries (interfaces) or clusters. Point defects can undergo recombination if they interact with the opposite defect or create more complex structures. The distance for which an object interacts with the surrounding defects/boundaries is given by its capture radius. This parameter is set as an input data for each mobile particle in the system; a default value for this parameter is the lattice parameter of the b.c.c. iron: 0.287 nm. In case of more complex structures such as voids, dislocation loops, etc., this parameter can be determined in different ways depending of the geometric description of the complex. Some OKMC codes use a capture sphere using a capture radius centered in the middle of the object (Caturla et al. 2000; Domain et al. 2004; Ortiz and Caturla 2007), while others use the envelope of the sum of capture radii of its constituents (Martin-Bragado et al. 2013). In the first description, clusters of defects are represented as a single point, located at the center of mass of the cluster, with the number of defects in that cluster and a capture radius that can depend on the number of defects and/or the cluster type. In the latter, the location of each point defect is actually kept in the simulation (Martin-Bragado et al. 2013).

There are also different ways when dealing with Cr in OKMC codes when modeling FeCr alloys. One option consists of considering Cr as an impurity that

can occupy substitutional sites in the lattice (Gámez et al. 2011), so it typically diffuses by paring itself with an I or V, until the latter point defect is emitted (Martin-Bragado et al. 2018). Another way consists of counting the number of both Fe and Cr atoms in each cell (hence allowing the calculation of their concentration in such cells when required), instead of recording individual positions for each Fe and Cr atoms (Dopico et al. 2015). The latter way treats global inhomogeneous alloys as discrete cells of homogeneous alloys, where the migration parameters of defects are modified depending on the concentration of solute in the cell they are contained. The mechanical properties of Fe-based and FeCr-based steels for nuclear applications are modified by the complex microstructures that are formed within them due to radiation. OKMC codes are able to model the formation and time evolution of these defects in order to reproduce actual transmission electron microscope observations of such irradiated steels and predict their behavior.

These complex structures (or clusters) of point defects and impurities can trap and/or emit its constituents attending to the energy barrier of the process. As it was previously mentioned, the capture radii of complex structures spatially determine the region where objects are allow to interact. In case of trapping a defect, the formation energies would be needed to compute the binding energy of such a defect to the cluster, while for detrapping (or emission) of defect constituents, the migration energy will be also required. Clusters can also experience other actions like transformation into another cluster, internal Frenkel pair recombinations of its constituents, and one- or three-dimensional migration in case it is a mobile defect. Some examples of these structures are three-dimensional vacancy clusters (voids), small clusters of self-interstitial atoms (SIA) with 3D migration, and dislocation loops of SIA with Burger's vectors along $\frac{1}{2}$ a_0 <111> and a_0 <100> directions. The latter defect has recently attracted scientific interest Marian et al. (2002), Marinica et al. (2012), Xu et al. (2013) and Granberg et al. (2017) to explain its formation since it is an immobile defect, commonly observed in TEM micrographs, that can act as defect sink in these steels, thus varying locally the mechanical behavior.

As mentioned in the introduction, information about migration energies of defects can be obtained from resistivity recovery experiments where first the sample is exposed to irradiation and then it is annealed with increasing temperature. In this process, different activation mechanisms are revealed as the temperature increases, which show as peaks when representing the derivative of the resistivity curve with temperature. Each one of these peaks is due to one (or several) recombination processes that result in the decrease of the resistivity of the sample due to defect annihilation (e.g., recombination between a single self-interstitial and a single vacancy). Therefore, each one of these peaks is normally related to the activation of an event such as the migration of a defect. The interpretation of these curves, and the association of each peak to a particular defect, is not straightforward. The use of combined DFT calculations and KMC simulations has proven to be very useful in the interpretation of these experiments (Fu et al. 2005; Gámez et al. 2011). Here, an example for pure Fe and FeCr in the dilute limit irradiated with electrons is presented. For more details of this calculation, see Gámez et al. (2011).

Table 1 Migration and dissociation energies for defects in pure Fe. Data from Fu et al. (2005)

Species	Migration energy (eV)	Dissociation energy (eV)
I	0.34	
I2	0.43	1.14
I3	0.43	1.26
I4	0.40	1.98
In	Immobile	
V	0.67	
V2	0.62	0.97
V3	0.35	1.04
V4	0.48	1.29
Vn	Immobile	

Table 2 Migration and dissociation energies for Cr in pure Fe

Species	Migration energy (eV)	Dissociation energy (eV)		Reference
ICr	0.23	$Cr + I$	0.42	(Olsson et al. 2007, Olsson 2009, Ortiz et al. 2011)
ICr2	0.30	$Cr + ICr$	0.31	(Olsson et al. 2007, Ortiz et al. 2011)
		$Cr_2 + I$	0.73	
I2Cr	Immobile	$ICr + I$	0.88	(Olsson et al. 2007, Ortiz et al. 2011)
		$Cr + I_2$	0.45	
I2Cr2	Immobile	$Cr + I_2Cr$	0.36	(Olsson et al. 2007, Ortiz et al. 2011)
		$Cr_2 + I_2$	0.75	
		$ICr + ICr$	0.86	
		$ICr_2 + I$	0.97	

The resistivity recovery curve can be modeled using information about migration energies of defects from DFT or from empirical potentials (for those clusters too large to study with DFT). In this case the migration energies of the different types of defects, vacancies, and self-interstitials are those given in Table 1 for the case of pure Fe, while Table 2 gives the values for defects containing Cr. Besides migration energies, it is also necessary to have information about the dissociation energy needed to detach a defect from a cluster. This requires knowledge of the binding energy as a function of number of defects in the cluster for all sizes and for all defect types. From the formation energy of clusters as a function of size, the binding energy of a single defect to a cluster can be calculated as:

$$E_b(n) = E_f(n) + E_f(1) - E_f(n+1) \tag{8}$$

Typically, values for small clusters can be obtained from DFT calculations (Fu et al. 2005). For larger cluster sizes, calculations with empirical potentials

that have been fitted to DFT are normally used. This information is then used to obtain a curve for all cluster sizes by fitting the data from the atomistic simulations. Tables 1 and 2 show the values for the example presented here. The prefactor for defect migration and dissociation is also another parameter in these simulations, and the most common procedure is to consider the vibration frequency of the lattice. The dissociation energy is calculated as the sum of the binding energy of the defect to the cluster plus the migration energy of the defect. For large clusters (in the calculations presented here, for clusters with more than four defects), the binding energy of a cluster with n defects is calculated as:

$$E_b(n) = E_f + \frac{\left(E_b(2) - E_f\right)\left(n^{2/3} - (n-1)^{2/3}\right)}{2^{2/3} - 1} \qquad (9)$$

with $E_f{}^V = 2.07$ eV and $E_f{}^I = 3.77$ eV, after (Fu et al. 2005). Therefore, the dissociation energy of a point defect from a cluster with n defects is $E_d(n) = E_b(n) + E_m$, where E_m is the migration energy of the point defect.

As mentioned above, there are also interaction rules that have to be taken into consideration in an off-lattice kinetic Monte Carlo calculation. On one hand, defect reactions must be defined for all types of defects and all defect sizes. For example, the recombination between self-interstitials (I) and vacancies (V), not just for single defects but for all cluster sizes, must be specified. That is, $V_n + I_m \rightarrow V_{n-m}$ (if $n > m$) or $\rightarrow I_{m-n}$ (if $m > n$). And the same for the interaction between clusters of the same type: $V_n + V_m \rightarrow V_{n+m}$, $I_n + I_m \rightarrow I_{n+m}$. In the example presented here, the calculations were done with the code Bigmac, developed at LLNL (Johnson et al. 1998) where clusters are represented by the number of defects and a capture radius. The capture radius for a cluster is often calculated as:

$$r_n = Z.\left(\frac{3n\Omega}{4\pi}\right)^{1/3} + r_0 \qquad (10)$$

where $r_0 = 3.3$ Å and obtained such that the distance between V and I is 3.3 a_0 (Fu et al. 2005), with $a_0 = 0.287$ nm, n is the number of defects, and Ω is the atomic volume of Z is the bias factor, taken as 1.15 for interstitials and 1 for vacancies.

The description above is for the simplest calculation of a metal under irradiation, only vacancies and self-interstitials are considered, and, moreover, only one type of vacancy clusters or self-interstitial clusters for each size is taken into account. However, in some metals different types of clusters of the same type can be observed after irradiation. For example, in Fe 1/2 <111> loops and <100> loops are often observed after ion irradiation (Yao et al. 2008), and their concentration, size, and ratios depend strongly on parameters such as irradiation energy, impurity concentration, proximity to surfaces, or irradiation type (electron, ion, or neutron). Introducing these details in the calculation imply extending the values of Table 1 for all the different clusters as well as all the different interactions. The same must be done when the effects of impurities such as carbon must be included.

When considering an alloy, different approaches can be used. If it is a diluted alloy, considering diluted as a concentration lower that 0.1%, the alloy element can be introduced in the same way as an impurity, that is, all the positions of the alloying atoms can be explicitly included in the calculation. In this case, the relevant migration and dissociation energies and reactions that involve the alloying atom and all the defects must be specified. For the case of FeCr in the dilute limit, the parameters in Table 2 have been considered in the calculation. These parameters, like in the case of pure Fe (Table 1), come mostly from DFT calculations. For the case of high concentration, this approach in not efficient if large system sizes have to be considered and other methods can be used such as those based on using concentration cells within the simulation box, as explained above (Dopico et al. 2015).

Possibly the simplest calculation of an irradiated metal using OKMC that can be directly compared to experiments is that of the resistivity recovery curve after electron irradiation. In this case defects are produced as Frenkel pairs, distributed randomly in the irradiated area and with a dose rate given by the electron irradiation rate. Results are presented here that follow the conditions in the experiments by Abe and Kuramoto (Abe and Kuramoto 1999) for three different FeCr alloys with concentrations between 0.01 and 0.1%. The initial condition in these simulations are pairs of vacancies and self-interstitials distributed randomly in the simulation box, which was a cube with a volume of $(287 \text{ nm})^3$. Cr atoms are also distributed randomly in the simulation box with the appropriate concentration, depending on the experimental conditions. Table 3 includes the number of Frenkel pairs considered for each calculation, obtained from the experimental initial resistivity of each sample and considering a resistivity per Frenkel pair of 30 $\mu\Omega$.cm/at% (12). An isochronal annealing is then performed, starting at 77 K and increasing the temperature 3 K every 3 minutes, following the experimental conditions in Abe and Kuramoto (1999).

Figure 6 shows the derivative with temperature of the total number of defects normalized to the initial number of Frenkel pairs obtained for the temperature range studied (80 K up to 150 K). Figure 6a shows the results for pure Fe and for three FeCr concentrations (0.019 at%, 0.047 at%, and 0.095 at%) which are those in the experiments by Abe and Kuramoto. The first peak is called the ID2 peak and it corresponds to the recombination between self-interstitials and vacancies that belong to the same Frenkel pair, which is known as correlated recombination. The second peak, IE, corresponds to recombinations between self-interstitials and vacancies after a longer migration of the self-interstitial. For the case of Fe, this peak is located at around 140 K and it has a smaller amplitude than the first peak. The first interesting feature of adding Cr is the shift of the IE peak toward lower temperatures, and the shift increases as the Cr concentration increases, as can be seen in Fig. 4a, and as observed experimentally (Abe and Kuramoto 1999; Maury et al. 1987). The reason for this shift is the formation of mixed dumbbells, that is, Fe-Cr self-interstitials that, as given in Table 2, have a lower migration energy than pure Fe dumbbells. However, they also have a low binding energy and therefore a low dissociation energy (see Table 2). At low temperatures, where the IE peak is located, these interstitials can perform several jumps before dissociating.
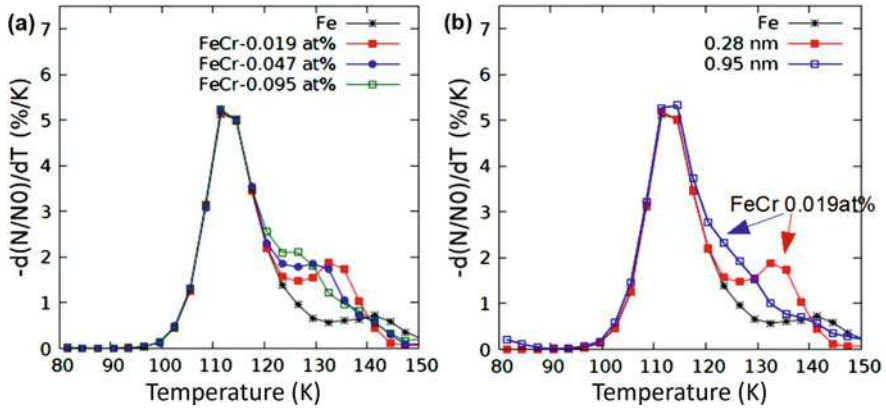
Fig. 6 Derivative of the number of defects as a function of temperature. (**a**) For the case of pure Fe and three FeCr concentrations. (**b**) For FeCr 0.019 at% and different values of the capture radius for I + Cr compared to the case of pure Fe

In these simulations, one parameter that played an important role was the capture radius between the substitutional Cr and self-interstitials. If a capture radius of 0.95 nm is used, like in the case of the interaction between a pure Fe self-interstitial and a vacancy, the IE peak almost disappears, as shown in Fig. 4b, while the best agreement with the experimental data is obtained for a recombination radius of 0.28 nm, that is, the second nearest neighbors distance. This short-range interaction between Cr and a self-interstitial atom can be interpreted as an effective value since according to DFT, not all orientations of these two defects will result into an attractive interaction, in some cases it is repulsive (Olsson et al. 2007; Terentyev et al. 2008).

As an example of the type of information that can be obtained from OKMC, the evolution of defects with temperature for the case of FeCr 0.019 at% is presented in Fig. 7: single vacancies (V), single self-interstitials (I), mixed dumbbells (ICr), vacancy clusters (Vclus), self-interstitial clusters (Iclus), and the different mixed clusters considered in these simulations (ICr2, Cr2I2, and CrI2).

This figure shows how at around 100 K, the number of single vacancies and single self-interstitials starts decreasing. This gives rise to the first recovery peak, ID2. At the same time, the number of interstitial clusters (Iclus) increases, while the number of vacancy clusters (Vclus) decreases. This implies that the species migrating are self-interstitials and not vacancies, which occurs at higher temperatures. The concentration of ICr dumbbells is always very low due to the fact that their dissociation energy is low (0.42 eV). However, they also have a low migration energy, lower than the self-interstitial dumbbell (0.23 eV and 0.34 eV, respectively); therefore, it can perform several jumps before dissociating giving rise to the shift of the IE peak with respect to pure Fe (see Fig. 4a). The low migration of the mixed dumbbell is also responsible for an enhanced recombination
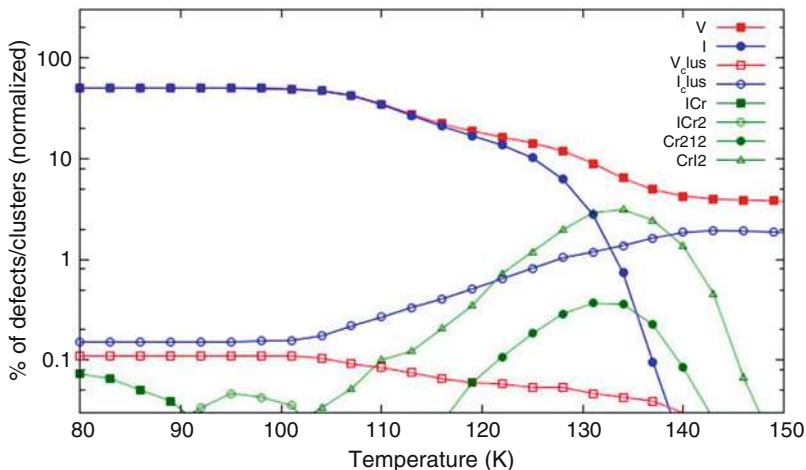
Fig. 7 Evolution of defects with temperature for the case of FeCr 0.019 at%

between vacancies and self-interstitials with respect to pure Fe. ICr can also interact with other ICr, with self-interstitials or with substitutional Cr to form higher order clusters. In this model only those clusters given in Table 2 are included, that is, Cr2I2, I2Cr, and CrI2. For more details on the contribution of these clusters to the recovery stages, see Gámez et al. (2011).

5   Conclusions

As shown above, KMC models have the advantage that the location of every defect (or cluster) in three dimensions is preserved, unlike in rate theory models (RT), where in most applications only the defect concentration is followed. However, KMC is computationally more expensive than RT models. For certain applications, the three-dimensional character of KMC is needed, since spatial correlations must be accounted for to be able to describe some of the features observed experimentally. This is the case of resistivity recovery experiments of irradiated Fe at very low temperature, where the first peaks measured in these experiments can only be explained with the use of KMC models. KMC, however, has the limitation of the system size that can be modeled for the time scales relevant for nuclear applications. Typical OKMC calculations have sizes of a few hundred of nanometers per side on a cubic lattice. In metals, one of the factors that limits the size (or time scale) of the OKMC calculations is the wide disparity of rates between different events. Both in b.c.c. and f.c.c. metals, there are self-interstitial clusters that migrate, according to molecular dynamics simulations, almost athermally, that is, their migration energies are very low (~ 0.1 eV). Moreover, these defects migrate in one-dimension which, for low doses, implies that these defects can travel over long distances without

interacting with other defects, therefore, slowing down the calculation. In these cases, EKMC and FPKMC approaches are more efficient, since defects can jump in one single simulation step several lattice distances.

The efficiency of these algorithms can also be improved by the use of parallel algorithms. Parallelization is relatively simple in an EKMC or FPKMC approach, and there are also algorithms for the case of OKMC such as the synchronous parallel algorithm of Martinez et al. (2008) or parallelization using GPUs (Jimenez and Ortiz 2016).

One of the key issues in KMC models is the accuracy of the parameters used. This is particularly important in the OKMC and EKMC/FPKMC algorithms, trying to avoid the "ad hoc" selection of the possible events included in the calculation. The on-the-fly kinetic Monte Carlo methods, for example, compute the rates of the specific processes at the same time as the algorithm is evolving so that event rates are not tabulated. There are different implementations of the on-the-fly KMC, one of the first ones due to Henkelman and Jonsson (2001). The self-evolving atomistic kinetic Monte Carlo (SEAKMC) developed by Xu et al. (2011), SEAKMC, is able to achieve long simulation times by defining "active volumes" around the defect of interest (Xu et al. 2012, 2015). When the systems to be studied are very complex, such as in defect evolution in concentrated alloys, special methods have been developed to produce the catalogue of transition rates in a more efficient way. One such method consists of using artificial neural networks to predict the values of the energy barriers (Castin and Malerba 2010).

There are still challenges in the development of KMC models such as the implementation of strain fields in a general form, although there are some recent advances (Mason et al. 2014).

References

Abe H, Kuramoto E (1999) Interaction of solutes with irradiation-induced defects of electron-irradiated dilute iron alloys. J Nucl Mater 271-272:209–213. https://doi.org/10.1016/S0022-3115(98)00741-7

Andreani R, Diegele E, Gulden W, Lässer R, Maisonnier D, Murdoch D, Pick M, Poitevin Y (2006) Overview of the European Union fusion nuclear technologies development and essential elements on the way to DEMO. Fus Eng Design 81(1–7):25–32. https://doi.org/10.1016/j.fusengdes.2005.09.005

Becquart CS, Barbu A, Bocquet JL, Caturla MJ, Domain C, Fu CC, Golubov SI, Hou M, Malerba L, Ortiz CJ, Souidi A, Stoller RE (2010) Modeling the long-term evolution of the primary damage in ferritic alloys using coarse-grained methods. J Nucl Mater 406(1):39–54. https://doi.org/10.1016/j.jnucmat.2010.05.019

Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. J Comp Phys 17:10–18. https://doi.org/10.1016/0021-9991(75)90060-1

Castin N, Malerba L (2010) Calculation of proper energy barriers for atomistic kinetic Monte Carlo simulations on rigid lattice with chemical and strain field long-range effects using artificial neural networks. J Chem Phys 132(7):074507. https://doi.org/10.1063/1.3298990

Caturla MJ, Soneda N, Alonso E, Wirth BD, Díaz de la Rubia T, Perlado JM (2000) Comparative study of radiation damage accumulation in Cu and Fe. J Nucl Mater 276(1–3):13–21. https://doi.org/10.1016/S0022-3115(99)00220-2

Dalla Torre J, Bocquet JL, Doan NV, Adam E, Barbu A (2005) JERK, an event-based kinetic Monte Carlo model to predict microstructure evolution of materials under irradiation. Phil Mag 85(4–7):549–558. https://doi.org/10.1080/02678370412331320134

Domain C, Becquart CS, Malerba L (2004) Simulation of radiation damage in Fe alloys: an object kinetic Monte Carlo approach. J Nucl Mater 335(1):121–145. https://doi.org/10.1016/j.jnucmat.2004.07.037

Donev A, Bulatov V, Opplestrup T, Gilmer GH, Sadigh B, Kalos MH (2010) A first-passage kinetic Monte Carlo algorithm for complex diffusion–reaction systems. J Comp Phys 229(9):3214–3236. https://doi.org/10.1016/j.jcp.2009.12.038

Dopico I, Castrillo P, Martin-Bragado I (2015) Quasi-atomistic modeling of the microstructure evolution in binary alloys and its application to the FeCr case. Acta Materalia 95:324–334. https://doi.org/10.1016/j.actamat.2015.05.040

Dudarev SL, Boutard JL, Lässer R, Caturla MJ, Derlet PM, Fivel M, Fu CC, Lavrentiev MY, Malerba L, Mrovec M, Nguyen-Manh D, Nordlund K, Perlado M, Schäublin R, Van Swygenhoven H, Terentyev D, Wallenius J, Weygand D, Willaime F (2009) The EU programme for modelling radiation effects in fusion reactor materials: an overview of recent advances and future goals. J Nucl Mater 386–388:1–7. https://doi.org/10.1016/j.jnucmat.2008.12.301

Fu CC, Dalla-Torre J, Willaime F, Bocquet JL, Barbu A (2005) Multiscale modelling of defect kinetics in irradiated iron. Nat Mat 4:68–74. https://doi.org/10.1038/nmat1286

Gámez L, Gámez B, Caturla MJ, Terentyev D, Perlado JM (2011) Object kinetic Monte Carlo calculations of irradiated Fe-Cr dilute alloys: the effect of the interaction radius between substitutional Cr and self-interstitial Fe. Nucl Inst Meth B 269(14):1684–1688. https://doi.org/10.1016/j.nimb.2010.12.044

Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comp Phys 22:403–434. https://doi.org/10.1016/0021-9991(76)90041-3

Granberg F, Byggmästar J, Sand AE, Nordlund K (2017) Cascade debris overlap mechanism of <100> dislocation loop formation in Fe and FeCr. EPL 119(5):56003. https://doi.org/10.1209/0295-5075/119/56003

Heinisch HL (1990) Computer simulation of high energy displacement cascades. Radiat Eff Def Solids 113:53–73. https://doi.org/10.1080/10420159008213055

Henkelman G, Jonsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115(21):9657. https://doi.org/10.1063/1.1415500

Hornbogen E, Glenn RC (1960) A metallographic study of precipitation of copper from alpha iron. Trans Metall Soc AIME 218:1064–1070

Hyde JM, Marquis EA, Wilford KB, Williams TJ (2011) A sensitivity analysis of the maximum separation method for the characterisation of solute clusters. Ultramicroscopy 111:440–447

Jenkins ML, English CA, Eyre BL (1976) Heavy-ion damage in α Fe. Nature 263:400–401. https://doi.org/10.1038/263400a0

Jimenez F, Ortiz CJ (2016) A GPU-based parallel object kinetic Monte Carlo algorithm for the evolution of defects in irradiated materials. Comput Mater Sci 113:178–186. https://doi.org/10.1016/j.commatsci.2015.11.011

Johnson MD, Caturla MJ, Díaz de la Rubia T (1998) A kinetic Monte–Carlo study of the effective diffusivity of the silicon self-interstitial in the presence of carbon and boron. J Appl Phys 84:1963

Jourdan T, Crocombette JP (2012) Rate theory cluster dynamics simulations including spatial correlations within displacement cascades. Phys Rev B 86:5411. https://doi.org/10.1103/PhysRevB.86.054113

Klaver TPC, Drautz R, Finnis MW (2006) Magnetism and thermodynamics of defect-free Fe-Cr alloys. Phys Rev B 74:094435

Klueh RL, Nelson A (2007) Ferritic/martensitic steels for next-generation reactors. J Nucl Mater 371:37

Knaster J, Moeslang A, Muroga T (2016) Materials research for fusion. Nat Phys 12:424–434. https://doi.org/10.1038/nphys3735

Kuriplach J (2002) Positron-defect interactions in complex systems. App Surf Sci 194:61–70. https://doi.org/10.1016/S0169-4332(02)00090-9

Lanore JM (1974) Simulation de l'evolution des defauts dans un reseau par le methode de monte-carlo. Rad Eff 22(3):153–162. https://doi.org/10.1080/10420157408230773

Levesque M, Martínez E, Fu CC, Nastar M, Soisson F (2011) Simple concentration-dependent pair interaction model for large-scale simulations of Fe-Cr alloys. Phys Rev B 84:184205

Malerba L, Caro A, Wallenius J (2008) Multiscale modelling of radiation damage and phase transformations: the challenge of FeCr alloys. J Nucl Mater 382(2–3):112–125. https://doi.org/10.1016/j.jnucmat.2008.08.014

Mansur LK, Rowcliffe AF, Nanstad RK, Zinkle SJ, Corwin WR, Stoller RE (2004) Materials needs for fusion, generation IV fission reactors and spallation neutron sources – similarities and differences. J Nucl Mat 329–333:166–172. https://doi.org/10.1016/j.jnucmat.2004.04.016

Marian J, Wirth BD, Perlado JM (2002) Mechanism of formation and growth of <100> interstitial loops in ferritic materials. PRL 88(25):255507. https://doi.org/10.1103/PhysRevLett.88.255507

Marinica MC, Williaime F, Crocombette JP (2012) Irradiation-induced formation of Nanocrystallites with C15 laves phase structure in bcc Iron. PRL 108(2):025501. https://doi.org/10.1103/PhysRevLett.108.025501

Martin-Bragado I, Rivera A, Valles G, Gomez-Selles JL, Caturla MJ (2013) MMonCa: an object kinetic Monte Carlo simulator for damage irradiation evolution and defect diffusion. Comp Phys Comm 184:2703–2710. https://doi.org/10.1016/j.cpc.2013.07.011

Martin-Bragado I, Borges R, Balbuena JP, Jaraiz M (2018) Kinetic Monte Carlo simulation for semiconductor processing: a review. Prog Mater Sci 92:1–32. https://doi.org/10.1016/j.pmatsci.2017.09.003

Martinez E, Marian J, Kalos MH, Perlado JM (2008) Synchronous parallel kinetic Monte Carlo for continuum diffusion-reaction systems. J Comp Phys 227:3804–3823. https://doi.org/10.1016/j.jcp.2007.11.045

Martinez E, Fu CC, Levesque M, Nastar M, Soisson F (2011) Simulations of decomposition kinetics of Fe-Cr solid solutions during thermal aging. Solid State Phenom 172-174:1016

Martínez E, Senninger O, Fu CC, Soisson F (2012) Decomposition kinetics of Fe-Cr solid solutions during thermal aging. Phys Rev B 86:224109. https://doi.org/10.1103/PhysRevB.86.224109

Martínez E, Senninger O, Caro A, Soisson F, Nastar M, Uberuaga BP (2018) Role of sink density in nonequilibrium chemical redistribution in alloys. Phys Rev Lett 120:106101

Mason DR, Yi X, Kirk MA, Dudarev SL (2014) Elastic trapping of dislocation loops in cascades in ion-irradiated tungsten foils. J Phy Condensed Matter 26:375701. https://doi.org/10.1088/0953-8984/26/37/375701

Matsui H, Takehana S, Guinan MW (1988) Resistivity recovery in high purity iron after fission- and fusion- neutron irradiation. J Nucl Mater 155-157:1284–1289. https://doi.org/10.1016/0022-3115(88)90512-0

Maury F, Lucasson P, Lucasson A, Faudot F, Bigot J (1987) A study of irradiated FeCr alloys: deviations from Matthiessen's rule and interstitial migration. J Phys F 17:1143. https://doi.org/10.1088/0305-4608/17/5/014

Miller MK (2000) Atom probe tomography. Kluwer Academic/Plenum, New York

Miller MK, Wirth BD, Odette GR (2003) Precipitation in neutron-irradiated Fe–Cu and Fe–Cu–Mn model alloys: a comparison of APT and SANS data. Mater Sci Eng A 353:133–139. https://doi.org/10.1016/S0921-5093(02)00679-2

Mirebeau I, Hennion M, Parette G (1984) First measurement of short-range-order inversion as a function of concentration in a transition alloy. Phys Rev Lett 53:687

Morley A, Sha G, Hirosawa S, Cerezo A, Smith GDW (2009) Determining the composition of small features in atom probe: bcc cu-rich precipitates in an Fe-rich matrix. Ultramicroscopy 109:535–540

Murty KL, Charit I (2008) Structural materials for Gen-IV nuclear reactors: challenges and opportunities. J Nucl Mat 383:189–195. https://doi.org/10.1016/j.jnucmat.2008.08.044

Novy S, Pareige P, Pareige C (2009) Atomic scale analysis and phase separation understanding in a thermally aged Fe–20at.% Cr alloy. J Nucl Mat 384(2):96–102

Olsson P (2009) Ab initio study of interstitial migration in Fe-Cr alloys. J Nucl Mater 386-388: 86–89. https://doi.org/10.1016/j.jnucmat.2008.12.065

Olsson P, Abrikosov IA, Wallenius J (2006) Electronic origin of the anomalous stability of Fe-rich bcc Fe-Cr alloys. Phys Rev B 73(10):104416

Olsson P, Domain C, Wallenius J (2007) Ab initio study of Cr interactions with point defects in bcc Fe. Phys Rev B 75:014110. https://doi.org/10.1103/PhysRevB.75.014110

Opplestrup T, Bulatov V, Gilmer GH, Kalos MH, Sadigh B (2006) First-passage Monte Carlo algorithm: diffusion without all the hops. Phys Rev Lett 97:230602. https://doi.org/10.1103/PhysRevLett.97.230602

Opplestrup T, Bulatov V, Donev A, Kalos MH, Gilmer GH, Sadigh B (2009) First-passage kinetic Monte Carlo method. Phys Rev E 80:066701. https://doi.org/10.1103/PhysRevE.80.066701

Ortiz CJ, Caturla MJ (2007) Simulation of defect evolution in irradiated materials: role of intracascade clustering and correlated recombination. Phys Rev B 75:184101. https://doi.org/10.1103/PhysRevB.75.184101

Ortiz C, Terentyev D, Olsson P, Vila R, Malerba L (2011) Simulation of defect evolution in electron-irradiated dilute FeCr alloys. J Nucl Mater 417(1–3):1078–1081. https://doi.org/10.1016/j.jnucmat.2010.12.188

Papamilhail K, Mergia K, Ott F, Serruys Y, Speliotis T, Apostolopoulos G, Messoloras S (2016a) Fe+ ion irradiation induced changes in structural and magnetic properties of iron films. Nucl Mat Energy 9:459–464. https://doi.org/10.1016/j.nme.2016.03.006

Papamilhail K, Mergia K, Ott F, Serruys Y, Speliotis T, Apostolopoulos G, Messoloras S (2016b) Magnetic effects undiced by self-ion irradiation of Fe films. Phys Rev B 93:100404. https://doi.org/10.1103/PhysRevB.93.100404

Phillips R (2001) Cristals, Defects and Microstructures: Modelling Across Scales. Cambridge University Press, Cambridge

Senninger O, Martínez E, Soisson F, Nastar M, Bréchet Y (2014) Atomistic simulations of the decomposition kinetics in Fe-Cr alloys. Influence of magnetism. Acta Mat 73:97–106

Senninger O, Soisson F, Martínez E, Nastar M, Fu CC, Bréchet Y (2016) Modeling radiation induced segregation in iron-chromium alloys. Acta Mater 103:1–11. https://doi.org/10.1016/j.actamat.2015.09.058

Soisson F, Becquart CS, Castin N, Domain C, Malerba L, Vincent E (2010) Atomistic kinetic Monte. Carlo studies of microchemical evolutions driven by diffusion processes under irradiation. J Nucl Mater 406:55

Swan-Wood TL, Delaire O, Fultz B (2005) Vibrational entropy of spinodal decomposition in FeCr. Phys Rev B 72:024305:1–7

Takaki S, Fuss J, Kugler H, Dedek U, Schultz H (1983) The resistivity recovery of high-purity and carbon doped iron following low-temperature electron-irradiation. Rad Effects 79(1–4):87–122. https://doi.org/10.1080/00337578308207398

Terentyev D, Olsson P, Klaver TPC, Malerba L (2008) On the migration and trapping of single self-interstitial atoms in dilute and concentrated Fe–Cr alloys: atomistic study and comparison with resistivity recovery experiments. Comput Mater Sci 43(4):1183–1192. https://doi.org/10.1016/j.commatsci.2008.03.013

Ullmaier H (ed) (1991) Atomic defects in metals. Landolt-Börnstein, Group III Condensed Matter, vol 25. Springer, Berlin

van der Schaaf B, Diegele E, Laesser R, Moeslang A (2006) Structural materials development and databases. Fus Eng Design 81(8–14):893–900. https://doi.org/10.1016/j.fusengdes.2005.08.040

Vurpillot F, Bostel A, Blavette D (2000) Trajectory overlaps and local magnification in three-dimensional atom probe. Appl Phys Lett 76(21):3127–3129

Woo CH, Singh B, Heinisch HL (1991) Diffusion-based evaluation of defect processes in cascade zones. J Nucl Mater 179-181:951–953. https://doi.org/10.1016/0022-3115(91)90247-5

Xu H, Osetsky YN, Stoller RE (2011) Simulating complex atomistic processes: on-the-fly kinetic Monte Carlo scheme with selective active volumes. Phys Rev B 84:132103. https://doi.org/10.1103/PhysRevB.84.132103

Xu H, Osetsky YN, Stoller RE (2012) Self-evolving atomistic kinetic Monte Carlo: fundamentals and applications. J Phys Condens Matter 24:375402. https://doi.org/10.1088/0953-8984/24/37/375402

Xu H, Stoller R, Osetsky Y, Terentyev D (2013) Solving the puzzle of <100> interstitial loop formation in bcc iron. Phys Rev Lett 110:265503. https://doi.org/10.1103/PhysRevLett.110.265503

Xu H, Stoller RE, Béland LK, Osetsky YN (2015) Self-evolving atomistic kinetic Monte Carlo simulations of defects in materials. Comp Mat Science 100:135–143. https://doi.org/10.1016/j.commatsci.2014.12.026

Yamamoto T, Odette GR, Kishimoto H, Rensman JW, Miao P (2006) On the effects of irradiation and helium on the yield stress changes and hardening and non-hardening embrittlement of ∼ 8Cr tempered martensitic steels: compilation and analysis of existing data. J Nucl Mat 356(1–3):27–49. https://doi.org/10.1016/j.jnucmat.2006.05.041

Yao Z, Hernández Mayoral M, Jenkins ML, Kirk MA (2008) Heavy-ion irradiations of Fe and Fe–Cr model alloys part 1: damage evolution in thin-foils at lower doses. Philos Mag 88(21):2851–2880. https://doi.org/10.1080/14786430802380469

Young WM, Elock EW (1966) Monte Carlo studies of vacancy migration in binary ordered alloys: I. Proc Phys Soc 89:735–746. https://doi.org/10.1088/0370-1328/89/3/329

Zinkle SJ, Busby JT (2009) Structural materials for fission & fusion energy. Mater Today 12(11):12–19. https://doi.org/10.1016/S1369-7021(09)70294-9

Zinkle SJ, Ice GE, Miller MK, Pennycook SJ, Wang XL (2009) Advances in microstructural characterization. J Nucl Mat 386–388:8–14. https://doi.org/10.1016/j.jnucmat.2008.12.302

Rate Theory of Radiation Damage

90

Nasr M. Ghoniem

Contents

Abstract

The theoretical foundations of rate theory applied to radiation damage are presented here. A general stochastic framework for the evolution of atomic clusters within the context of death-and-birth processes is first given. This leads to the well-known master equation, which, in the limiting case of a Poisson process, tends in the macroscopic limit to the mean-field approximation. The background of classical nucleation theory is given, pointing out some of the difficulties and challenges. When atomic clusters grow or shrink by the absorption of single atomic defects, a continuum Fokker-Planck approximation can be derived. This is followed by the mean-field theory of void growth that has been widely used to study the phenomena of swelling and creep. In materials with

N. M. Ghoniem (✉)
Department of Mechanical and Aerospace Engineering, University of California, Los Angeles (UCLA), Los Angeles, CA, USA
e-mail: ghoniem@ucla.edu

high generation rates of helium gas, the more fundamental approach of "cluster dynamics" is developed, and under certain approximations shown to describe continuous nucleation and growth of helium bubbles. Some aspects of spatial behavior of clustering are discussed, particularly the role of cluster motion and fluctuations that lead to spatial self-organization and pattern formation.

1   Introduction and the Physics of Radiation-Generated Defects

Clustering of defects in materials is a process that embodies a rich variety of physical mechanisms and which has significant implications to a wide range of material technologies. Strong deviations from thermodynamic equilibrium conditions drive atomic-size defects into agglomeration, as an attempt to reduce the system's free energy. In nuclear energy applications (both fission and fusion), materials used as nuclear fuels, structural materials, the pressure vessel, as well as instrumentation materials are all subjected to the generation of non-equilibrium concentrations of atomic defects. Energetic neutron collisions with lattice atoms produce intrinsic lattice defects, as well as extrinsic impurities or insoluble gas atoms. This leads to the formation of pure vacancy clusters (voids), clusters of vacancies and gas atoms (bubbles), small interstitial-type and vacancy-type loops, and stacking fault tetrahedra (SFTs).

The generation of point defects in materials stems from two primary types of neutron interactions with bulk materials, known as elastic and inelastic interactions. These are caused by fundamentally different processes and lead to the production of different types of defects within materials. Elastic interactions occur when incoming neutrons transfer their kinetic energy to a host lattice atom. If the energy transferred by the neutron in this collision is greater than the displacement threshold energy, E_{th}, of the lattice atom, then the atom will be ejected from its lattice site. This initial ejected atom is called the primary knock-on atom (PKA). This process leads to the production of a vacancy and a self-interstitial atom (SIA) within the host lattice. Upon ejection from its lattice site, if the PKA still carries greater kinetic energy than E_{th}, it will lead to the ejection of a secondary knock-on atom and so on. This successive chain of elastic collisions and scattering is referred to as cascade damage and has been studied extensively both experimentally and computationally. Inelastic interactions occur under high-energy irradiation conditions leading to (n, He) and (n, p) transmutation reactions within the host lattice. These interactions lead to the production of significant quantities of solute gas atoms within the bulk, namely, helium and hydrogen. Helium is highly insoluble in metals and therefore highly mobile inside matrix materials.

The physics of point defects has been a topic of numerous studies within the field of radiation damage for some time. It is well known that both vacancies and self-interstitial atoms (SIAs) agglomerate into clusters of various sizes. Vacancies combine to form three-dimensional (3-D) voids in the vicinity of the primary damage site, while SIAs typically combine to form 2-D glissile or sessile dislocation

loops on the periphery. Studies have shown that glissile loops are highly mobile and perform 1-D random motion in their slip direction. The production of Frenkel pairs and resulting clusters are known to lead to hardening, swelling, and creep phenomena in irradiated metals. In irradiated materials, point defects and small defect clusters are produced by collision cascades initiated by primary knock-on atoms (PKAs). Following this event, point defects agglomerate, and their rates of agglomeration are determined by fundamental energies of migration, formation, and binding. In this chapter, a systematic presentation of the rate theory of atomic clustering, is given proceeding from the general framework of cluster dynamics to the specific methods of solutions under a variety of approximations.

2   Theory of Cluster Dynamics

Let $C(n, t)$ represent the number of clusters of size n that exist in the system at time t per unit volume. Generally, n can be a composition of various species, such as vacancies and helium atoms with a bubble or several species that cluster together in a precipitate. Let's assume first that n is for one single species (e.g., vacancies forming a void). We will now develop *master equation* to determine $C(n, t)$. A master equation is a set of first-order differential equations describing the time evolution of the probability of a system to occupy each one of a discrete set of states. In the context of atomic clustering, if the total number of atoms is fixed, the probability of finding a defect cluster containing n atoms is proportional to its volumetric concentration.

   We would like to know how many of the clusters will change size so that they will move into the size range ω. Let $W(n; n', \Delta t)$ be the probability that a cluster of size n' will grow to be size n in the time interval between t and $t + \Delta t$. The number of clusters per unit volume growing into the interval ω can be called N^+ and is defined as follows:

$$N^+ = \int_\omega dn \int_\Omega C(n', t)W(n; n', \Delta t)dn' \tag{1}$$

The number of clusters leaving the size interval ω is N^- and is given by:

$$N^- = \int_\omega C(n, t)dn \int_\Omega W(n'; n, \Delta t)dn' \tag{2}$$

In the limit $\Delta t \to 0$, we have $W(n'; n, \Delta t) \to R(n'|n)\Delta t$, and $W(n; n', \Delta t) \to R(n|n')\Delta t$. The net rate of change in the concentration of clusters is thus:

$$\lim_{\Delta t \to 0} \frac{N^+ - N^-}{\Delta t} = \omega \frac{\partial C}{\partial t} = \int_\omega dn \int C(n', t)R(n|n')dn' - \int_\omega dnC(n, t) \int R(n'|n)dn' \tag{3}$$

In the limit of $\omega \to 0$

$$\frac{\partial C(n, t)}{\partial t} = \int_1^\infty [C(n', t) R(n|n') - C(n, t) R(n'|n)] dn' \qquad (4)$$

Equation 4 is the basis for all efforts to find the concentration of defects in irradiated materials and is now commonly known as *cluster dynamics*. In other literature, this is also known as the Smoluchowski-Chapman-Kolmogorov equation (SCK) in probability theory. In its raw form given by Eq. 4, the cluster dynamics equation is not solvable! The reason is that we need to determine transition rates between clusters of arbitrary size, containing possibly several atomic species and existing in widely separated spatial locations within a structure. The number of combinations involved would be overwhelming and would result in an astronomical number of coupled rate equations. For these practical reasons, a variety of approximations have been developed over the years to cast simplified and computable forms of Eq. 4 with varying degrees of approximations. We outline some of the key developments below.

3 Nucleation Theory

3.1 Classical Theory

We will now turn our attention to the specific application of the SCK equation to the nucleation of atomic clusters in a general environment, for example, agglomeration in the gas, liquid, or solid state of small atomic clusters (Katz and Wiedersich 1971; Russell 1971). To this end, we follow a development of *classical nucleation theory*. The basic assumptions of this theory are (1) atomic clusters originate when two atoms come together. (2) These clusters are in equilibrium with the surroundings and can grow or shrink by the addition or removal of single atoms. (3) Most clusters will break up into single atoms, but there is a finite probability that a cluster will grow large enough (i.e., to the *critical size*) that the number of atoms that leave the cluster will just equal those that come to it. (4) If the cluster grows larger than this critical size, it will not shrink but will grow to a macroscopic size. We would expect that the probability of an atom leaving the cluster to decrease, as the cluster grows bigger. Qualitatively this can be seen from considerations of the surface curvature. As the cluster grows, the curvature of the surface will decrease resulting in an increase in the number of nearest neighbors for each surface atom. The more nearest neighbor bonds that have to be broken to free the particle, the less likely it is to break away. This is the origin of the *Gibbs-Thompson effect*. At a point in the growth, the probability that an atom will hit the cluster will just be equal to the probability that one will break away. Once the cluster has grown larger than this size, there will only be a small probability that it will decrease in size, since fewer atoms will leave the cluster than will hit it.

3.1.1   The Becker-Döring Formulation

Richard Becker and Werner Döring constructed a widely used theory of nucleation in the context of ferromagnetism (Becker and Döring 2013). This formulation, now known as classical nucleation theory, has been widely applied in many fields (Ball et al. 1986). In this theory, the transition probabilities (i.e., the Rs) can be calculated in terms of two parameters – the *impingement frequency* β and the *evaporation frequency* α. The transition probability $R(n + 1|n) = \beta A'(n)$ where $A'(n)$ is the area of space around the cluster which will hold an impinging monomer. Similarly, $R(n - 1|n) = \alpha A'(n - 1)$. We can now write for the transition probabilities in the clustering dynamics Eq. 4 as:

$$R(n'|n) = \quad \beta A'(n)\delta(n' - (n + 1)) + \alpha A'(n - 1)\delta(n' - (n - 1))$$
$$R(n|n') = \quad \beta A'(n')\delta(n - (n' + 1)) + \alpha A'(n' - 1)\delta(n - (n' - 1))$$
$$(5)$$

where δ is the Dirac delta function, and we only consider first-order growth (or shrinkage) terms. Substituting these probabilities into Eq. 4, and utilizing the sifting property of the Dirac delta function, results in:

$$\frac{\partial C}{\partial t} = \beta A'(n-1)C(n-1, t) + \alpha A'(n)C(n+1, t) - \beta A'(n)C(n, t) - \alpha A'(n-1)C(n, t)$$
$$(6)$$

This procedure is called *detailed balancing*. At equilibrium, $\frac{\partial C}{\partial t} = 0$ and C can be called the *constrained distribution function*, $C_\circ$. Thus, at equilibrium, we would have:

$$\beta A'(n-1)C_\circ(n-1) + \alpha A'(n)C_\circ(n+1) = \beta A'(n)C_\circ(n) + \alpha A'(n-1)C_\circ(n) \quad (7)$$

If there is no steady growth of clusters through the size classes (i.e., for steady-state conditions), the Eq. (7) can be made more restrictive and would have the form:

$$\beta A'(n - 1)C_\circ(n - 1) = \alpha A'(n - 1)C_\circ(n) \quad (8)$$

or, $\alpha = \beta \frac{C_\circ(n-1)}{C_\circ(n)}$. It should be noted that by using an equilibrium condition, α has been determined by invoking the principle of thermodynamic reversibility. This principle requires that, at equilibrium, every microscopic process takes place at the same rate as its reverse. While α has been found for an equilibrium condition, it is assumed to have the same value under non-equilibrium conditions. Substituting for α in the rate equation (Eq. (6)), we obtain:

$$\frac{\partial C}{\partial t} = \beta A'(n)C_\circ(n)\left[\frac{C(n + 1, t)}{C_\circ(n + 1)} - \frac{C(n, t)}{C_\circ(n)}\right]$$
$$- \beta A'(n - 1)C_\circ(n - 1)\left[\frac{C(n, t)}{C_\circ(n)} - \frac{C(n - 1, t)}{C_\circ(n - 1)}\right] \quad (9)$$

The above equation is a difference-differential equation for nucleation. This equation can be greatly simplified if we assume that $C/C_\circ$ and $DC_\circ$, defined below, vary slowly with n and if we disregard terms of higher than second order. To proceed, we define:

$$D(n) \equiv \beta A'(n) \simeq \beta A(n) \tag{10}$$

Now, expand in a Taylor series about n the following functions (note that the time dependence is implied): $\frac{C(n+1)}{C_\circ(n+1)}$; $\frac{C(n-1)}{C_\circ(n-1)}$; and $D(n-1)C_\circ(n-1)$. Let's perform the first expansion, and one can easily do the other two in a similar fashion.

$$\frac{C(n+1)}{C_\circ(n+1)} = \frac{C(n)}{C_\circ(n)} + \left[\frac{\partial \frac{C(n+1)}{C_\circ(n+1)}}{\partial n}\right]_{n+1=n} (n+1-n) + \frac{1}{2}\left[\frac{\partial^2 \frac{C(n+1)}{C_\circ(n+1)}}{\partial n^2}\right]_{n+1=n} (n+1-n)^2$$

Substituting these expansions into the difference-differential equation yields:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial n}\left[D(n)C_\circ(n)\frac{\partial(\frac{C(n)}{C_\circ(n)})}{\partial n}\right] \tag{11}$$

The above is known as the *isothermal nucleation equation* and has the form of Fick's second law of diffusion in a force field with a variable diffusion coefficient. This observation is consistent with the view that nucleation is a random-walk process through cluster size space. We can now rewrite the nucleation equation as an equation of continuity in *size space*, as follows:

$$\frac{\partial C}{\partial t} + \frac{\partial J(n)}{\partial n} = 0 \tag{12}$$

if we let $J(n) = -DC_\circ \frac{\partial}{\partial n}(C/C_\circ)$, where $J(n)$ is then the cluster current in size space, which is the net number of clusters growing from size n to $n+1$ per unit time, the "nucleation current $J(n)$ can also be written as:

$$J(n) = -D\frac{\partial C}{\partial n} + uC \tag{13}$$

$$= (\text{random walk term}) + (\text{drift velocity, } u, \text{ term})$$

with the "velocity" u given by: $u = D\frac{1}{C_\circ}\frac{\partial C_\circ}{\partial n} = D\frac{\partial}{\partial n}\ln C_\circ$ In principle, we can generalize the above procedure to derive a governing equation for the nucleation of atomic clusters, which contain atomic species α, β, γ, etc., as a continuity equation in an n-dimensional size space, given by:

$$\frac{\partial C(\mathbf{n}, t)}{\partial t} + \nabla \cdot \mathbf{J}(\mathbf{n}) = 0 \tag{14}$$

where $\mathbf{J}$ is the generalized cluster flux in size space, as a function of the vector $\mathbf{n} = \{x_1, x_2, \ldots\}$, which is composed of x_1 of the α-species, x_2 of the β-species, etc.

3.1.2   The Steady State

We now consider the possibility that a steady current of clusters exists without any change in C (i.e., $\partial C/\partial t = 0$). In order to produce such a situation, we must continuously supply monomers to keep $C(1)$ constant and remove all the clusters equal to or larger than a certain size, $\hat{n}$, so that size will not go to infinity (Becker and Doring 1935). Using the equation for the nucleation rate, we have:

$$J_S = -DC_\circ \frac{\partial}{\partial n}(C_S/C_\circ) \tag{15}$$

where the subscript S indicates steady state. Also, from the steady-state model, we can write:

$$(C_S/C_\circ)_{n=1} = 1 \quad \text{and} \quad (C_S/C_\circ)_{n=\hat{n}} = 0 \tag{16}$$

Integrating Eq. (15) for J_S yields: $\int\limits_{n=1}^{n=\hat{n}} -\frac{J_S}{DC_\circ} dn = \int\limits_{1}^{\hat{n}} d\frac{C_S}{C_\circ} = -1$. Or, the steady-state nucleation current is simply given by:

$$J_S = \left[\int\limits_{1}^{\hat{n}} \frac{dn}{DC_\circ} \right]^{-1} \tag{17}$$

In general, we can show that the steady-state *constrained* distribution function is given by:

$$C_\circ(n) = C(1) \exp[-\Delta G(n)/kT] \tag{18}$$

where $C(1)$ is the number of monomers and $\Delta G(n)$ is the free energy to form an atomic cluster of size n. $C_\circ$ has a sharp minimum when ΔG is maximized or when the cluster reaches its *critical size*, $n = n_C$. Thus, we can set $D = D_C$ and approximate $C_\circ$ by expanding $\Delta G(n)$ about the critical size n_C to the second order to get:

$$J_S = D_C C_\circ(n_C) \left[\left(\frac{1}{2\pi k_B T} \right) \frac{\partial^2 \Delta G(n_C)}{\partial n^2} \right]^{1/2} = D_C C_\circ(n_c) \mathcal{Z} \tag{19}$$

where

$$\mathcal{Z} = \left[\left(\frac{1}{2\pi k_B T} \right) \frac{\partial^2 \Delta G(n_C)}{\partial n^2} \right]^{1/2} = \left[\frac{-1}{2\pi} \left(\frac{\partial^2 \ln C_\circ(n_C)}{\partial n^2} \right) \right]^{1/2} \qquad (20)$$

$\mathcal{Z}$ is generally called the *Zeldovich factor* (Zeldovich 1943).

3.1.3   Time Lag

We have not yet made any statements about the speed with which the process of nucleation takes place. In order to make such estimates, we must examine the variation of C and $C_\circ$ with cluster size and introduce a new concept, a critical region of size Δ. The region around $n_C \, \partial C_\circ / \partial n$ will be very small and that J will be dependent mostly upon the random walk term. The first problem is to determine a reasonable width, Δ, for this region, i.e., how far can we go from n_C in cluster size before the increasing drift term begins to have a dominant effect on the direction of cluster growth. A reasonable estimate for Δ can be made by reexamining the nucleation equation as follows:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial n} \left[D C_\circ \frac{\partial C / C_\circ}{\partial n} \right] = \frac{\partial}{\partial n} \left[D \frac{\partial C}{\partial n} - D \frac{C}{C_\circ} \frac{\partial C_\circ}{\partial n} \right] \qquad (21)$$

In the vicinity of n_C, we have $\frac{\partial C_\circ}{\partial n} \cong \left. \frac{\partial C_\circ}{\partial n} \right|_{n_C} + (n - n_C) \frac{\partial}{\partial n} \left. \frac{\partial C_\circ}{\partial n} \right|_{n_C}$, $C_\circ = C(1) \exp [-\Delta G / k_B T]$, and $\ln C_\circ = \ln C(1) - \Delta G / k_B T$. Therefore, $k_B T \frac{\partial^2 \ln C_\circ}{\partial n^2} = -\frac{\partial^2 \Delta G}{\partial n^2}$. However, $\frac{\partial^2 \ln C_\circ}{\partial n^2} = -\frac{1}{C_\circ^2} \left(\frac{\partial C_\circ}{\partial n} \right)^2 + \frac{1}{C_\circ} \left(\frac{\partial^2 C_\circ}{\partial n^2} \right)$. And at $n = n_C$, we have: $\frac{\partial^2 C_\circ}{\partial n^2} = C_\circ \frac{\partial^2 \ln C_\circ}{\partial n^2}$. Thus, $\frac{\partial C_\circ}{\partial n} = (n - n_C) 2\pi \mathcal{Z}^2 C_\circ$ If we set $D = D_C$, the nucleation equation becomes:

$$\frac{\partial C}{\partial t} = D_C \frac{\partial^2 C}{\partial x^2} - u \frac{\partial C}{\partial x}. \qquad (22)$$

where $u = 2\pi D_C \mathcal{Z}^2 x$ and $x = n - n_C$. The nucleation Eq. 22 can now be used to see how the random walk and drift terms interact. Consider what happens to a group of clusters placed at size $n_C + \Delta / 2$. From diffusion theory, the random walk term will cause a spread of clusters in the form of a Gaussian distribution. The drift term, on the other hand, will have the effect of moving the whole distribution to higher value of x. We can see these movements quantitatively if we assume that u is $\approx$ constant and solve the nucleation equation for C.

$$C = \frac{1}{\sqrt{4\pi D_C t}} \exp \left[\frac{-(x - ut - \Delta / 2)^2}{4 D_C t} \right] \qquad (23)$$

A measure of the amount of spread of the clusters is given by the root mean square deviation of x, i.e., $d = (\overline{x^2} - \overline{x}^2)^{1/2} = \sqrt{2D_C t}$, where $\overline{x^2} = \int_{-\infty}^{+\infty} x^2 C(x,t) dx$ to express this quantitatively let $u \cong \pi D_C \mathcal{Z}^2 \Delta =$ constant. Then if we have the correct $\Delta/2$; $\Delta/2 > \sqrt{2D_C t} - ut$. Since $\sqrt{2D_C t} - ut$ will have a maximum at $t = \tau_{\Delta/2}$:

$$\frac{d}{dt}(\sqrt{2D_C t} - ut)\bigg|_{t=\tau_{\Delta/2}} = 0 \qquad \text{or} \qquad \tau_{\Delta/2} = D_c/2u^2 \qquad (24)$$

$\tau_{\Delta/2}$ is thus the time at which the maximum number of clusters have crossed n_C. The inequality then becomes:

$$\frac{\Delta}{2} > \frac{D_C}{u} - \frac{D_C}{2u} = \frac{D_C}{(2\pi D_C \mathcal{Z}^2 \Delta)} \qquad (25)$$

$$\frac{\Delta}{2} > \frac{1}{2\mathcal{Z}\sqrt{\pi}} \qquad (26)$$

The same procedure could be used to find $n_C - \Delta/2$. Thus we may define the critical region as being of width Δ and centered on n_C. Those clusters which are smaller than $n_C - \Delta/2$ will be affected mostly by the drift term and dissolve. Likewise, those clusters larger than $n_C + \Delta/2$ will tend to grow, and those between $n_C - \Delta/2$ and $n_C + \Delta/2$ will tend to be governed by random walk. For completeness, we should mention the usual approach to defining Δ. Consider $\Delta G(n)$ as given in the equation for $C_\circ(n)$, i.e.:

$$C_\circ(n) = C(1) \exp[-\Delta G(n)/k_B T] \qquad (27)$$

The dependence of $\Delta G(n)$ on n is schematically shown in Fig. 1. If a cluster reaches a size which has an energy closer than $k_B T$ to $\Delta G(n_C)$, then there is a good probability that it will move over the free energy *hill*. Thus, we define Δ by the following equation:

$$\Delta G(n_C + \Delta/2) - \Delta G(n_C) = k_B T \qquad (28)$$

Expanding $\Delta G(n_C + \Delta/2)$ about n_C, we get:

$$\Delta G(n_C + \frac{\Delta}{2}) - \Delta G(n_C) = \frac{\Delta}{2} \frac{\partial \Delta G(n_C)}{\partial n}\bigg|_{n_C} + \frac{\Delta^2}{8} \frac{\partial^2 \Delta G(n_C)}{\partial n^2}\bigg|_{n_C} \qquad (29)$$

or $\Delta/2 = \frac{1}{2\mathcal{Z}\sqrt{\pi}}$. This value of Δ is very close to the one we found before. The advantage of the previous treatment was that we obtained a value for the time, $\tau_{\Delta/2}$, which we cannot get from the last procedure. We can now postulate that $C_S/C_\circ$ is approximated by:

$$C_S/C_\circ = \qquad 1 \qquad \text{when} \qquad n < (n_C - \Delta/2)$$

$$C_S/C_\circ = \qquad 0 \qquad \text{when} \qquad n > (n_C + \Delta/2) \qquad (30)$$
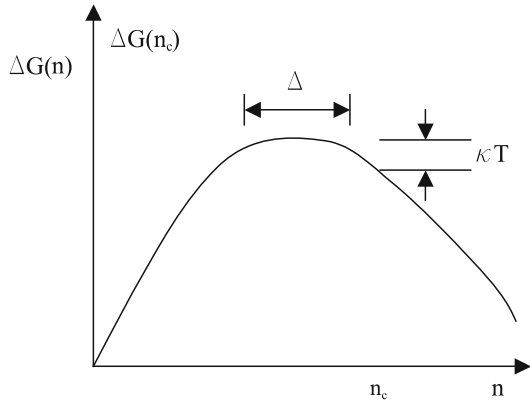
$$\frac{\partial}{\partial n}(C_S/C_\circ) = \quad -\frac{1}{\Delta} \quad \text{when} \qquad (n_C - \Delta/2) < n < (n_C + \Delta/2)$$

Thus using the steady-state equation we get:

$$J_S = -D_C C_\circ(n_C)(-\frac{1}{\Delta}) = D_C C_\circ(n_C)/\Delta \qquad (31)$$

If we then accept as a reasonable value of Δ, $\Delta = 1/\mathscr{Z}$, we get the same result for J_S as we had for the Zeldovich treatment (Eq. 19). We have not directly examined the length of time that would exist between a rapid change in supersaturation and the appearance of visible clusters, but we have all of the information that we need to calculate this delay time. It can be divided into three parts: (1) The time τ_N to nucleate to a size $n_C - \Delta/2$, (2) the time τ_Δ to grow from $(n_C - \Delta/2)$ to $(n_C + \Delta/2)$, and (3) the time τ_G to grow from $(n_C + \Delta/2)$ to visible size. The times τ_N and τ_G will be related to the drift velocity which we have already derived, and the time τ_Δ will be related to $\tau_{\Delta/2}$. Examining τ_Δ first, we remember that $\tau_{\Delta/2}$ was the time for 16% of the clusters to grow from $(n_C + \Delta/2)$ to less than n_C. No more than this number would ever be smaller than n_C because of the effect of the drift velocity to larger sizes.

By symmetry, we see that the same arguments could be applied to clusters growing from $(n_C - \Delta/2)$ to n_C. Therefore, if a particular cluster grows from $(n_C - \Delta/2)$ to n_C, and then on to $(n_C + \Delta/2)$, the total time for this growth will be $2\tau_{\Delta/2}$. Note we have used here the principle of time reversibility, i.e., if a cluster shrinks from $(n_C + \Delta/2)$ to n_C in time $\tau_{\Delta/2}$; the reverse process will take the same amount of time. In summary, we have $\tau_\Delta = D_C/u^2 = \frac{1}{\pi D_C \mathscr{Z}^2}$. We can proceed to find τ_N from the drift velocity, u. By definition, $u(n) = dn/dt$. Therefore, we

have $\int\limits_{0}^{\tau_N} dt = \int\limits_{1}^{(n_C-\Delta/2)} \frac{dn}{-u(n)}$. Here, we use the growth velocity $-u$ or alternatively reverse the time axis and consider the time to dissolve a cluster of size $(n_C - \Delta/2)$. In the same manner, we can calculate $\tau_G = \int\limits_{n_C+\Delta/2}^{n_V} dn/u(n)$. From these times, we can also estimate the time τ_S to reach a steady-state distribution of clusters (of sizes up to $n_C + \Delta/2$). Thus, $\tau_S = \tau_G + \tau_\Delta$.

3.2 Extension to Radiation Damage

Interest in the application of classical nucleation theory to radiation damage grew rapidly in the early 1970s. The first such undertakings were accomplished by Katz and Wiedersich (1971) and Russell (1980) in 1971. These studies extended the classical theory to materials in an irradiation field. This development required consideration of two species (vacancies and interstitials) that are continuously produced as voids nucleate. We start by writing and expression for the Gibb's free energy (GFE) of a system containing a concentration distribution, $N^{eq}(m)$, of clusters of sizes m, can be expressed as

$$G = G_0 + \sum_m N^{eq}(m)g_m - kT \sum_m \ln W_m, \qquad (32)$$

where G_0 is the free energy of the perfect lattice, g_m is the GFE required to form a void of size m, and the last term is the absolute temperature multiplied by the configurational entropy. We note that g_m is typically estimated as the void formation energy (E_m^f) under the assumption of negligible volume and entropy changes in the formation process. Also, W_m represents the number of possible ways in which a void of size m can be distributed in the lattice, which can be shown to be

$$W_m = \frac{m^{N^{eq}(m)}(\frac{1}{m\Omega})!}{[(\frac{1}{m\Omega} - N^{eq}(m)]![N^{eq}(m)]!}. \qquad (33)$$

The chemical potential of a void of size m, volume $m\Omega$, where Ω is the atomic volume, can be defined as the partial derivative of G with respect to the equilibrium concentration of size m voids, or:

$$\mu_m = \frac{\partial G}{\partial N^{eq}(m)} = E_m^f + kT \ln\left[\Omega N^{eq}(m)\right]. \qquad (34)$$

The chemical potential of a single vacancy can therefore be taken as $\mu_v = E_v^f + kT \ln[\Omega C_v]$, where the vacancy formation energy can be substituted from the equilibrium vacancy concentration, given as $C_v^{eq} = \Omega^{-1} \exp\left(-\frac{E_v^f}{kT}\right)$. The vacancy chemical potential can therefore be expressed as:

$$\mu_v = kT \ln \left[\frac{1}{\Omega C_v^{eq}} \right] + kT \ln [\Omega C_v] = kT \ln S_v. \tag{35}$$

where $S_v = C_v/C_v^{eq}$ is the vacancy supersaturation in the lattice. For the vacancies and voids to be in chemical equilibrium, the chemical potential of voids of size m must be equal to m times the chemical potential of an individual vacancy. Therefore, the equilibrium void distribution function can be solved for from (34) and (35) as:

$$N^{eq}(m) = \Omega^{-1} \exp(m \ln S_v - \frac{E_m^f}{kT}). \tag{36}$$

Under non-equilibrium conditions, we can define the nucleation current (in the phase space of cluster size) as the rate at which clusters of size m grow to $m + 1$. This can be expressed as:

$$I = \beta_v(m)N(m) - \alpha_v(m+1)N(m+1) - \beta_i(m+1)N(m+1). \tag{37}$$

Here, the emission of interstitials as a void growth process is neglected due to the large formation energy of interstitials. By the condition of thermodynamic reversibility, the rate of capture of vacancies to clusters of size m should equal the rate of emission of vacancies of clusters of size $m + 1$, and therefore α_v can be solved for, with $C_0(m)$ replaced with $N^{eq}(m)$. Plugging this back into Eq. 37, we obtain:

$$I = -\beta_v(m)h(m) \left[\frac{N(m+1)}{h(m+1)} - \frac{N(m)}{h(m)} \right], \tag{38}$$

where the function $h(m)$ is defined in:

$$\frac{h(m)}{h(m+1)} = \frac{N^{eq}(m)}{N^{eq}(m+1)} + \frac{\beta_i}{\beta_v}. \tag{39}$$

The ratio of impingement frequencies β_i/β_v is referred to as the arrival-rate ratio and plays a large role in determining the critical void size, m_c, which occurs at the minimum of $h(m)$. Approximating the difference in the brackets of (38) as the differential $d(N/h)/dm$, the resulting differential equation can be solved under the boundary conditions that $N/h \to 1$ and 0 as m goes to 1 and ∞, respectively. The nucleation current can then be calculated as:

$$I = \left[\frac{1}{2\pi} \left(\frac{d^2 \ln h}{dm^2} \right)_{m_c} \right]^{1/2} \beta_v(m_c)h(m_c). \tag{40}$$

where the first term on the right-hand side is called the Zeldovich factor and corrects for the fact a fraction of clusters which have grown past the critical size will still shrink to smaller sizes.

The extension of the theory to include the effects of gas atoms resulting from transmutation reactions was also accomplished by Wiedersich et al. (1974) and Russell (1978). The derivation of these heterogeneous nucleation models follows the same structure as the preceding model, under the assumption that gas atoms are static nucleation sites for voids. The number of clusters with j gas atoms per unit volume is taken as M_j, making the total concentration of gas atoms, M, equal to $\sum\limits_{j=1} jM_j$. The Gibb's free energy expression takes the modified form:

$$G = G_0 + \sum_j \sum_m \left[N_j^{eq}(m) g_{mj} - kT \sum_m \ln W_{mj} \right], \tag{41}$$

where N_j^{eq} is the equilibrium distribution of gas-vacancy clusters per unit volume containing m vacancies and j gas atoms. The configurational term can be calculated as:

$$W_{mj} = \frac{M_j!}{[M_j - N_j^{eq}(m)]![N_j^{eq}(m)]!}. \tag{42}$$

The presence of gas atoms in voids leads to an internal pressure that aids in stabilizing voids against dissociation. This has the effect of leading to smaller values of the critical size, as the pressure reduces the effect of surface tension against collapse. Thus, the reversible work of formation g_{mj} will contain a positive contribution due to surface tension and a negative contribution due to the gas pressure. Assuming the gas atoms in the matrix are at an effective pressure p^{eq}, corresponding to the temperature and their concentration, the reversible work of adding gas atoms to a cluster at pressure p is $jkT \ln(p^{eq}/p)$. Assuming the ideal gas law holds, the pressure can be determined from $pV = jkT$, where $V = m\Omega$. Estimates for p^{eq} can be obtained from statistical mechanics models. The reversible work of cluster formation can then be expressed as $g_{mj} = 4\pi R^2 \gamma - jkT \ln\left(\frac{p_{eq}}{p}\right)$, where γ is the surface energy and the cluster radius R can be taken as $(3m\Omega/4\pi)^{1/3}$. For chemical equilibrium to be satisfied, the condition $m\mu_v = \mu_{mj}$ must be satisfied, where the chemical potential of gas-filled clusters is:

$$\mu_{mj} = \frac{\partial G}{\partial N_j^{eq}(m)} = g_{mj} + kT \ln\left[\frac{N_j^{eq}(m)}{M_j} \right]. \tag{43}$$

Using Eq. (35) and assuming chemical equilibrium, the equilibrium distribution of gas atoms in voids can be solved for as:

$$N_j^{eq}(m) = M_j \exp(m\ln S_v - \xi m^{2/3} + j\ln(p^{eq}/p)). \tag{44}$$

where $\xi = (4\pi/kT)(3\Omega/4\pi)^{2/3}\gamma$. The exponential term represents the free energy of cluster formation. Whereas in the homogeneous void nucleation case, this was a function of m alone, this term represents an energy surface dependent on both m and j. Solution of the nucleation current I_j follows (40), with the critical cluster size now a function of j. Following a similar procedure as the homogeneous case, it can be expressed as:

$$I_j = \left[\frac{1}{2\pi}\left(\frac{d^2 \ln h_j}{dm^2}\right)_{m_{cj}}\right]^{1/2} \beta_v(m_{cj})h_j(m_{cj}). \tag{45}$$

where the function h_j is defined in (39) with j subscripts added to $h(m)$ and $N^{eq}(m)$.

4   Mean-Field Theory of Void Growth

The aim of the previous treatment of defect cluster nucleation is to predict the *nucleation rate* of voids or bubbles of a critical size but does not attempt to follow the size distribution to the growth regime. The basic assumption of void growth theory is that a discrete array of point defect sinks can be approximated by a continuous distribution of an absorbing medium into which an *average* void or bubble is placed. This is the key idea of the "mean-field approximation," which we will follow here (Brailsford and Bullough 1972).

In this approach, all sinks are homogenized in an effective medium (that includes, voids, precipitates, bubbles, dislocations, grain boundaries, etc.). The absolute loss rate of interstitials and vacancies within this effective medium is written as $(D_i c_i k_i^2)$ and $(D_v c_v k_v^2)$, respectively, where the D's are diffusion coefficients and the C's are fractional concentrations. The quantities (k_i^{-1}) and (k_v^{-1}) are the mean fee path distance an interstitial or vacancy moves in the medium before getting absorbed. The objective of the theory is to determine the sink strengths (the k's) for each type of sink independently and then add those together is an effective loss rate. Each considered sink is surrounded by a region free of sinks and then enclosed within an effective medium, and a boundary value problem is solved. We follow this approach for a single void of a mean size, but the method has been generalized for all other sinks (see Brailsford and Bullough 1972 for details).

The neutral void sink is assumed to be a small sphere of radius r_{s1} at the center of a larger spherical region of radius R_1 that is free of all other sinks. Voids do not have a preferential attraction toward interstitials over vacancies (unlike dislocations), and hence we will assume that the fractional concentration of each is c. If the defect generation rate is K (displacements per atom per second), then the steady-state concentration c of defects in the sink free region is:

$$\frac{D}{r^2}\frac{d}{dr}\left(r^2\frac{dc}{dr}\right) + K = 0, \quad r_{s1} \leq r \leq R_1 \tag{46}$$

with the boundary condition: $D\frac{dc}{dr} = \bar{K}(c - \bar{c}_1)$ at $r = r_{s1}$. $\bar{K}$ is the transfer velocity across the matrix-void interface. Since K is the jump rate into the sink per available site, the jump rate out (i.e., thermal emission) is $\bar{K}\bar{c}_1$, where $\bar{c}_1 = c^e \exp(F_1\Omega/kT)$. c^e is the thermal equilibrium concentration of defects, and $F_1\Omega$ is the change in energy of the void (volume V and surface area A), and Ω is the atomic volume. $F_1 = \pm\partial(V\delta P + \gamma A)/\partial V = \pm(\delta P + 2\gamma/r_{s1})$. where γ is the surface tension and δP is the excess internal pressure in the void due to any enclosed gas. The $\pm$ sign relates to interstitial and vacancy emission, respectively. At $r = R_1$, the concentration c and its first derivative are continuous with the steady-state concentration in the medium ($r > R_1$). If we assume that n type of sinks simultaneously exist, then $k^2 = k_1^2 + k_2^2 + k_3^2 + \ldots = k_1^2 + k_n^2$, where Dck_1^2 is the loss rate of point defects due to voids. Far away from the void, and neglecting point defect recombination, Eq. 46 now takes the form:

$$\frac{D}{r^2}\frac{d}{dr}\left(r^2\frac{dc}{dr}\right) + K' - DCk^2 = 0, \quad r \geq R_1. \tag{47}$$

The effective defect generation rate in the medium, K_1, is the increased rate due to emission from the neutral sinks, and K_n is the additional increase due to emission from the other $(n-1)$ sink types. Thus, $K' = K + K_1 + K_n$. The solution to Eq. 46 in the boundary layer and Eq. 47 in the lossy medium outside the region R for a dilute sink concentration (i.e., neglecting (r_{s1}/R) compared to unity) is given by (Brailsford and Bullough 1972):

$$c = \mu - \lambda\left(\frac{r_{s1}}{r}\right) - \frac{Kr^2}{6D}, \quad r_{s1} \leq r \leq R_1 \tag{48}$$

$$c = \frac{K'}{Dk^2} + \frac{AR_1}{r}\exp[-k(r - R_1)], \quad r \geq R_1 \tag{49}$$

where μ, A, and λ are given by:

$$\mu = \frac{KR_1^2(3 + kR_1)}{6(1 + kR_1)D} + \frac{K'}{Dk^2}, \quad A = \frac{KR^2}{3D}, \tag{50}$$

$$\lambda = \left[\left(\frac{K'}{Dk^2} - \bar{c}_1\right) + \frac{KR_1^2}{6D}\frac{(3 + kR_1)}{(1 + kR_1)}\right] \tag{51}$$

If C_s is the average concentration of neutral sinks (e.g., voids) in the medium, the flux of point defects can be calculated and from which the "sink strength" of neutral sinks is obtained. The sink strength k_1^2 and the thermal emission rate K_1 from neutral sinks are given by (Brailsford and Bullough 1972):

$$k_1^2 = 4\pi r_{s1}C_{s1}; \quad K_1 = 4\pi r_{s1}C_{s1}D\bar{c}_1. \tag{52}$$

Following similar treatments, the sink strengths of other types of microstructures (e.g., dislocations and precipitates) have been derived for both vacancies and for interstitials. Dislocations are found to attract more interstitials than vacancies, hence creating an imbalance in point defect flows to other features, notably voids. The sinks strengths are shown to be (Brailsford and Bullough 1972):

$$k_i^2 = Z_i \rho_d + 4\pi r_s C_s + Y_i 4\pi r_p C_p \tag{53}$$

$$k_v^2 = Z_v \rho_d + 4\pi r_s C_s + Y_v 4\pi r_p C_p \tag{54}$$

where r_s is the radius of neutral sinks at a concentration of C_s and r_p is the radius of the coherent sinks (precipitates) at a concentration of C_p, and $Y_v = 1$, $Y_i = 1 + (Z_i - Z_v)\rho_d/(Z_v\rho_d + 4\pi r_s C_s)$, $Z_v = 1$, and $1.02 \le Z_i \le 1.08$. The effective defect generation rate from all sinks (K') is obtained by summing all thermal emission terms, thus $K_i' - K = 0$ and $K_v' - K = D_v \left[Z_v\{\rho_d^L \bar{c}_{vl} + \rho_d^N c_v^e\} + 4\pi r_s C_s \bar{c}_{vs} \right]$. The superscripts L and N are for irradiation-induced dislocation loops and network, respectively. Incorporating mutual recombination between vacancies and interstitials with the rate constant α, the steady-state point defect concentrations in the medium (average or mean-field) are given by:

$$K - D_i c_i k_i^2 - \alpha c_i c_v = 0 \tag{55}$$

$$K - D_v c_v k_v^2 - \alpha c_i c_v = 0 \tag{56}$$

Solution of these two equations leads to:

$$c_i = (D_v k_v^2/2\alpha)[-1(1 + \mu) + \{(1 + \mu)^2 + \eta\}^{\frac{1}{2}}], \tag{57}$$

$$c_v = (D_i k_i^2/2\alpha)[-1(1 - \mu) + \{(1 + \mu)^2 + \eta\}^{\frac{1}{2}}] \tag{58}$$

where $\eta = 4\alpha K / D_i D_v k_i^2 k_v^2$ and $\mu = (K' - K)\eta/4K$.

Now, one calculates the excess accumulation rate of vacancies at neutral sinks (voids), which would be given by $\{D_v c_v - D_i c_i - D_v \bar{c}_{vs}\}4\pi r_s C_s$. This is also the volume swelling rate, $\frac{d}{dt}\frac{\Delta V}{V}$. After some algebra, we find $\frac{d}{dt}(\frac{\Delta V}{V}) = \left[\frac{d}{dt}(\frac{\Delta V}{V})\right]_0 F(\eta)$, where:

$$\left[\frac{d}{dt}(\frac{\Delta V}{V})\right]_0 = K(Z_i - Z_v)\rho_d 4\pi r_s C_s (Z_v\rho_d + 4\pi r_s C_s)^{-1/2}$$

$$\times \left\{ Z_i\rho_d + 4\pi r_s C_s + 4\pi r_p C_p \left[1 + \frac{(Z_i - Z_v)\rho_d}{Z_v\rho_d + 4\pi r_s C_s} \right] \right\}^{-1} \tag{59}$$

is the volume swelling rate in the absence of recombination and thermal emission:

$$F(\eta) = (2/\eta)[-(1 + \mu) + \{(1 + \mu)^2 + \eta\}^{1/2} - \zeta\eta], \tag{60}$$
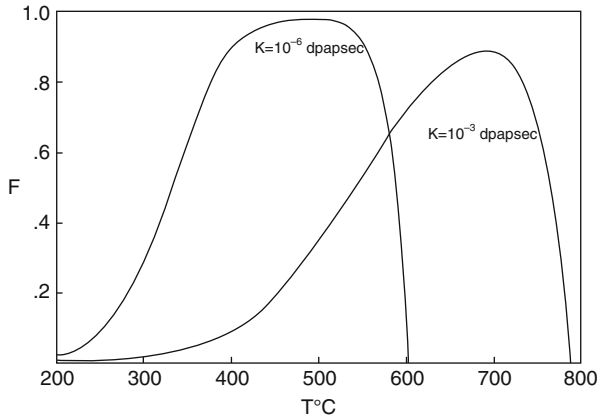
Fig. 2 Variation of the function F with temperature for 316 stainless steel parameters: $C_p = 0$, $\rho_d^L = 10^{-2} \rho_d = 10^9$ cm^{-2}, $Z_i = 1.01$, $Z_v = 1.0$, $4\pi r_s C_s = 10^{11}$ cm^{-2}, $\alpha/D_i = 10^{17}$ cm^{-2}, $D_v = 0.6 \exp(-1.4\,eV/kT)$ cm^{-2} sec^{-1}. (Figure from Brailsford and Bullough 1972)

and

$$\zeta(T) = \frac{D_v(Z_v\rho_d + 4\pi r_s C_s)\{Z_i\rho_d + 4\pi(r_s C_s + r_p C_p)\}}{2K(Z_i - Z_v)\rho_d\{Z_v\rho_d + 4\pi(r_s C_s + r_p C_p)\}} \tag{61}$$

Equation 4 is further approximated with a simple linear time dependence for $[\frac{\Delta V}{V}]_0$, such that:

$$\left(\frac{\Delta V}{V}\right)_0 \approx \frac{K(t - t_0)(Z_i - Z_v)\rho_d 4\pi r_s C_s}{(\rho_d + 4\pi r_s C_s)(\rho_d + 4\pi r_s C_s + 4\pi r_p C_p)} \tag{62}$$

where Kt_0 is an incubation dose (1–10 displacements per atom (dpa)), and is obtained from experiments, and Kt is the current dose in dpa. The main role that the function $F(\eta)$ plays is the temperature dependence of swelling. Calculations of the temperature dependence of the function $F(\eta)$ at two different dose rates are displayed in Fig. 2, taken from reference (Brailsford and Bullough 1972).

5   Coupled Nucleation-Growth Rate Theory

Classical nucleation theory tends to be inadequate for modeling cluster growth because of its high sensitivity to physical parameters, which are typically not known to a high degree of accuracy. This aspect renders it virtually non-predictive. In addition, the theory calculates nucleation rates but does not predict defect cluster density nor size as function of irradiation dose. On the other hand, the mean-field void growth theory presented above assumes that defect clusters (e.g., voids,

bubbles, and dislocation loops) have already nucleated and that their concentrations are constant, independent of dose and spatial location within the material. To remedy these deficiencies and at the same time have a tractable system of cluster dynamics equations, Ghoniem et al. (1985) developed a minimal set of cluster dynamics equations, where assumptions can be made about nucleation of bubbles under high rates of gas production due to nuclear reactions. In this method, a list of coupled rate equations for the concentrations of species is developed. Each equation represents a mass balance equation for a particular defect cluster species as the sum of all production and loss rates of the species to be equal to its concentration rate of change. Under the mean-field approximation, where statistical variations in time and space are averaged, these equations are only time-dependent. The general structure of the cluster dynamics equations reads:

$$\partial_t C_{ij} + \nabla \cdot \mathbf{J}_{ij} = S_{ij} - L_{ij} \tag{63}$$

The volumetric concentration of a cluster containing i Frenkel pairs (negative i for interstitials, and positive i for vacancies) and j helium atoms is C_{ij}; its flux $\mathbf{J}_{ij}$ is composed of two terms: a diffusive flux $-D_{ij}\nabla C_{ij}$ and a convective flux $\mathbf{V}_{ij}C_{ij}$, where the velocity $\mathbf{V}_{ij}$ is proportional to a configurational force $\mathbf{F}_{ij} = -\nabla E_{ij}$ where E_{ij} is an elastic interaction energy and the proportionality parameter (mobility) is $M_{ij} = \frac{D_{ij}}{kT}$. The source and loss terms S_{ij} and L_{ij} can include the processes of absorption, coalescence, and emission. The system of clustering equations represented by Eq. 63 is very generic and cannot be solved without specific details. In practice, however, self-interstitial atom clusters (SIAs) do not associate with helium atoms (energetically unfavorable), and thus negative values of i result in separate equations for SIAs. For positive values of i (i.e., vacancies), they agglomerate with helium atoms j to produce helium-filled bubbles. The practical combination of i and j is determined by the nucleation condition for a stable bubble nucleus, which can be quite large if there is not enough gas atoms to stabilize bubbles. However, when the gas generation rate is high (as in fusion energy neutron spectrum), a critical stable bubble nucleus can have only 1 vacancy and 2–3 helium gas atoms bound to it. Divacancies may likewise bind several gas atoms to form stable bubble nuclei. Because the concentration of single vacancies is always much higher than that of di- or tri-vacancies, an assumption can be made that the bubble nucleus size contains only 1 vacancy and 2–3 gas atoms. Once a critical bubble is formed, the *mean-field* or average bubble can be described by a growth equation similar to our previous discussion of void growth, and there would not be need to determine the concentrations of various size helium-vacancy clusters within the spirit of this description.

Following Ghoniem et al. (1985), rate equations for the concentrations of vacancies ($C_{10} = C_v$), self-interstitials ($C_{\bar{1}0} = C_i$), helium interstitials ($C_{01} = C_g$), helium substitutional atoms ($C_{11} = C_g v$), di-helium interstitials ($C_{02} = C_{2g}$), di-helium mono-vacancy ($C_{12} = C_{2g}v$), the critical nucleus (C^*), the bubble density (C_b), the number of helium atoms in a bubble (m), and the average bubble radius (R) can be expressed as rate equations. Cluster mobility is a very strong

function of its size, and as such only single vacancies, interstitials, and helium gas atoms are assumed to be mobile. The *reduced* set of rate equations derived by Ghoniem et al. (1985) are:

$$\frac{\partial C_v}{\partial t} = D_v \nabla^2 C_v + fG + (\beta e_1 + \delta)C_{gv} - \left[\alpha C_i + \beta C_g + \gamma \left(C_s^v + C_{gv} + 2C_{2g} \right. \right.$$
$$\left. \left. + 2C_{2gv} + 3C^* \right) \right] C_v$$

$$\frac{\partial C_i}{\partial t} = D_i \nabla^2 C_i + fG - \alpha C_i \left(C_v + C_{gv} + 2C_{2gv} + 3C^* + C_s^i \right)$$

$$\frac{\partial C_g}{\partial t} = D_g \nabla^2 C_g + G_{He} + \beta \left(e_1 C_{gv} + e_2 C_{2gv} + 2e_5 C_{2g} \right) + \delta \left(C_{gv} + 2C_{2gv} \right.$$
$$+ 2(2C_{2g}) + 3C^* + m_1 C_b \right)$$
$$+ \alpha C_i \left(C_{gv} + 3C^* \right) - \beta C_g \left(C_v + 4C_g + C_{gv} + 2C_{2gv} \right.$$
$$+ 2C_{2g} + 3C^* + \epsilon C_b \right)$$

$$\frac{\partial C_{gv}}{\partial t} = \beta e_2 C_{2gv} + 2\delta C_{2gv} + \beta C_g C_v + 2\gamma C_v C_{2gv} - C_{gv}(\beta e_1 + \delta + \gamma C_v$$
$$+ \alpha C_i + \beta C_v)$$

$$\frac{\partial C_{2gv}}{\partial t} = 3\delta C^* + \beta C_g C_{gv} + 2\gamma C_v C_{2g} - C_{2gv}(\beta e_2 + 2\delta + 2\beta C_v + 2\alpha C_i + 2\gamma C_v)$$

$$\frac{\partial C_{2g}}{\partial t} = 2\alpha C_i C_{2gv} + 2\beta C_g^2 - C_{2g}(\beta e_5 + 2\delta + 2\gamma C_v + 2\beta C_g)$$

$$\frac{\partial C^*}{\partial t} = 2\beta C_g \left(C_{2gv} + C_{2g} \right) - 3C^* \left(\delta + \alpha C_i + \beta C_g + \gamma C_v \right)$$

$$\frac{\partial C_b}{\partial t} = \frac{4}{m} \left(3\beta C_g C^* \right) + \frac{3}{m} \left(3\gamma C_v C^* \right)$$

$$\frac{\partial m}{\partial t} = \epsilon \beta C_g - \delta m$$

$$\frac{\partial R}{\partial t} = \frac{1}{R} \left(\frac{a_0^2}{48} \right) (\gamma C_v - \alpha C_i - \gamma (e_3 - e_4))$$

$$(64)$$

A constraint on the solution of the reduced set of rate equations above is that the total helium gas concentration must be conserved. Therefore, the total volumetric concentration of helium gas, M should be equal to that produced by irradiation $G_{He} \times t$, that is, $M = G_{He} \times t = (C_g + C_{gv} + 2C_{2gv} + 3C^* + mC_b)$.

Accurate numerical solution of the reduced mean-field rate equations should satisfy this constraint at all times. In solving this system of equations, other microstructure features (e.g., network dislocations, grain boundaries, and precipitates) are assumed to be distributed, and their interaction rates with point defect and helium-vacancy clusters is taken from the mean-field approximation

of rate theory. Expressions for the concentration of helium atoms at precipitates (M_{ppt}) and grain boundaries (M_{GB}) are given as:

$$dM_{ppt}/dt = \varepsilon_{ppt}\beta C_{ppt}C_g - \delta M_{ppt}, \quad dM_{GB}/dt = \beta C_{GB}C_g - \delta M_{GB}.$$

The reaction frequencies of self-interstitials (α), helium (β), and vacancies (γ) and the thermal emission probabilities of vacancies (e_1, e_2) in the above equations are given as $\alpha = 48\nu_i \exp(-E_i^m/kT)$, $\beta = 48\nu_g \exp(-E_g^m/kT)$, $\gamma = 48\nu_v \exp(-E_v^m/kT)$, $e_1 = \exp(-E_{v,g}^b/kT)$, $e_2 = \exp(-E_{v,2g}^b/kT)$, and $\delta = bG$, where the factor of 48 is the assumed combinatorial number, ν is the atomic jump frequency, and E^m and E^b represent migration and binding energies, respectively.

The factor δ is called the re-solution frequency and represents the probability per displacement for dissolving a helium atom back into the matrix (b) times the displacement damage rate (G). Descriptions of the remaining symbols are given in the following nomenclature. The fraction of surviving Frenkel pairs per displacement is f, the displacement damage rate G, the helium production rate G_H, the equivalent matrix-distributed sink density for interstitials C_s^i, the equivalent matrix-distributed sink density for vacancies C_s^v, the equivalent grain boundary sink concentration C_{GB}, the precipitate cavity combinatorial number ε_{ppt}, the equivalent precipitate sink concentration C_{ppt}, and the average number of helium atoms per cavity m. More details of numerical calculations are found in reference Ghoniem et al. (1985). Figure 3 shows the distribution of helium in clusters, bubbles, and grain
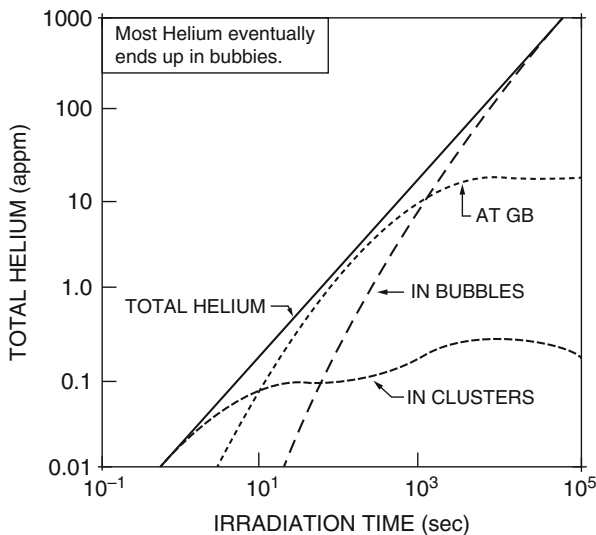
Fig. 3 Distribution of helium gas in the matrix among microstructure features. The irradiation conditions are for 316 stainless steel at 625 °C at a dose rate of 3×10^{-3} dpa/s and a helium/dpa ratio of 5 appm He/dpa (Ghoniem et al. 1985)

boundaries as a function of time. Since the helium injection rate is constant, the total amount of helium is linear in time. During early irradiation times (<0.01 dpa), most of injected helium resides in small helium-vacancy clusters. These are converted to bubbles at a higher dose. The largest proportion of helium ends up in matrix bubbles at doses greater than about 10 dpa. It is observed that during the early stages of irradiation, helium is contained in small clusters. Later, a large proportion goes to grain boundaries.

6 Defect Cluster Size Distribution

Beyond the mean-field theory of nucleation and growth, additional information on defect cluster size or spatial distributions may be necessary. Several approaches to determine the size distribution of SIA (Ghoniem 1989; Kiritani 1973; Hayns 1976) and gas-filled bubbles (Ghoniem and Sharafat 1980; Golubov et al. 2001) have been developed. These approaches are all attempts to develop numerical solutions of the SCK cluster dynamics Eq. 4. We follow here an approach that emphasizes the observation that transitions in cluster size are more often small, either by absorption/emission of single defects or by small size fluctuations resulting from nearby collision cascades. Under these assumptions, the SCK Eq. 4 is approximated as a continuous Fokker-Planck (FP) equation that can be solved numerically. We follow the work of Ghoniem and Sharafat (1980) here.

Let the concentration of clusters containing h helium atoms and u vacancies at time t is given by $C(h, v, t)dhdv = C_{tot}(t)P(h, v, t)dhdv$, where $P(h, v, t)dhdv$ is the probability of finding a cluster in the size interval ($h, u \rightarrow h + dh, v + dv$). The time evolution of the probability function in terms of all possible transitions in the helium-vacancy phase space is given by writing Eq. 4 as:

$$\partial_t P = \sum_k W\{[(h - \Delta h_k), (v - \Delta v_k)] \rightarrow (h, v)\}P[(h - \Delta h_k), (v - \Delta v_k), t]$$

$$= \sum_k W\{(h, v) \rightarrow [(h + \Delta h_k), (v + \Delta v_k)]\}P(h, v, t) \qquad (65)$$

Here Δh_k and Δv_k are fluctuations in the helium and vacancy contents, respectively, which are caused by a stochastic process, k. The transition probability, W, and the cluster probability density, P, can both be expanded around the point (h, v), similar to the procedure that we developed in Sect. 3 to obtain the 2D FP equation given by:

$$\partial_t P + \nabla \cdot (\boldsymbol{F} - \boldsymbol{D} \cdot \nabla)P = 0 \qquad (66)$$

where $\boldsymbol{F}$ and $\boldsymbol{D}$ are drift vector and diffusion coefficient tensor (in size space), respectively. The components of $\boldsymbol{F}$ and $\boldsymbol{D}$ represent first and second moments of the transition probability, respectively, and are given by (Ghoniem and Sharafat 1980).

$$\boldsymbol{F} = \begin{bmatrix} a_{1h} \\ a_{1v} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{D} = \begin{bmatrix} a_{2hv} & a_{2hh} \\ a_{2vv} & a_{2vh} \end{bmatrix} \qquad (67)$$

It is noted, however, that while cascade-induced fluctuations do not influence the magnitudes of the first moments (a_{1h}, a_{1v}), they increase the magnitudes of the second moments (a_{2hh}, a_{2vv}). Numerical solution of Eq. 66 on an infinite quarter phase plane is given in Ghoniem and Sharafat (1980), where a two-moment solution is presented for the specific case that size transitions are only single-step (i.e., no large fluctuations due to cascades). Under these conditions, the rates of helium (k^{gc}), vacancy (k^{vc}) and interstitial (k^{ic}) capture, helium replacement (k^{gr}), and vacancy emission (k^{ve}) can be used to compute the elements of $\boldsymbol{F}$ and $\boldsymbol{D}$. These are given by:

$$a_{1h} = k^{gc} - (k^{ge} + k^{gr}), \tag{68}$$

$$a_{1v} = k^{vc} - (k^{ic+ve} + k^{gr}), \tag{69}$$

$$a_{2hh} = \frac{1}{2}[k^{ge} + k^{gr} + k^{ge}], \tag{70}$$

$$a_{2vv} = \frac{1}{2}[k^{ic+ve} + k^{gr} + k^{vc}], \tag{71}$$

$$a_{2hv} = a_{2vh} = k^{gr} \tag{72}$$

A numerical solution to the FP equation, Eq. 66, has been developed by Ghoniem and Sharafat (1980). Denoting the cluster size by the vector $\boldsymbol{x}$, such that $\boldsymbol{X} = h\boldsymbol{i} + v\boldsymbol{j}$, where $\boldsymbol{i}$ and $\boldsymbol{j}$ are unit vectors along the h and v directions in cluster size space. Taking the first moment of Eq. 66, we obtain $d < \boldsymbol{X} > /dt = < F(\boldsymbol{X}) >$, where the right-hand side represents the drift vector averaged over the probability distribution function. The symbol $< \quad >$ is used to denote this average. This equation is not a closed because we do not know the probability distribution function P a priori. However, to lowest order, one can approximate this equation as:

$$\frac{d < \boldsymbol{X} >}{dt} = F(< \boldsymbol{X} >) \tag{73}$$

The integration of Eq. 73 gives the trajectory of the average cluster in the growth regime of the flow field in vacancy-helium size space. Let us define the variance matrix by (Ghoniem 1989):

$$< \delta X_i \delta X_j > = < X_i X_j > - < X_i >< X_j >, \quad i, j = h, v. \tag{74}$$

Kinetic equations for the variance matrix have been developed in reference Ghoniem (1989) and are given by:

$$\frac{d < \delta X_i \delta X_j >}{dt} = < a_{1j} X_i > - < X_i >< a_{1j} > + < a_{1i} X_j >$$
$$- < X_j >< a_{1i} > + < a_{2,ij} > . \tag{75}$$

Equation 75 is again not closed, and expansions of the parameters around their values at the average trajectory, (x) would result in an open-ended set of moment equations. Although it is possible to develop coupled equations for higher order moments (see Ghoniem 1989), it is sufficient here to develop a lowest order expansion of Eq. 75:

$$\frac{d < \delta X_i \delta X_j >}{dt} \approx a_{2,ij}(< \boldsymbol{X} >). \tag{76}$$

Equation 76 is the lowest order evolution equation for the variance matrix. The second moments of the transition probabilities, $a_{2,ij}$, are the components of the diffusion tensor (68) and are to be evaluated at the average trajectory, $< \boldsymbol{X} >$. We can reconstruct the probability distribution function from its zeroth, first, and second moments. The simplest reconstruction procedure can be based on Gaussian functions, i.e.:

$$P(h, v, t) \approx \left(\delta X_h \delta X_v \sqrt{2\pi} \right)^{-1} \exp\left(-y^2/2 \right),$$

$$\text{where} \quad y = \left[\left(\frac{h - < h >}{\delta X_h} \right)^2 + \left(\frac{v - < v >}{\delta X_v} \right)^2 \right]^{\frac{1}{2}} \tag{77}$$

This FP formulation permits study of the effects of simultaneous helium generation and cascade-induced point defect fluctuations on cavity evolution. This is certainly useful for extrapolating existing radiation damage data to anticipated fusion conditions where the primary differences in damage parameters are the helium generation rate and the cascade size. Cascade effects on the cavity size distribution are included through the second moments of the transition probabilities in the Fokker-Planck equation. Cascades induce fluctuations in point defect concentrations. Since the RMS value of fluctuations depends on the cascade size, we must determine a spectrally averaged value for the variances δX_v, and δX_h,.

The FP model has been applied to the study of interstitial loop size distribution in irradiated materials and has been compared directly to experiments. Figure 4 shows the evolution of the interstitial loop size distribution at various irradiation doses (a) and the effects of stochastic cascade fluctuations on the size distribution in (b). It is clear that collision cascades have a dramatic effect on the broadening of the size distribution as compared to single-step events by point defect absorption/emission.

7   Instabilities and Pattern Formation

Experimental observations have demonstrated the existence of fully or partially ordered (patterned) void, bubble, and dislocation loop microstructure in irradiated materials. These experimental observations are particularly striking for the spatial ordering of bubble and void microstructure in irradiated materials (Sass and Eyre
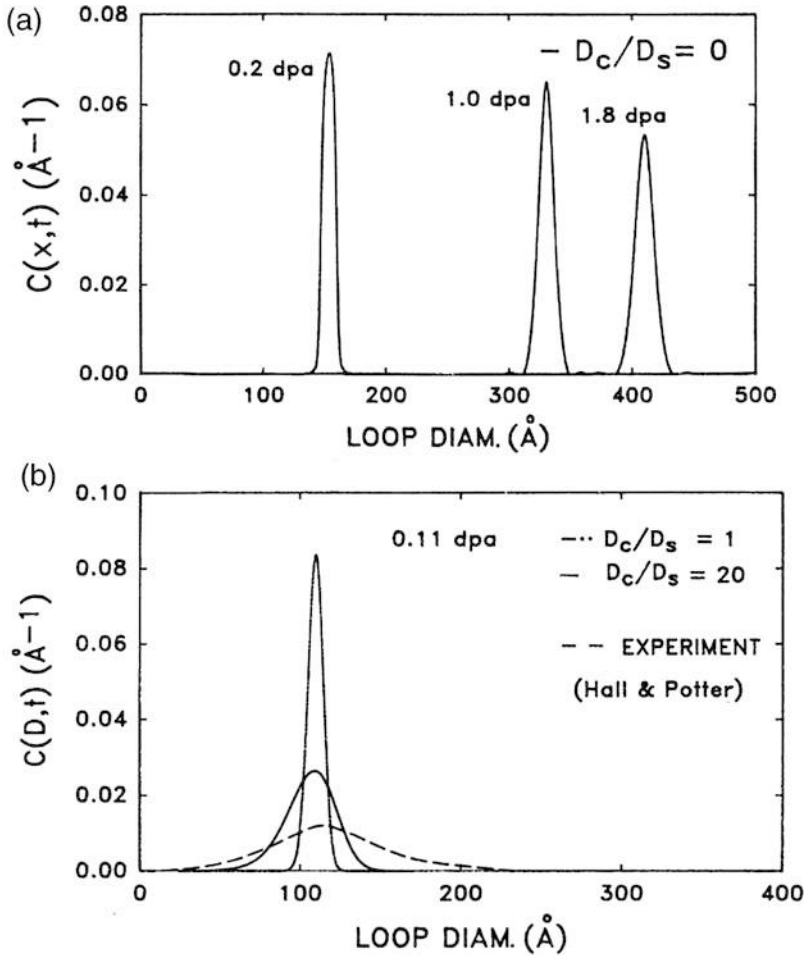
Fig. 4 (**a**) Evolution of the interstitial loop size distribution for single-step point defect absorption/emission. (**b**) Influence of collision cascades on the spread of the size distribution and comparison with the experiments of Hall and Potter (1979)

1973; Johnson et al. 1983; Evans 1971; Mazey and Evans 1986; Kulcinski and Brimhall 1973; Wiffen 1972). An important characteristic of void and bubble lattices is their spatial orientation along crystallographic directions, at least for the three main crystal structures (FCC, BCC, and HCP). Concepts on the behavior of dynamical system have been successfully applied to the problem of microstructure evolution in irradiated materials. Self-organization of irradiated microstructures can thus be viewed as one of the few examples where patterning can be directly linked to the collective dynamical behavior of defects. The void lattice parameter is typically about two orders of magnitude larger than the atomic lattice parameter and about

four times larger than the void diameter in fully developed void superlattices. The void lattice parameter has been reported to initially decrease with increasing dose as the superlattice is being formed (Loomis et al. 1977) and then attains a constant value once the superlattice is fully developed (Evans 1979; Arsenault 1973; Loomis et al. 1977).

Murphy described the coupled evolution of point defects and vacancy loops in the framework of a one-dimensional rate theory dynamical model (Murphy 1987). In this description, point defects are the only mobile species, which diffuse in space. They may annihilate by recombination or through absorption by network dislocations and vacancy loops, which are immobile and are supposed to be created instantaneously by cascade effects.

$$\partial_t c_i = K - \alpha c_i c_v + D_i \partial_x^2 c_i - D_i c_i (Z_{iN} \rho_N + Z_{iV} \rho_V)$$

$$\partial_t c_v = K(1 - \epsilon) - \alpha c_i c_v + D_v \partial_x^2 c_v$$
$$- D_v (Z_{vN}(c_v - \bar{c}_{vN})\rho_N + Z_{vV}(c_v - \bar{c}_{vV})\rho_V)$$

$$\partial_t \rho_V = \frac{1}{|\vec{b}| r_V^0}[\epsilon K - \rho_V (D_i Z_{iV} c_i - D_v Z_{vV}(c_v - \bar{c}_{vV}))] \tag{78}$$

where K is the displacement damage rate of point defects, ϵ the cascade collapse efficiency, responsible for the trapping of part of the vacancies in dislocation loops. ρ_N is the line density of network dislocations, and ρ_V is the line density of vacancy loops, given by $\rho_V = 2\pi r_V N_V$, where r_V and N_V are the mean radius and number density of loops, respectively. The mean loop radius is approximately constant and equal to $r_V^0/2$, where r_V^0 is the initial loop radius. b is the Burgers vector and $Z_{vr} \simeq 1$ and $Z_{ir} \simeq 1 + B$. Furthermore, the network dislocation density is assumed to be uniform and constant in time.

The system (78) admits uniform steady-state solutions c_i^0, c_v^0 and ρ_V^0. Their linear stability analysis has been performed on defining the spatial fluctuations of defect densities through $c_r(x, t) = c_r^0(1 + \bar{c}_r(x, t))$ and $\rho_V(x, t) = \rho_V^0(1 + \bar{\rho}_V(x, t))$ (Murphy 1987). The linear evolution matrix for a Fourier mode of wavenumber k of these fluctuations is given by:

$$\mathbf{M} = \begin{pmatrix} -\bar{K}/c_v^0 - k^2 D_v & -\alpha c_i^0 & -D_v Z_{vV}(c_v^0 - \bar{c}_{vV})\rho_V^0/c_v^0 \\ -\alpha c_v^0 & -K/c_i^0 - k^2 D_i & -D_i Z_{iV} rho_V^0 \\ \dfrac{D_v Z_{vV} c_v^0}{|\vec{b}| r_V^0} & -\dfrac{D_i Z_{iV} c_i^0}{|\vec{b}| r_V^0} & -\dfrac{\epsilon K}{|\vec{b}| r_V^0 \rho_V^0} \end{pmatrix} \tag{79}$$

where $\bar{K} = K(1 - \epsilon) + D_v(Z_{vN} \rho_N \bar{c}_{vN} + Z_{ivV} \rho_V^0 \bar{c}_{vV})$.

If at least one eigenvalue of this matrix has positive real part at finite k, the corresponding eigenmode will grow and destabilize uniform solutions by giving rise to spatial oscillations. The determination of the eigenvalues of this matrix implies the resolution of a cubic characteristic equation. However, since the response of vacancy and interstitial densities to small perturbations are much more rapid

than the response of vacancy loop densities, this characteristic equation may be reduced to a linear one giving the linear growth rate of linear perturbations. This linear growth rate has been computed numerically (Murphy 1987) for parameter values corresponding to 316 stainless steel (Foreman and Makin 1979). It has been found positive for wavelengths in the micron range and network dislocation densities lower than some critical value. An interesting aspect of this analysis is also the temperature dependence of the wavelength of the fastest growing perturbations. For network dislocation densities in the range 10^{12}–$10^{13} \mathrm{m}^{-2}$, this wavelength is slightly increasing with temperature for temperatures less than 500 °C and strongly increasing for temperatures larger than 500 °C, in agreement with experimental results obtained in similar materials (Murphy 1987). It may also be shown that the wavelength decreases with increasing displacement rate. Similar calculations have also been made for neutron-irradiated copper at 250 °C, where, for a network dislocation density of $10^{11} \mathrm{m}^{-2}$, instability was found for a wavelength of approximately 3.4 μm, in agreement with experimental observations (Muncie et al. 1985). Hence, it has been shown here that rate theory is able to predict an instability in the dislocation structure of irradiated materials.

8   Conclusions

Classical nucleation theory has provided easy concepts for understanding the clustering physics of point defects in irradiated materials. However, even with extensions to describe point defect supersaturation and the presence of helium gas in under irradiation, it is not adequately predictive. A more phenomenological treatment of void growth using a mean-field approximation has resulted in qualitative comparisons with many experiments, although such an approach is also limited in its fundamental reach. Coupling nucleation and growth in one single model has also been successful in overcoming the traditional limitations of separate nucleation and growth theories. However, the approach relies on specific assumptions on the critical nucleus size of a bubble being small enough that tractable computations can be performed. Cavity evolution under conditions of cascade damage and high helium generation, which are typical of fusion reactor conditions, is shown to be a continuous process which cannot be easily separated into the classical nucleation and growth regimes. It is also emphasized that the rate of helium transport, which is described as an effective helium mobility, has a strong dependence on bubble evolution. The long transient, associated with high helium mobility, is a result of the concentration buildup of small bubbles clusters which act as the primary traps for migrating helium. However, the quasi-steady-state mobility can be understood in terms of three simple mechanisms: radiation displacement at low temperatures, interstitial replacement at intermediate temperatures, and thermal desorption at high temperatures. The displacement of helium by cascades from bubbles results in significant effects on cavity evolution. First, this displacement process provides for an internal source of helium production which, in turn, causes continuous cavity nucleation. Second, the resulting fine distribution of helium nuclei in the

matrix is associated with slower growth rates for helium-filled cavities. Finally, the stochastic framework represented by Fokker-Planck theory is shown to be a convenient approach to the analysis of interstitial loop and bubble evolution under neutron and ion irradiation conditions.

References

Arsenault RJ (ed) (1973). Defects and defect clusters in bcc metals and their alloys. In: Proceedings of the 1973 international conference held at Gaithersburg, Maryland, August 14–16. Nucl Met Met Soc AIME, vol 18

Ball JM, Carr J, Penrose O (1986) The Becker-Döring cluster equations: basic properties and asymptotic behaviour of solutions. Commun Math Phys 104:657–692

Becker R, Döring W (1935) Kinetische Behandlung der Keimbildung in übersättigten Dämpfen. Annalen der Physik 416(8):719–752

Becker R, Döring W (2013) Ferromagnetismus. Springer, Berlin

Brailsford A, Bullough R (1972) The rate theory of swelling due to void growth in irradiated metals. J Nucl Mater 44:121–135

Evans J (1971) Observations of a regular void array in high purity molybdenum irradiated with 2 MeV nitrogen ions. Nature 229:403

Evans J (1979) Irradiation behavior of metallic materials. In: Poirier J, Dupouy J (eds) Proceedings International Conference on Fast Reactor Core Components CEA, Gif-Sur-Yvette, p 225

Foreman A, Makin M (1979) The effect of vacancy loops on the swelling of irradiated materials. J Nucl Mater 79:43

Ghoniem NM (1989) Stochastic theory of diffusional planar-atomic clustering and its application to dislocation loops. Phys Rev B 39:11810

Ghoniem NM, Sharafat S (1980) A numerical solution to the Fokker-Planck equation describing the evolution of the interstitial loop microstructure during irradiation. J Nucl Mater 92:121–135

Ghoniem NM, Alhajji JN, Kaletta D (1985) The effect of helium clustering on its transport to grain boundaries. J Nucl Mater 136:192–206

Golubov S, Ovcharenko A, Barashev A, Singh B (2001) Grouping method for the approximate solution of a kinetic equation describing the evolution of point defect clusters. Philos Mag A 81:643–658

Hall BO, Potter D (1979) Microstructural development during low-dose irradiation. In: Sprague J, Kramer D (eds) Effects of radiation on structural materials. ASTM International, Philadelphia, pp 32–45

Hayns M (1976) On the group method for the approximate solution of a hierarchy of rate equations describing nucleation and growth kinetics. J Nucl Mater 59:175–182

Johnson P, Mazey D, Evans J (1983) Helium gas bubble superlattice in copper and nickel. Radiat Eff 78:147

Katz JL, Wiedersich H (1971) Nucleation of voids in materials supersaturated with vacancies and interstitials. J Chem Phys 55:1414–1425

Kiritani M (1973) Analysis of the clustering process of supersaturated lattice vacancies. J Phys Soc Japan 35:95–107

Kulcinski GL, Brimhall JL (1973) in American Society for Testing and Materials, Report No.ASTM.STP 529:227–258

Loomis B, Gerber S, Taylor A (1977) Void ordering in ion-irradiated Nb and Nb-1% Zr. J Nucl Mater 68:19

Mazey D, Evans J (1986) Bubble lattice formation in titanium injected with krypton ions. J Nucl Mat 138:16

Murphy S (1987) Spatial instability in dislocation structure under irradiation. Europhys Lett 3:1267

Muncie J, Eyre B, English C (1985) Low-dose neutron irradiation damage in copper. I. Vacancy loop production. Philos Mag A 52:309

Russell KC (1971) The role of quenched-in embryos in solid-state nucleation process. Metall Trans 2:512

Russell KC (1978) The theory of void nucleation in metals. Acta Metallurgica 26:1615–1630

Russell KC (1980) Nucleation in solids: the induction and steady state effects. Adv Colloid Interf Sci 13:205–318

Sass S, Eyre B (1973) Diffraction from void and bubble arrays in irradiated molybdenum. Philos Mag 27:1447

Wiedersich H, Burton J, Katz JL (1974) Effect of mobile helium on void nucleation in materials during irradiation. J Nucl Mater 51:287–301

Wiffen F (1972) The effect of alloying and purity on the formation and ordering of voids in BCC metals. In: Proceedings of the 1971 international conference on radiation-induced voids in metals. Albany, New York, p 386

Zeldovich J (1943) Acta Phys-Chem URSS 18:7

Discrete Dislocation Dynamics Simulations of Irradiation Hardening in Nuclear Materials

91

Jaime Marian, Steve Fitzgerald, and Giacomo Po

Contents

J. Marian (✉)
Department of Materials Science and Engineering and Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA

Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA
e-mail: jmarian@ucla.edu

S. Fitzgerald
School of Mathematics, University of Leeds, Leeds, UK
e-mail: S.P.Fitzgerald@leeds.ac.uk

G. Po
Department of Mechanical and Aerospace Engineering, University of California Los Angeles, Los Angeles, CA, USA
e-mail: gpo@ucla.edu

2243

Abstract

Neutron irradiation can severely impact the mechanical behavior of nuclear structural materials. Irradiation introduces a high density of nanometric defects that block dislocation motion and result in hardening and a loss of ductility often associated with the onset of localized plastic flow in a highly heterogeneous manner. In our hierarchy of numerical methods used to understand and quantify mechanical property degradation under irradiation, discrete dislocation dynamics (DD) provides a window into the time and length scales where critical interactions between dislocations and defects occur. In the present chapter, we discuss the current state of the art of DD simulations applied to irradiation scenarios, introducing the theoretical models devised to deal with dislocation-defect interactions, crystal structure particularities, and the most salient aspects of several highlighted applications. We also discuss current limitations and what new understanding has been gained vis-a-vis serviceable nuclear materials using these techniques.

1   Introduction

Irradiation hardening is a complex phenomenon involving the interplay between a material's microstructure and irradiation-generated defects, and its effects on mechanical behavior. This complexity is generally owed to the existence of nontrivial dependencies between the mechanical response of a material and a relatively large number of internal (temperature, chemical composition, internal microstructure, etc.) and external (dose rate, dose, external stress, etc.) variables, which are often correlated with one another. The reader is referred to reviews of the evolution in our understanding of radiation damage processes and mechanical properties that exist on the topic for more details (Matzke 1992; Sickafus et al. 1999; Odette and Lucas 2001; Ackland 2010).

Before the advent of large-scale simulations and high-performance computing, the prevalent thinking in irradiation hardening of materials was dominated by two main models. The first one is commonly known as the *dispersed barrier hardening* (DBH) model, originally proposed by Seeger in the 1950s (Seeger 1958). The main idea behind the DBH model is that dislocation loops and vacancy clusters produced directly within displacement cascades act as barriers to gliding dislocations, resulting in hardening (defect production in the cascade region is also the main premise behind other irradiation damage theories such as the *production bias model* and swelling). Seeger's approach is substantially equivalent to the well-known Orowan mechanism for hardening due to precipitates and dispersed particles (Singh et al. 1997; Lambrecht et al. 2010) applied to irradiated materials. Consequently, the yield strength increase from irradiation is then determined from the bowing stress required to overcome the irradiation obstacles. The second model used to explain irradiation hardening is known as the *loop decoration hardening* model, which assumes the formation of defect "clouds" around the cores of ingrown

dislocations (i.e., dislocations present in the material before irradiation) (Blewitt et al. 1960). These dislocations cannot then act as dislocation sources until they can pull away or free themselves from the clouds of defects (Young 1962; Holmes 1964). In such way, the stress necessary to unlock these dislocations so that they can act as dislocation sources again is then likely to determine the yield strength of the irradiated crystals. This mechanism of radiation hardening is commonly known as "source" hardening and is conceptually equivalent to the formation of Cottrell atmospheres around dislocations by solute atoms during plastic flow (Cottrell and Bilby 1949).

With the emergence of large-scale molecular dynamics (MD) simulations, it was soon established that dense cascades in metals result in damage in the form of self-interstitial atom (SIA) clusters (Marian et al. 2003; Bacon et al. 2006; Terentyev et al. 2008, 2010) and various forms of vacancy-type clusters ranging – depending on the material and the cascade energy – from small nanovoids, stacking fault tetrahedra (SFT) (Wirth et al. 2002; Osetsky et al. 2005, 2006; Zinkle 2012), and vacancy loops (Diaz de La Rubia et al. 1989; Calder and Bacon 1993; Phythian et al. 1995). Moreover SIA clusters can evolve into dislocation loops displaying one-dimensional migration, while vacancy-type clusters typically display little or no mobility. Among others, these observations led to the adoption of the combined *dislocation* and *production* bias models as the most successful theories to explain irradiation damage accumulation that we have today (Woo and Goesele 1983; Woo and Singh 1990; Golubov et al. 2000). While information from MD cascade simulations has helped inform our understanding of irradiation hardening, it has not settled the question as to what is the prevalent mechanism behind yield strength increases. In addition, experimental observations have clearly established the existence of a sharp and prominent yield drop both in irradiated face- and body-centered cubic (FCC, BCC) metals, associated with the onset of plastic flow localization via narrow "defect-free" dislocation channels above a certain dose (cf. Sect. 4.2). These channels concentrate all deformation in a highly heterogeneous way leading to a plastic instability that is seen to severely compromise irradiated materials' ductility. This instability is also observed in polycrystalline metals and alloys in the form of the absence of work hardening. While some of these features are a priori incompatible with a DBH interpretation, irradiation can also result in sub-thermodynamic precipitation and/or the formation of gas bubbles, which are perfectly consistent with the DBH model.

In any case, irradiation hardening and flow localization are collective plastic phenomena that cannot be approached solely from an atomistic simulation point of view. Kinetic methods such as kinetic Monte Carlo (KMC) and cluster dynamics can push the time- and length-scale envelope beyond those attainable atomistically, although often at the expense of physical accuracy (see corresponding chapters on KMC and rate theory methods in this Volume. Ideally, the simulation tool of choice should capture long-range elastic interactions between dislocations and local (short-range) effects describing defect-dislocation interactions. Dislocation dynamics (DD) is such a technique, designed as an accurate method to calculate strain hardening and collective dislocation processes. From their original inception

in the late 1980s, the last two decades have seen tremendous improvements in the development of efficient DD algorithms applied to all kinds of different scenarios. Indeed, DD suggests itself as the ideal technique to study irradiation hardening, as it naturally captures both long- and short-range interactions mentioned above. The purpose of this chapter is to introduce the technique, discuss its theoretical underpinning, and review the most notable applications within the realm of irradiated materials.

Notwithstanding earlier advances in the development of closed-form expressions for the stress fields of arbitrary dislocation segments, the method of discrete dislocation dynamics was originally introduced in the late 1980s/early 1990s independently by researchers in France and the United States, first as a 2D method (Lépinoux and Kubin 1987; Ghoniem and Amodeo 1988; Gulluoglu et al. 1989), quickly followed by extensions to three dimensions (Amodeo and Ghoniem 1990; Kubin and Canova 1992; Kubin et al. 1992). The technique enjoyed early success fueled by a wide range of applications and improvements in computational power and was picked up by multiple groups worldwide who have used it and made notable contributions to our understanding of, among others, dislocation junction formation, strain hardening, plasticity in confined volumes, patterning, etc. For more information, the reader is referred to the chapter on dislocation dynamics simulations in the accompanying volume to this series on multiscale modeling of materials.

In terms of the use of DD for the simulation of irradiation hardening in materials, the early works started around the late 1990s and relied heavily on MD simulations of dislocation-obstacle interactions to provide the physical nexus between cascade defects and gliding dislocations. Therefore, the evolution of the DD technique as it relates to nuclear material applications is strongly linked to the availability of reliable descriptions of the interactions between dislocations and irradiation obstacles, which have been typically simulated atomistically. Currently, there are examples in the literature of simulations including point defects, dislocation loops, stacking fault tetrahedra, voids, precipitates, and grain boundaries, as well as demonstrations of the formation of defect-free channels following the hardening peak. This is the object of Sects. 3 and 4 in this chapter.

2 Brief Overview of DD Theory and Its Implementation

Discrete dislocation dynamics is a method intended for the study of the collective evolution of ensembles of crystal dislocations under applied loads. In DD, macroscopic system properties (e.g., stress-strain response) are computed directly from the evolution of the dislocation microstructure, therefore establishing a direct link between microstructure and properties. In contrast to atomistic methods where the degrees of freedom of the system are individual atomic positions, in DD, the degrees of freedom are the positions of line defects (dislocations) in three-dimensional space (two-dimensional implementations of DD are not considered in this chapter). The main objective of the DD method is to evolve the discrete dislocation configuration

in time. Typically it is assumed that dislocation motion is overdamped, so that it is possible to define the dislocation velocity of an elementary line segment as a function of local quantities such as stress σ and line orientation ξ and possibly other parameters like Burgers vector $\boldsymbol{b}$, slip plane normal $\boldsymbol{n}$, and temperature T:

$$\boldsymbol{v} = \hat{\boldsymbol{v}}(\sigma, \xi, \boldsymbol{b}, \boldsymbol{n}, T) . \tag{1}$$

The function $\hat{\boldsymbol{v}}$ is known as the *mobility law*. The stress field is composed of the stress field exerted by the dislocation network and either a uniform "external" stress associated with elastic deformation of the supercell in periodic simulations or the elastic "correction" field which accounts for mechanical boundary conditions in finite-volume simulations. In irradiated materials, irradiation-induced defects such as voids and precipitates may also contribute to the local stress field.

The dislocation stress is computed from the *static* elastic theory of dislocations; hence rendering DD is a quasistatic method, despite its name. The elastic fields engendered by a dislocation loop $\mathscr{L}$ can be derived in the infinitesimal-deformation eigendistortion theory of Mura (1987). For an infinite isotropic medium, the displacement and stress fields generated by the loop are, respectively,

$$\boldsymbol{u}^{\infty}(\boldsymbol{x}) = -\frac{\boldsymbol{b}\Omega}{4\pi} - \frac{1}{8\pi(1-v)} \oint_{\mathscr{L}} \frac{1}{R} \left\{ (1-2v)\,\boldsymbol{b} \times \hat{\boldsymbol{\xi}}' + \left[\hat{\boldsymbol{R}} \cdot \left(\boldsymbol{b} \times \hat{\boldsymbol{\xi}}' \right) \right] \hat{\boldsymbol{R}} \right\} d\ell' \tag{2}$$

$$\begin{aligned} \boldsymbol{\sigma}^{\infty}(\boldsymbol{x}) = {} & \frac{\mu}{4\pi(1-v)} \oint_{\mathscr{L}} \frac{1}{R^2} \left\{ (1-v) \left[\hat{\boldsymbol{\xi}}' \otimes (\boldsymbol{b} \times \hat{\boldsymbol{R}}) + (\boldsymbol{b} \times \hat{\boldsymbol{R}}) \otimes \hat{\boldsymbol{\xi}}' \right] \right. \\ & + \left[(\boldsymbol{\xi}' \times \boldsymbol{b}) \otimes \hat{\boldsymbol{R}} + \hat{\boldsymbol{R}} \otimes (\boldsymbol{\xi}' \times \boldsymbol{b}) \right] \\ & \left. + \hat{\boldsymbol{R}} \cdot (\boldsymbol{b} \times \hat{\boldsymbol{\xi}}') \left[3\hat{\boldsymbol{R}} \otimes \hat{\boldsymbol{R}} + \boldsymbol{I} \right] \right\} d\ell' . \end{aligned} \tag{3}$$

In Eqs. (2) and (3), $\boldsymbol{b}$ is the Burgers vector, μ is the shear modulus, and v the Poisson ratio. Moreover, $\boldsymbol{R} = \boldsymbol{x} - \boldsymbol{x}'$, $R = \|\boldsymbol{R}\|$, $\hat{\boldsymbol{R}} = \boldsymbol{R}/R$, and $\Omega = \int_{\mathscr{S}} R_j/R^3 \, dA'_j$ is the solid angle subtended by the loop. Note that these fields are singular when computed at any point of the dislocation line itself ($\boldsymbol{R} = 0$) and regularization methods are necessary (Cai et al. 2006; Po et al. 2014a, 2017).

The mobility law, Eq. (1), defines the velocity point-wise along a dislocation line. However, because the line is a flexible object with a potentially infinite number of degrees of freedom, the numerical solution of the loop evolution equation requires that a discretization procedure be introduced to reduce the number of degrees of freedom to a finite set. To this end, an arbitrary dislocation configuration is discretized in line elements having constant Burgers vector, called *segments*, connected at *nodes*. Conservation of the Burgers vector requires that the vectorial sum of the Burgers vector vanishes for each node. Different discretization methods and evolution strategies have been implemented, and we distinguish the following three methods.

1. *Segment-based DD*. In the segment-based implementation, the dislocation network is composed of piecewise linear segments. Within a certain glide plane, segment orientation varies discretely among a predetermined set, which includes edge, screw, and junction orientations (Devincre et al. 2011). This pre-defined set of segment orientations is used to tabulate and speed up the calculation of the stress field which each segment generates. For each segment, the sum of the external stress field and the stress field due to nearby segments is computed at one point along the segment, typically the midpoint. This midpoint stress is then used to compute the segment velocity based on material-specific mobility laws, which in general can be expressed as Eq. (1). The midpoint velocity is then used to advance the position of each segment, and finally nodes are reconnected taking advantage of the discrete set of segment orientations, as illustrated in Fig. 1a.

2. *Node-based DD*. The starting point of the nodal DD method (Weygand et al. 2002; Arsenlis et al. 2007) is the equation expressing the condition of force equilibrium for an elementary dislocation line. If we assume that the drag force exerted by the lattice is proportional to the velocity $\boldsymbol{v}$, then the condition of force equilibrium reads

$$\boldsymbol{B}(\boldsymbol{\xi}, \boldsymbol{n}, T)\boldsymbol{v} = \boldsymbol{\sigma b} \times \boldsymbol{\xi} \tag{4}$$

where $\boldsymbol{B}(\boldsymbol{\xi}, \boldsymbol{n}, T)$ is a drag coefficient possibly dependent on slip system, line orientation, and temperature and $\boldsymbol{f}^{PK} = \boldsymbol{\sigma b} \times \boldsymbol{\xi}$ is the mechanical force exerted by the stress field on the dislocation (Peach-Koehler force). Clearly the
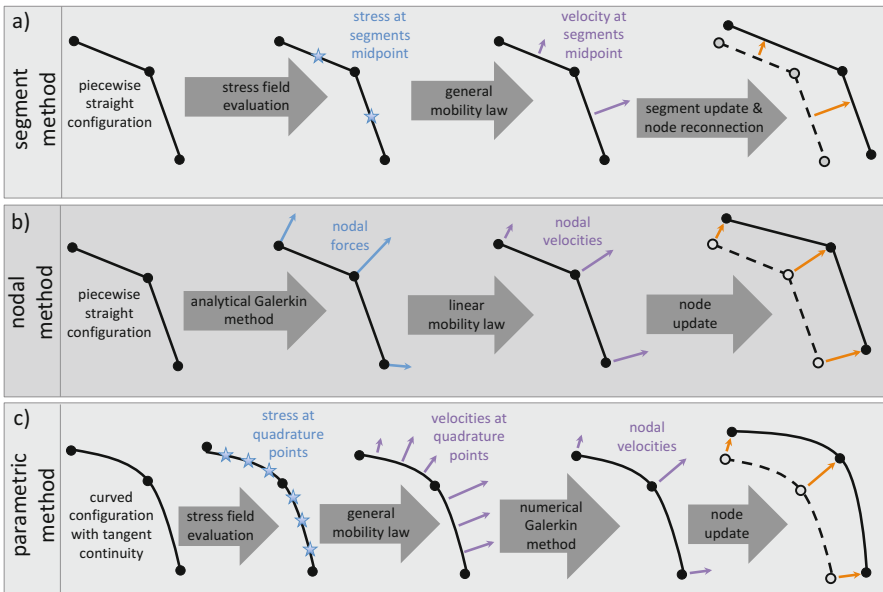
Fig. 1 Comparison between (**a**) segment, (**b**) nodal, and (**c**) parametric DD methods

condition (4) can be regarded as a special case of the mobility law (1) if we take $\hat{\boldsymbol{v}}(\boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{b}, \boldsymbol{n}, T) = \boldsymbol{B}^{-1}(\boldsymbol{\sigma}\boldsymbol{b} \times \boldsymbol{\xi})$. In order to solve (4) numerically, the dislocation network is discretized into piecewise straight segments with constant Burgers vector meeting at nodes. Equation (4) is then assembled as a discrete system of equations for the nodal velocities:

$$\mathbb{B}\boldsymbol{V} = \boldsymbol{F} \tag{5}$$

which can be solved via discrete numerical methods such as finite elements. The mobility matrix $\mathbb{B}$ is sometimes approximated by its diagonal in order to speed up the computation of the nodal velocities (Arsenlis et al. 2007). The nodal method is sketched in Fig. 1b.

3. *Parametric DD*. The parametric DD uses the Galerkin finite element method to obtain nodal velocities (Ghoniem et al. 2000; Po and Ghoniem 2014; Po et al. 2014b). In general, however, the weak form is constructed from the general mobility law (1) instead of the linear law (4). The parametric method allows for spline segments with tangent continuity across nodes, which in general require a smaller number of nodal degrees of freedom to approximate curved loops compared to linear segments. The use of nonlinear mobility laws is particularly important in BCC metals, where non-Schmid stress components influence the mobility of screw dislocations.

The above methods are discretization strategies suited to obtain time-marching schemes for the degrees of freedom describing the dislocation network configuration. In addition to the time-marching scheme, a typical DD timestep comprises several discrete events which are now briefly summarized.

- *Network re-discretization*. The length of individual dislocation segments may change drastically during the course of a simulation. Dedicated algorithms are therefore necessary to locally increase or decrease the network discretization by local addition or removal of nodes and segments.
- *Cross-slip*. Screw dislocation segments can change their glide plane by a mechanism called *cross-slip*. Cross-slip is one of the most important length multiplication and obstacle avoidance mechanisms available to gliding dislocations. In FCC metals, cross-slip may require recombination of the Shockley partial dislocations and is typically implemented as a thermally activated phenomenon (Kubin 2013).
- *Junction formation and breaking*. When pairs of attractive dislocation segments collide, it is numerically advantageous to replace the original segments with a new segment having Burgers vector equal to the vectorial sum of the original Burgers vectors. This operation requires the implementation of both efficient collision detection algorithms between segments and topological operations that change the connectivity of the network. Dislocation junctions are passively released by unzipping of the junction segments or actively by dedicated algorithms.

2.1   Considerations Related to Boundary Conditions

2.1.1   Using Periodic Boundary Conditions for Strain Hardening Simulations

From its original limitations as a 2D method or a 3D method consisting only of closed dislocation loops, DD has been successfully extended to study bulk representative volumes by the use of periodic boundary conditions (PBC) (Fivel and Canova 1999; Bulatov et al. 2000; Madec et al. 2004). Dislocation configurations compatible with the use of periodic boundary conditions must result in net zero integral over the primary box of the so-called dislocation density or Nye's tensor. Although dislocation structures that satisfy this condition are not unique, calculations have shown that the degree of arbitrariness introduced in periodic reconstructions is small for large-scale DD simulations (Bulatov et al. 2000). The general idea is to use closed loops or infinite dipoles that ensure the translational periodicity along each spatial dimension. Using PBC, however, introduces an artifact in the form of self-annihilation of dislocation lines. Nevertheless, by setting the values of the self-annihilation distances and ensuring that the resulting dislocation mean free path is physical, one can mitigate the implications of this effect (Madec et al. 2004), and indeed highly accurate DD simulations using PBC are now commonplace in the literature.

2.1.2   Coupling to Finite Elements for Simulations of Confined Volumes

The dislocation stress and displacement fields $\boldsymbol{u}^{\infty}(\boldsymbol{x})$ and $\boldsymbol{\sigma}^{\infty}(\boldsymbol{x})$ given in (2) and (3) are valid for an infinite medium. In the case of a finite domain $\mathscr{B}$, the state of internal stress driving dislocation motion is the superposition of the stress $\boldsymbol{\sigma}^{\infty}(\boldsymbol{x})$ and a "correction" stress $\boldsymbol{\sigma}^{c}(\boldsymbol{x})$ which results from the solution of an elastic boundary value problem (BVP) (Eshelby 1979; Van der Giessen and Needleman 1995; Deng et al. 2008; Weygand et al. 2002; Po et al. 2014b). The BVP solves for the displacement field $\boldsymbol{u}^{c}(\boldsymbol{x})$ which satisfies

$$\begin{cases} \sigma_{ij}^{c} = c_{ijkl}u_{k,l}^{c} & \text{in } \mathscr{B} \\ \sigma_{ij,j}^{c} = 0 & \text{in } \mathscr{B} \\ \sigma_{ij}^{c}\hat{n}_{j} = p_{i} - \sigma_{ij}^{\infty}\hat{n}_{j} & \text{on } \partial_{N}\mathscr{B} \\ u_{i}^{c} = \bar{u}_{i} - u_{i}^{\infty} & \text{on } \partial_{D}\mathscr{B} \end{cases} \tag{6}$$

In (6) $\partial_{N}\mathscr{B}$ is the portion of the boundary over which the external traction p_{i} is applied, while $\partial_{D}\mathscr{B}$ is the portion of the boundary over which a displacement boundary condition $\bar{u}$ is prescribed. Once the solution of this BVP is found, the displacement field and stress field in the crystal are computed as $u_{i} = u_{i}^{\infty} + u_{i}^{c}$ and $\sigma_{ij} = \sigma_{ij}^{\infty} + \sigma_{ij}^{c}$, respectively. The BVP (6) is typically implemented using the Galerkin FEM. It can be shown that the total solution is independent of virtual loops
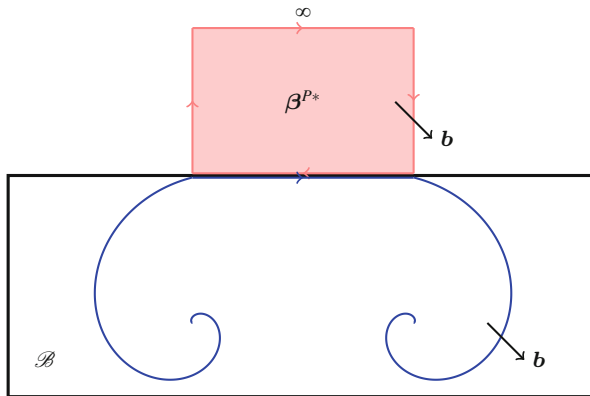
Fig. 2 Construction used for the solution of the BVP (6). For each dislocation segment on the domain boundary, a loop with the same Burgers vector and opposite line direction is constructed outside the integration domain and encloses the virtual slip surface β^{P*}. One portion of the external loop coincides with the boundary segment, and the opposite side is pushed to infinity in the direction of the boundary normal

added outside the domain. This fact can be used to cancel the stress field of internal segments deposited on the boundary, at the cost of virtual segments extending away from the boundary, as sketched in Fig. 2.

2.1.3   Anisotropic Elasticity

The majority of DD implementations use the isotropic elasticity approximation, due to the considerable technical simplification it affords. Real materials, however, are generally anisotropic to some extent, and for some applications, this cannot be ignored. As well as leading to quantitative differences in mechanical response, large elastic anisotropy allows certain unique phenomena to emerge, in particular dislocation instabilities, leading to sharp corners on the dislocation network, and enhance dislocation multiplication due to weakened Frank-Read sources. For a comprehensive introduction to anisotropic elasticity, the reader is referred to Balluffi (2016) and Bacon et al. (1980), and for its implementation in DD simulations, see Bacon et al. (1980), Han et al. (2003), Rhee et al. (2001), Fitzgerald and Aubry (2010), Yin et al. (2012), and Aubry and Arsenlis (2013).

One of the most notable examples where elastic anisotropy effects are most important in the context of nuclear materials is the behavior of Fe alloys at high temperature. In prismatic loops generated under irradiation, the lowest-energy shape does not necessarily correspond to a circular loop. Figure 3 (top) shows the anisotropic elastic line tension, Γ, vs. orientation for two situations of interest: $\langle 111 \rangle$ loops at 400°C and $\langle 100 \rangle$ loops at 500°C in iron (Dudarev et al. 2008). Although far from the constant value isotropy predicts, Γ remains positive for the former case, and no sharp corners appear. A barely discernable flattening of the circle toward a hexagon is present. For the second case, Γ dips below zero, and

Fig. 3 Top: line tension vs. orientation for ⟨111⟩ loops at 400 °C and ⟨100⟩ loops at 500 °C. Bottom: TEM micrographs confirming the predicted loop shapes
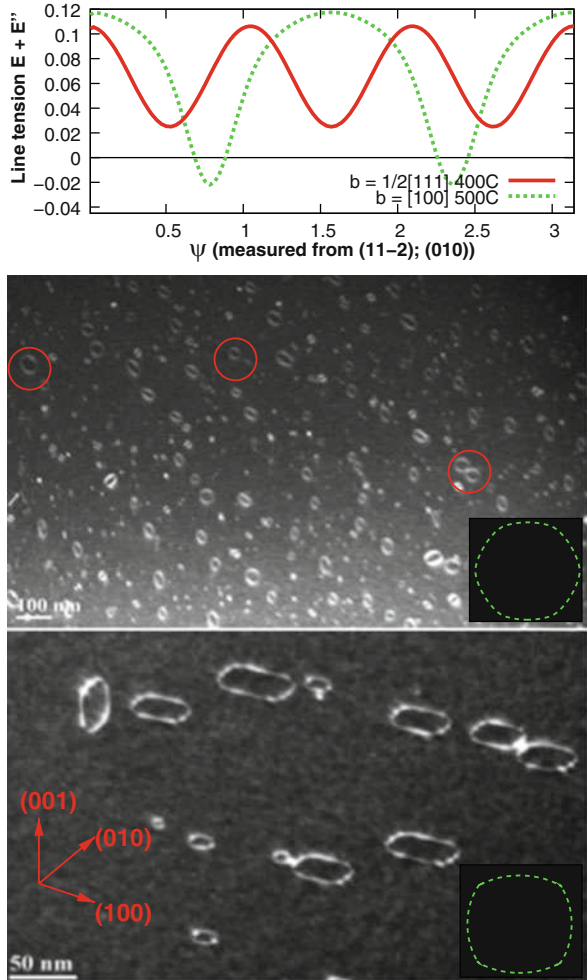
corners emerge. TEM micrographs confirm the prediction. The anisotropic elastic calculation reproduces the hexagonal and square symmetries of the ⟨111⟩ and ⟨100⟩ orientations, respectively.

A special case worth mentioning is that of hexagonal close-packed (HCP) metals, where the absence of cubic symmetry calls for the consideration of elastic anisotropy as an intrinsic lattice feature. A number of DD implementations that capture this aspect of HCP plasticity have been proposed (Monnet et al. 2004; Drouet et al. 2014; Aubry et al. 2016), although the application of DD to irradiated hexagonal metals is much more scant (cf. Sect. 4.1.3).

The effects of elastic anisotropy can be striking and are non-negligible in many metals of interest to radiation materials science. In particular, α-Fe (and also ferritic steels) become highly anisotropic at high temperature. This leads to different types

of prismatic loops appearing and sharp corners emerging on dislocation networks. This in turn significantly reduces the stress required to operate certain orientations of Frank-Read sources, enabling the onset of plasticity at much lower applied stresses than the isotropic approximation would suggest (Fitzgerald 2010; Fitzgerald et al. 2012).

2.2   Direct Consideration of Dislocation Partials for DD Simulations of FCC Metals

The dissociation of perfect dislocations into partials is a principal feature of close-packed lattices such as FCC and HCP structures. The importance of this dissociation is typically predicated upon a single parameter, the stacking fault energy, which is known to control critical features of plastic deformation in FCC metals such as the stability of dislocation junctions (and, by association, stage II hardening), cross-slip (and, by association, recovery), etc. Dislocation partials also enable the formation of structures of particular importance in irradiation scenarios such as stacking fault tetrahedra (SFT), which form from the collapse of vacancy platelets condensed during the cooling of the thermal spike of displacement cascades. These SFT are known to be sources of considerable hardening in irradiated FCC metals and therefore attracted significant amounts of research from both experimental and simulation points of view. However, while experimental advances have provided unique insights into the interaction mechanisms between network dislocations and SFT, a detailed step-by-step sequence of dislocation reactions has remained elusive, and the venerable "Kimura-Maddin" mechanism (Kimura et al. 1963) has remained the accepted interaction mechanism up to the development of reliable simulation techniques. One key feature of SFT is that they are dislocation structures themselves and thus amenable to study by MD and DD simulations. However, the explicit consideration of dislocation partials in DD simulations of FCC crystals was not considered a priority in the DD community up until very recently, and before that, the interaction between dislocations and SFT was only captured in an effective manner, informed by MD simulation results.

The first comprehensive effort to include dislocation partials in DD simulations of FCC crystals was done by Shenoy et al. (2000). However, a comprehensive model of dissociated dislocation dynamics and its application to irradiated materials was subsequently developed by Martinez, Arsenlis, and Marian in a series of papers where the isotropic elasticity-based DD technique was effectively generalized to any crystal lattice (Martinez et al. 2008a, b; Marian et al. 2009). The splitting of perfect dislocations into partials at physical nodes is achieved by considering the following additional vector conservation rule:

$$\sum_i \vec{n}_i = 0 \tag{7}$$

where $\vec{n}_i$ is a unit vector representing planes where dislocations partials are glissile, e.g., for Shockley partials, $\vec{n}_{\mathrm{Sh}} \equiv \langle 1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3} \rangle$. Note that for certain types of partials, $\vec{n}$ represents planes that are not lattice planes, and therefore the associated dislocation is sessile (e.g., a *stair rod* dislocation). As well, the model includes an additional force acting on partial dislocation segments:

$$\vec{f}_i = \vec{\xi}_i \times \gamma \vec{n}_i \tag{8}$$

where γ is the stacking fault energy. Choosing the correct sense for the $\vec{n}$ vectors ensures that this force is always attractive between partials. The inclusion of the above two expressions into the standard elastic DD formulation has been shown to suffice to predict the most important dislocation partials in the FCC lattice (Martinez et al. 2008a). By way of example, Fig. 4a shows the unfaulting of a triangular Frank loop into a stacking fault tetrahedron by the decomposition of Frank dislocation segments into Shockley and stair rod partials, such as the following reaction (in Thompson notation ($\mathrm{A}\alpha : \frac{1}{3}\left[\bar{1}1\bar{1}\right]$, $\mathrm{A}\delta : \frac{1}{6}\left[\bar{1}2\bar{1}\right]$, $\delta\alpha : \frac{1}{6}\left[\bar{1}0\bar{1}\right]$.)):

$$\mathrm{A}\alpha \rightarrow \mathrm{A}\delta + \delta\alpha$$

where it has been assumed that the dissociation takes place on the (111) (d) plane. Equivalent reactions can be written for the other planes.
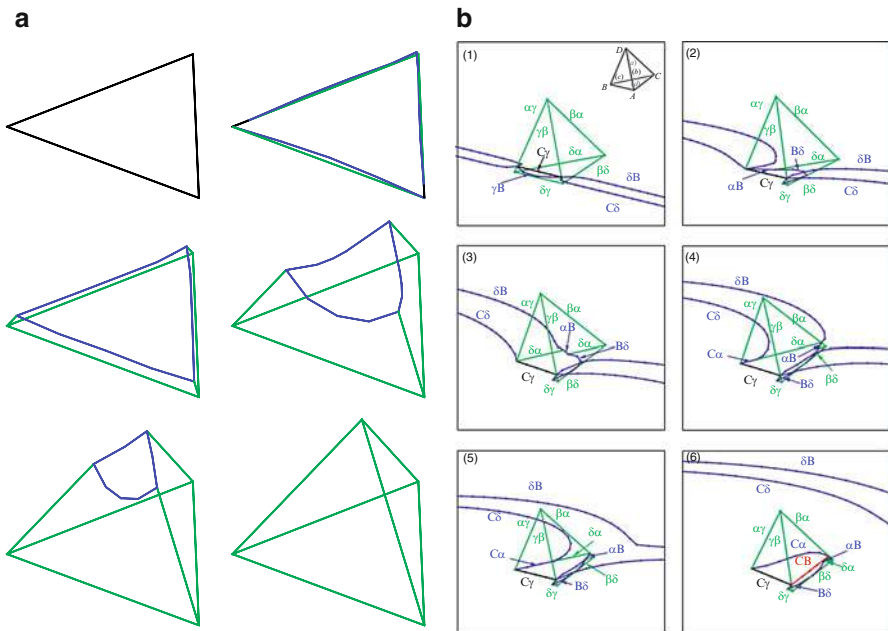
Fig. 4 (**a**) Unfaulting of a triangular Frank loop into a stacking fault tetrahedron. (From Martinez et al. 2008a). (**b**) Sequence of simulation snapshots of the interaction between a 60° dislocation with a 4.7-nm SFT. (From Martinez et al. 2008b)

The technique allows for detailed studies of dislocation-SFT interactions, an example of which is shown in Fig. 4b as a sequence of simulation snapshots of the interaction between a 60° mixed dislocation with a 4.7-nm SFT (from Martinez et al. 2008b). The quantitative implications of these studies will be discussed in Sect. 3.2.

3 Dislocation Interactions with Irradiation Obstacles

In general, irradiation-induced defect clusters constitute obstacles for gliding dislocations, but the fate and strength of the interaction depend on specific atomic mechanisms, which are best revealed by MD simulations. These mechanisms have been reviewed and classified by Bacon and Osetsky (2005) and Bacon et al. (2009). However, sometimes MD cannot capture the details of the interaction mechanisms, particularly those that emerge at low strain rates and quasistatic conditions or those involving many-body effects. In such cases, DD itself can provide information about the process, and here we review some of the most important examples.

3.1 Prismatic Loops

Despite their complexity, recent studies have shown that the reaction pathway between gliding dislocations and prismatic loops is well represented within DD simulations, hence indicating that these reactions are governed by elastic interactions. For example, Shi et al. (2015) have considered all possible reactions between a $1/2\langle 111\rangle\{110\}$ edge dislocation and a square $\langle 100\rangle\{100\}$ interstitial loop in α-Fe and compared their result to corresponding MD simulations carried out by Terentyev et al. (2008). It was found that DD simulations are able to reproduce the reaction pathways observed in the atomistic simulations. Examples of such reaction pathways are shown in Fig. 5. Similar conclusions were reached by Cui et al. (2017), who considered DD simulations of $1/2\langle 111\rangle\{110\}$ edge and screw dislocations interacting with $1/2\langle 111\rangle\{111\}$ prismatic loops in comparison to the MD results of Liu and Biner (2008).

DD simulations of bulk α-Fe deformation (Arsenlis et al. 2012), confined grain deformation (Gururaj et al. 2015b), and micropillar compression (Cui et al. 2017) have considered different populations of the two types of prismatic dislocation loops typically observed in irradiated Fe. In these simulations, prismatic loops are resolved individually, with typical loop densities in the range $10^{20} \sim 10^{22}$ m^{-3}. These models do not require any particular assumptions, and therefore they are well-suited to predict the collective behavior of dislocation-obstacle interactions at the microscale. However, the discrete representation of loops limits their scope of applicability because of the high computational cost associated with the increased number of elastic segments. Higher defect densities and larger simulation time and spatial scales can be achieved if the discrete representation is replaced by a phenomenological description of obstacle-dislocation interaction, with parameters determined
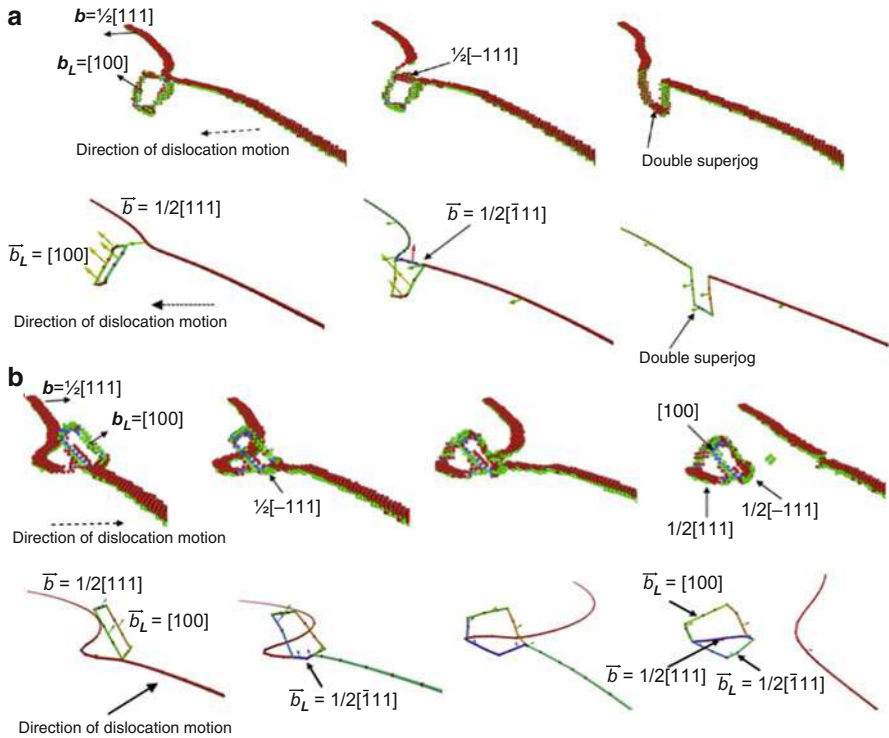
Fig. 5 Examples of corresponding reaction pathways in DD and MD simulations between a $1/2\langle111\rangle\{110\}$ edge dislocations and a square $\langle100\rangle\{100\}$ interstitial loop in α-iron (Shi et al. 2015). (**a**) Prismatic loop incorporated into the mobile edge dislocation in the form of a double super-jog. (**b**) Pinning-reaction-unpinning of the glissile dislocation, with transformation of the prismatic loop

from lower-scale simulations. For example, Ghoniem et al. (2001) performed early DD simulations to capture the strength of the interaction between dislocations and arrays of dipolar loops. Terentyev et al. (2013) have used simple DD calculations to extract the parameters necessary to describe $1/2\langle111\rangle$ dislocation loops in Fe as stochastic thermally activated obstacles to be used in larger-scale simulations. Along the same lines, Cui et al. (2018) have extracted the parameters that govern the SIA loop density evolution in Fe and have turned the dislocation-defect interaction into a continuum field problem coupled to DD simulations.

3.2   Stacking Fault Tetrahedra

The implementation of the model described in Sect. 2.2 has enabled the study of dislocation-SFT interactions as the building block for irradiation hardening in metals with moderate-to-low stacking fault energies. While several of these
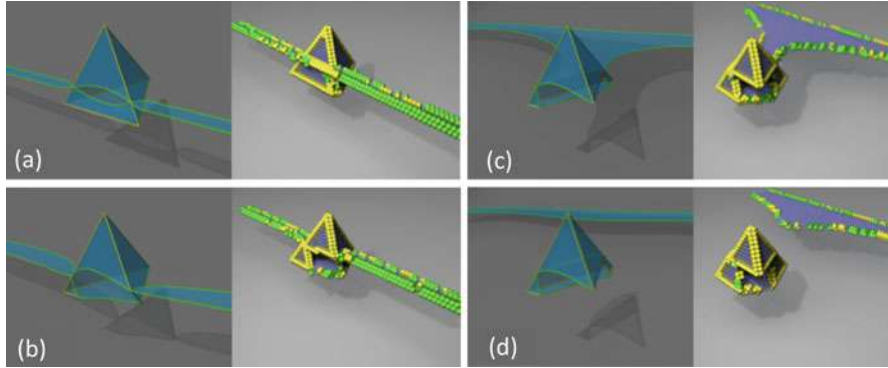
Fig. 6 Examples of DD simulations accounting for partial dislocations explicitly. Comparison of quasistatic atomistic simulations and DD simulations described in Sect. 2.2. Atomistic simulations from Marian et al. (2009)

simulations have been verified by dedicated MD simulations (Marian et al. 2009), as shown in Fig. 6, DD allows for a more efficient exploration of the relevant parametric space, to which – in addition to dislocation character, line length, temperature, stress, and obstacle size – one must add the surface of the area intersected by the glide plane on the SFT. As well, using the FCC/DD model gives access to strain rates not captured by MD simulations, thus providing an extra window into the dynamics of irradiation hardening in FCC metals. One of the most interesting physical insights gained from FCC/DD simulations was that the strength of SFT does not depend on their size but only on the intersected triangular area, such that equal areas in two tetrahedra of different sizes would give rise to the same response (Marian et al. 2009). This is also discussed in Sect. 4.1.1.

Unfortunately, the timesteps achievable by explicit integration of the equations of motion in the FCC/DD system drop to values that preclude the direct use of the technique to simulate strain hardening in a statistically representative manner. Until these issues are resolved, the explicit consideration of partials is likely to be limited to the study of unit mechanisms.

3.3   Voids and Precipitates

The mechanisms of interaction between dislocations and voids or precipitates have been explored by MD simulations by Osetsky and Bacon (2003) and Osetsky and Bacon (2010). For α-Fe, it was found that small coherent precipitates ($d < 2$ nm) can be sheared by moving dislocations, while larger precipitates and voids are overcome by more complex mechanisms, which may include dislocation climb, jog formation, and phase change of the precipitates. Although dislocations do not leave Orowan loops behind voids and sufficiently large coherent precipitates, the strengthening effects of these strong obstacles are compatible with the

Bacon-Kocks-Scattergood (BKS) strengthening model (see Sobie et al. (2015) for a comparison between strengthening models) (Bacon et al. 1973). This model predicts an obstacle resistance:

$$\tau_c = \frac{\mu b}{2\pi L} \left[\ln \left(D^{-1} + L^{-1} \right)^{-1} + B \right], \tag{9}$$

where D and L are the obstacle spacing and diameter, respectively, and B is an obstacle-specific fitting constant. Figure 7 shows the agreement of Eq. (9) in the case of voids and coherent Cu precipitates in α-Fe. Small precipitates offer a weaker resistance because obstacle shearing allows for large breakaway angles, which are inconsistent with Eq. (9). The situation is somewhat more complicated in FCC metals, because partials interact individually with voids when their spacing is larger than the obstacle size. In addition, the dissociated core prevents climb over voids by vacancy absorption, ultimately resulting in a larger deviation of the obstacle strength from Eq. 9 (Bacon and Osetsky 2005; Lehtinen et al. 2018). Moreover, obstacle strength decreases with temperature, because local energy fluctuations assist dislocation unpinning from obstacles (Monnet et al. 2010).

DD simulations carried out by Haghighat et al. (2009) have shown that elastic calculations predict a behavior of dislocations in the proximity of a void which is consistent with MD results when the image forces due to the void are taken into account. In large-scale calculations, however, accurate image force calculations cannot be performed for every void, and atomistically informed phenomenological interaction laws are once again more feasible. For example, Queyreau et al. (2011) considered the temperature dependence of void strengthening in DD simulations with a void density of $2 \times 10^{23} \mathrm{m}^{-3}$ in α-iron. Voids were assumed to be pinning points within the simulation domain, and they were released based on a probability following an Arrhenius-type law with activation energy dependent on the local stress and parameters extracted from MD simulations.

Also in the case of precipitates, accurate elastic calculations which account for image forces were performed to study the parameters governing the strength of both impenetrable (Shin et al. 2003) and penetrable (Takahashi and Ghoniem 2008) particles. A variety of techniques to account for impenetrable particles in larger-scale DD simulations have been proposed (Mohles 2001, 2004; Monnet 2006; Monnet et al. 2011; Ringdalen et al. 2017). Queyreau et al. (2010) have analyzed the effects of combined forest hardening and Orowan strengthening due to non-coherent carbides in reactor pressure vessel (RPV) steels. In this simulation, spherical carbide precipitates with average density of 3×10^{19} m^{-3} and mean precipitate diameter of $0.1\,\mu\mathrm{m}$ are regarded as volumes impenetrable to dislocations, and they can only be overcome by the formation of Orowan loops (Fig. 7b). It was found that although precipitate hardening is well described by the BKS model, when precipitate hardening is combined to forest hardening, the overall strengthening effect is best described by a quadratic rule of mixing between the two effects.
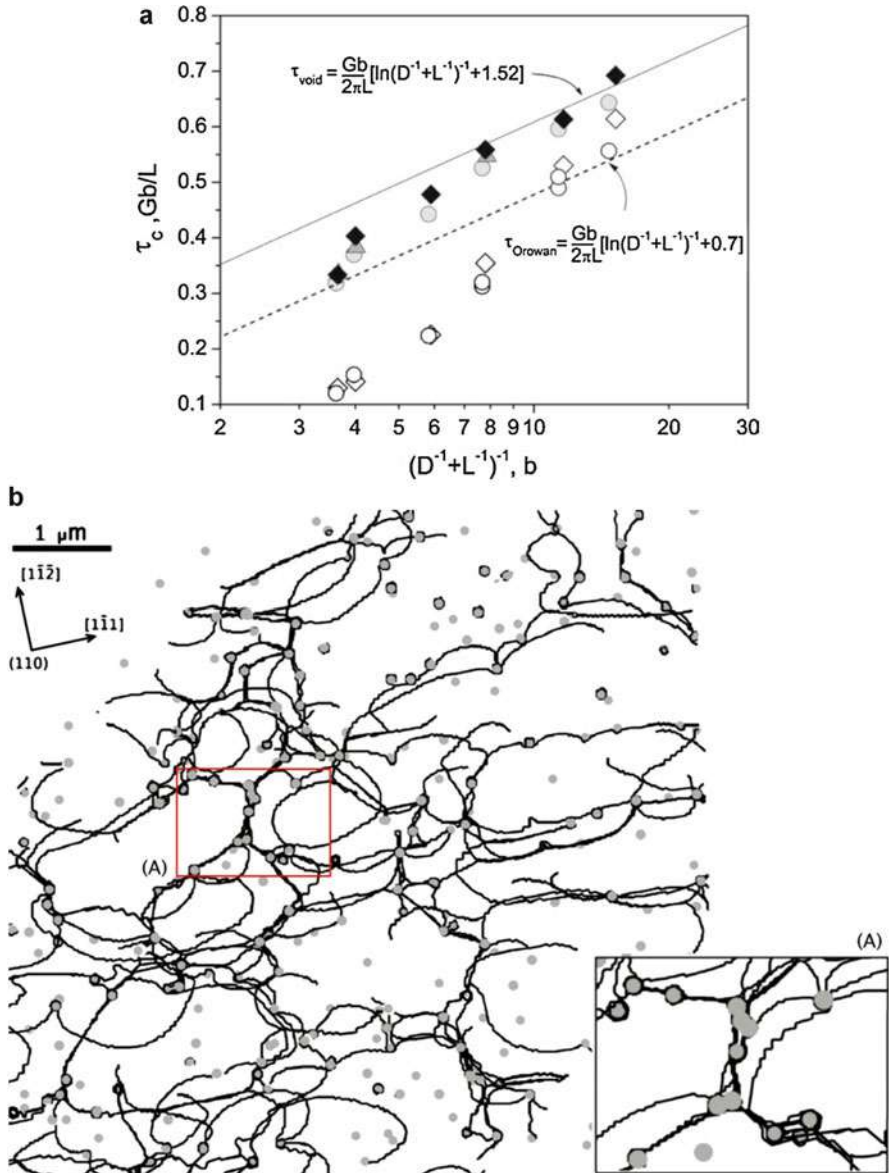
Fig. 7 (**a**) Strengthening effects of voids (full symbols) and coherent Cu precipitates (hollow symbols) in α-Fe as a function of the harmonic average between their diameter D and spacing L (Osetsky and Bacon 2003). (**b**) Snapshot of DD simulations of combined forest hardening and Orowan strengthening due to non-coherent carbides in RPV steels (Queyreau et al. 2010)

4   Applications

4.1   Homogeneous Irradiation Hardening

Here we discuss simulations in FCC Cu and BCC Fe as representative of each of their own material classes in terms of irradiation hardening.

4.1.1   Simulations in Irradiated Cu

One of the main advantages of DD simulations over atomistic simulations is the in principle much higher computational efficiency of the former, which allows for rapid exploration of the parametric space of the dislocation-defect interaction process. Variables such as defect size, dislocation length, dislocation character, defect orientation, stress, temperature, etc. create a many-dimensional landscape that is very difficult to probe using just atomistic simulations. With dislocation dynamics, one can explore these dependencies more extensively and enable a deeper understanding of the key variables affecting the process. By way of example, Fig. 8 shows the strength of SFT in Cu as a function of several key parameters as obtained with DD simulations methods such as that described in Sect. 2.2 (Marian et al. 2009).

Moreover, higher-scale models such as parametric or stochastic dislocation dynamics or crystal plasticity can be informed with multidimensional correlations, further reducing the computational overhead and reach larger scales. For example, one of the things that we have learned from DD simulations in Cu is that the strength of SFT does not directly depend on SFT size but on the area intersected by the dislocation glide plane and the defect, such that equal areas on SFT of different sizes give rise to the same response (Marian et al. 2009). This has enabled the development of universal laws that can be used in the coarser models, thus
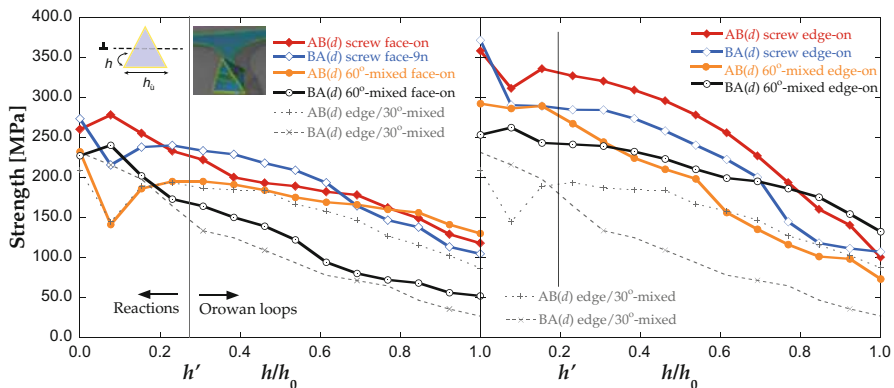
Fig. 8 Strength of stacking fault tetrahedra in Cu to dislocation glide as a function of the intersection point between the defect and the glide plane, dislocation character, and SFT orientation

eliminating the need to simulate individual dislocation-SFT interactions directly and quantifying bulk hardening on experimental scales (Arsenlis et al. 2004; Robertson et al. 2005; Barton et al. 2013).

4.1.2   Simulations in Fe

DD simulations of bulk hardening in Fe can be used to investigate the dichotomy existing in the irradiation hardening community which distinguishes between dispersed-barrier hardening and dislocation decoration (cf. Sect. 1). This has motivated significant work recently (Gururaj and Robertson 2011; Arsenlis et al. 2012; Gururaj et al. 2015b), from which some interesting conclusions have emerged. First, the hardening levels calculated using various defect densities are consistent with a DBH approach, whereby $\Delta \tau = \alpha \mu b \sqrt{d \rho_d}$, where d and ρ_d are the defect average size and number density, respectively (Arsenlis et al. 2012; Gururaj et al. 2015b). From these simulations, it is clear that the source hardening (dislocation decoration) model results in more elevated hardening levels than the DBH model for a given irradiation dose and, consequently, can trigger flow-localized behavior earlier (Gururaj et al. 2015b) (this is the object of the next subsection). Simulations including a combination of oxide-dispersed particles and irradiation defects have also revealed that particles can actually delay flow localization by holding the uniformity of the dislocation substructure against spatial fluctuations in the strain brought about by fine-scale irradiation defect distributions (Gururaj and Robertson 2011).

   As well, the well-known phenomenon of loop coarsening and "raft" formation can be simulated by enabling climb processes for dislocation segments (Keralavarma et al. 2012; Bakó et al. 2011).

4.1.3   Simulations in Zr

Although the volume of work is considerably smaller, there have been DD studies of irradiated Zr, which is the basis of Zircaloys used in nuclear fuel cladding. Drouet et al. (2014) have studied several configurations involving dislocations and irradiation loops as a function of loop geometry, glide plane orientation, and dislocation character. Their simulations have shown various interaction mechanisms between both objects including formation of double super-jogs, formation of helical turns, or sweeping of a loop by a gliding dislocation. They have also shown that the clearing of loops is more favorable when the dislocation glides in the basal plane than in the prismatic plane, consistent with observations of dislocation channeling in the basal plane observed after neutron irradiation by transmission electron microscopy.

4.2   Strain Localization

Strain localization is one of the main detrimental effects associated with high-dose irradiation in nuclear materials, as it has been extensively documented in the literature (Victoria et al. 2000; Hashimoto et al. 2004, 2006; Yao et al. 2004; Byun

and Farrell 2004; Byun et al. 2004, 2006b, a; Zinkle and Singh 2006; McMurtrey et al. 2014). Above a critical defect density, plastic flow in many materials is seen to occur along "defect-free" dislocation channels that result in a highly spatially heterogeneous deformation usually accompanied by a severe decrease in total elongation. These channels are seen to typically range between 150 and 250 nm in width and are generally oriented along {111} and {100} crystallographic orientations in FCC metals and {112} and {110} in BCC (Hashimoto et al. 2006; Zinkle and Singh 2006). Modeling this particular type of deformation presents many challenges as several phenomena taking place at sharply different time and length scales must be considered, including channel initiation, channel growth and propagation, etc.

The first efforts at modeling dislocation channeling using dislocation dynamics were conducted for irradiated Cu, assuming SFT removal mechanisms observed in MD simulations (Diaz de la Rubia et al. 2000; Sun et al. 2000; Khraishi et al. 2002). These simulations were effective in demonstrating that channel initiation results as a consequence of spatial fluctuations that trigger an instability characterized by a post-yield softening. However, the removal of objects by dislocations was based on effective contact rules based on MD simulations performed under strikingly different conditions than those of DD (Crosby et al. 2014). This introduces an incompatibility that questions the physical fidelity of the simulations. These physical models have eventually been extended in more recent DD works that treat dislocation-SFT interactions as true dislocation reactions (Martinez et al. 2008b, c) (cf. Sect. 2.2).

The other type of primary irradiation defect considered is dislocation loops. Being dislocation structures themselves, loops can be naturally captured within the linear elastic framework of DD, which allows for a direct treatment of dislocation-loop interactions (Nogaret et al. 2008; Arsenlis et al. 2012; Gururaj et al. 2015a). While the global mechanism that leads to plastic flow localization is fundamentally the same – i.e., local removal of irradiation defects either by absorption into the dislocation line or elastic repulsion – this type of simulations allows for a more detailed investigation of the channel nucleation and propagation mechanism. For example, for the case of 25-nm dislocation loops in Fe, it has been seen that the channel size is governed by the formation of a local curved dislocation source in the vicinity of a defect-depleted region grown by the absorption of and coalescence with irradiation defects (Arsenlis et al. 2012). This source can be activated at stresses significantly lower than the yield stress set by a homogeneous distribution of defects, resulting in softening and localized deformation in the vicinity of the source. The system behaves in an unstable manner in the sense that, as the source propagates throughout the simulation box, it triggers the absorption of more defects, which in turn makes the source segment grow and the required activation stress decrease. Some other works suggest that it is the level of defect hardening – rather than the defect density per se – that determines the onset of subsequent plastic localization (Gururaj et al. 2015b). In any case, channel initiation is a *rare* event, and large computational cells must be deformed to plastic strains above approximately 1% for channels to be incubated. This often requires massively parallel simulations and data analysis. Figure 9 shows the dislocation microstructure in an iron single crystal

Fig. 9 Dislocation microstructure in an iron single crystal after deformation with a defect density of 1.6×10^{22} m^{-3} loops, clearly showing the existence of a defect-free dislocation channel. (From Arsenlis et al. 2012)
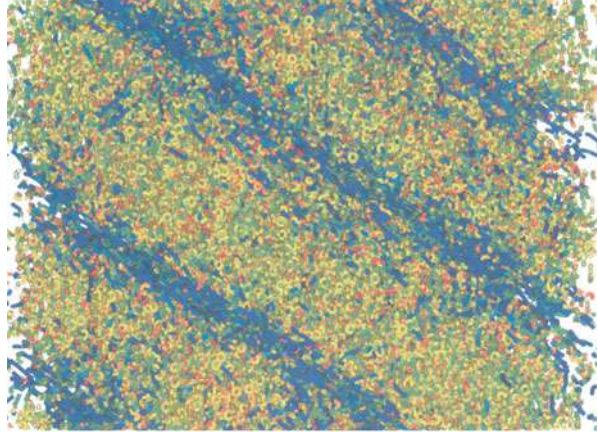
populated with a density of 1.6×10^{22} m^{-3} dislocation loops. The width of the channel is approximately 150 nm, and its orientation is along the [211] direction (Arsenlis et al. 2012).

While much work remains to be done to improve the meaningfulness of DD simulations to continue studying plastic flow localization in irradiated materials, existing models have been capable of offering useful information of the process such as the observation that channels can be and are nucleated in bulk grains, not just at grain boundaries (McMurtrey et al. 2014). This can be important to understand important phenomena such as irradiation-assisted stress corrosion cracking initiation.

4.3 Irradiated Thin Films

The foils used in transmission electron microscopy for irradiation studies may be as thin as 40 nm, and hence dislocations and other defects within it can be strongly influenced by elastic image forces due to the free surfaces. These fictitious image forces arise due to the zero traction boundary condition ($\boldsymbol{\sigma} \cdot \hat{\boldsymbol{n}} = 0$) imposed at the free surface. While such a scenario is perfectly well-captured within the method for finite geometries described in Sect. 2.1.2, the high symmetry of the thin film geometry allows for a more direct approach that results in significant computational gains over the finite element model (Weinberger et al. 2009). The equilibrium problem for the displacement vector on a flat infinite surface can be solved analytically in Fourier space (Weinberger et al. 2009; Ferroni et al. 2014), giving a solution that decays exponentially along the direction normal to the free surface. The resulting image stresses can then be straightforwardly calculated and added to DD nodal force contributions via the Peach-Koehler formula like any other stress. The exploitation of the analytical Fourier solutions yields a significant

computational cost saving over the general FEM approach but can only be used for domains with sufficiently high symmetry that such an analytical solution exists.

Complications arise when dislocation segments reach the free surfaces, an issue that can be solved by extending terminating segments out of the crystal to infinity by defining *virtual segments* (Weinberger et al. 2009). The stress fields of the virtual segments must be included when computing nodal forces on the real segments inside the crystal, but they themselves do not move dynamically and are simply redefined at each timestep (Tang et al. 2006; Yoffe 1961). Ferroni et al. (2014) used this approach to perform DD simulations of irradiated thin films in Fe. The simulations concluded that radiation damage loops interact strongly with the surface when they approach to within $3\times$ the linear size of the loop. For situations where climb is negligible, e.g., low temperatures, free surfaces play a dominant role in the evolution of thin film dislocation structures. When climb was present however, many small prismatic loops had self-annealed or coalesced into larger loops before reaching the surface.

4.4   Irradiated Micropillars

Micro-mechanical testing is an emerging area of investigation in the field for irradiated materials, which prospects the benefit of a reduced amount of radioactive material that needs to be handled during experiments (Hosemann 2018). *In situ* mechanical tests such as nanoindentation, micro tension/compression, and micro bending also have the ability to directly observe isolated changes in the defect microstructure in relation to the applied load. Experimental micro-mechanics is therefore expected to provide new insights on the mechanics of plastic deformation and failure of irradiated materials. Moreover, because these experiments are typically conducted at length scales amenable by DD simulations, they also provide specialized data useful to advance materials modeling.

Cui et al. (2017) considered the problem of intermittent plastic flow in irradiated micropillars and performed a comprehensive DD study of the phenomenon as a function of material type (BCC Fe vs. FCC Cu), loading mode (strain vs. stress control), and irradiation dose. In these simulations, irradiation-induced defects were represented explicitly as dislocation debris within the simulation volume, namely, SFT for Cu and interstitial dislocation loops for Fe. The elastic BVP correction method described in Sect. 2.1.2 was applied to obtain uniaxial compression in both strain and stress control modes. It was found that, in general, defects hinder dislocation motion because they decrease the dislocation mean free path by participating in junctions. In stress control mode, where no stress relaxation takes place during the burst, the burst's size tends to decrease with dose. In strain control, however, the burst size is controlled by the difference between the burst stress, which increases with dose, and the deactivation stress of the active source. Therefore irradiation may result in larger dislocation bursts under strain control conditions, as shown in Fig. 10.
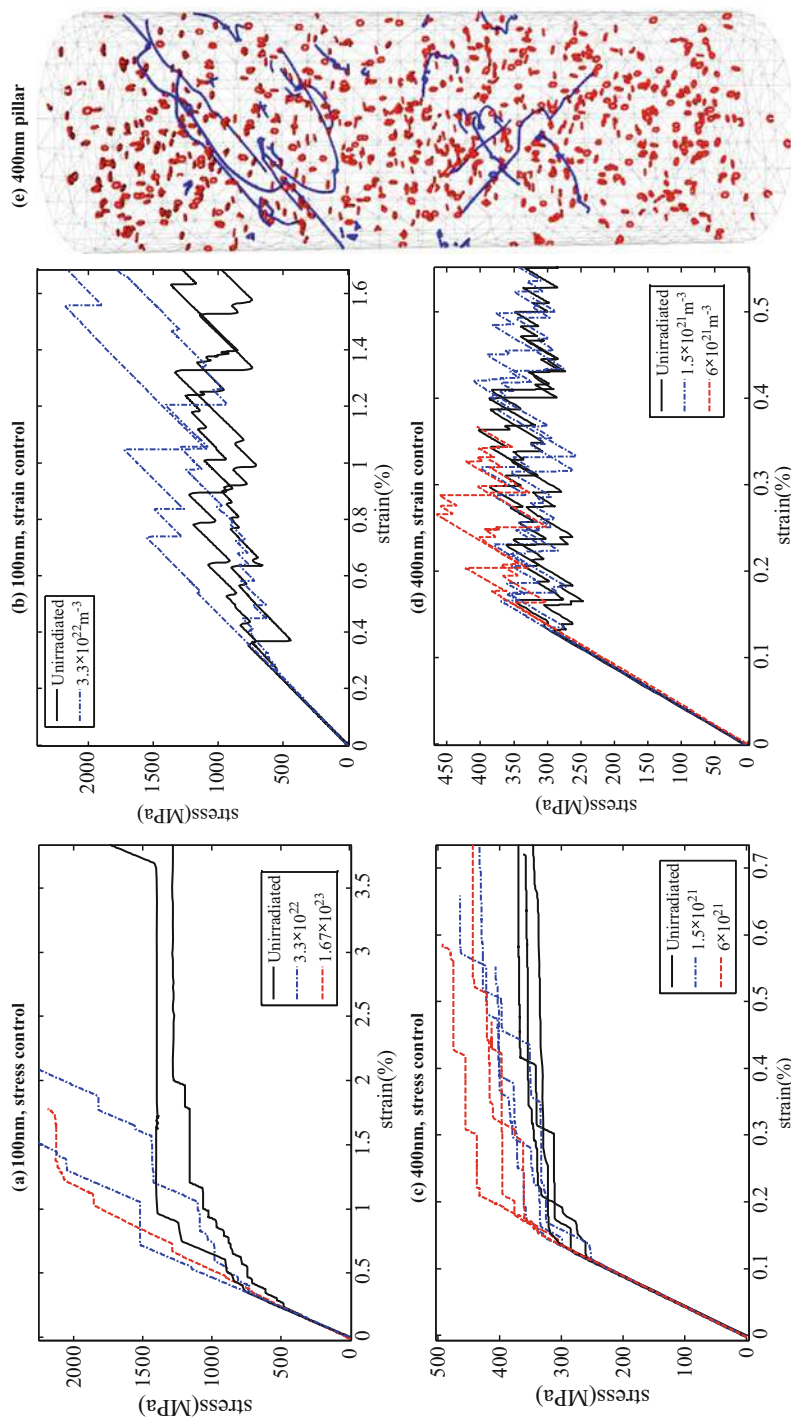
Fig. 10 Strain bursts in DD simulations of irradiated Fe micropillars (Cui et al. 2017)

5   Summary and Future Directions

This chapter attempts to present an overview of the most salient developments concerning the use and application of discrete dislocation dynamics for irradiated materials. While the progress achieved over the last two decades has been remarkable, there are still several aspects of irradiation hardening lacking to be studied. By way of example, we give here a non-exhaustive list of issues needing to be resolved:

- Pushing the time-/length-scale horizon to achieve usable strains $>5\%$ and realistic strain rates ($<10^{-1}$ per second).
- Simulations considering complex precipitate geometries (needle shaped, elongated, faceted, etc.), as well as in nano-layered materials, have not yet been the focus of substantial work Ringdalen et al. (2017).
- Implementation of realistic models of irradiation-assisted dislocation climb.
- Except for some noted efforts Geantil et al. (2013), dislocation dynamics has been generally unsuccessful in predicting the formation of dislocation substructures at high deformations. More work in this area would be a welcome addition.
- Simulations of shock deformation using DD must be done including time-dependent displacement fields (in 3D in addition to 2D (Gurrutxaga-Lerma et al. 2013)) and accounting for inertial effects on dislocation motion.

Beyond these and other aspects meriting more study, we believe that DD has been extremely successful in providing critical insights into fundamental defect-dislocation processes with high relevance in nuclear materials.

References

Ackland G (2010) Controlling radiation damage. Science 327(5973):1587–1588

Amodeo RJ, Ghoniem NM (1990) Dislocation dynamics. I. A proposed methodology for deformation micromechanics. Phys Rev B 41:6958–6967. https://doi.org/10.1103/PhysRevB.41.6958

Arsenlis A, Wirth B, Rhee M (2004) Dislocation density-based constitutive model for the mechanical behaviour of irradiated cu. Philos Mag 84(34):3617–3635

Arsenlis A, Cai W, Tang M, Rhee M, Oppelstrup T, Hommes G, Pierce TG, Bulatov VV (2007) Enabling strain hardening simulations with dislocation dynamics. Model Simul Mater Sci Eng 15(6):553

Arsenlis A, Rhee M, Hommes G, Cook R, Marian J (2012) A dislocation dynamics study of the transition from homogeneous to heterogeneous deformation in irradiated body-centered cubic iron. Acta Mater 60(9):3748–3757

Aubry S, Arsenlis A (2013) Use of spherical harmonics for dislocation dynamics in anisotropic elastic media. Model Simul Mater Sci Eng 21(6):065013

Aubry S, Rhee M, Hommes G, Bulatov V, Arsenlis A (2016) Dislocation dynamics in hexagonal close-packed crystals. J Mech Phys Solids 94:105–126

Bacon DJ, Osetsky YN (2005) Modelling dislocation–obstacle interactions in metals exposed to an irradiation environment. Mater Sci Eng A 400:353–361

Bacon DJ, Kocks UF, Scattergood RO (1973) The effect of dislocation self-interaction on the Orowan stress. Philos Mag 28(6):1241–1263

Bacon D, Barnett D, Scattergood RO (1980) Anisotropic continuum theory of lattice defects. Prog Mater Sci 23:51–262

Bacon DJ, Osetsky YN, Rong Z (2006) Computer simulation of reactions between an edge dislocation and glissile self-interstitial clusters in iron. Philos Mag 86(25–26):3921–3936

Bacon D, Osetsky Y, Rodney D (2009) Dislocation–obstacle interactions at the atomic level. Dislocat Solids 15:1–90

Bakó B, Clouet E, Dupuy LM, Blétry M (2011) Dislocation dynamics simulations with climb: kinetics of dislocation loop coarsening controlled by bulk diffusion. Philos Mag 91(23): 3173–3191

Balluffi RW (2016) Introduction to elasticity theory for crystal defects. World Scientific Publishing Company, Singapore

Barton NR, Arsenlis A, Marian J (2013) A polycrystal plasticity model of strain localization in irradiated iron. J Mech Phys Solids 61(2):341–351

Blewitt T, Coltman R, Jamison R, Redman J (1960) Radiation hardening of copper single crystals. J Nucl Mater 2(4):277–298

Bulatov VV, Rhee M, Cai W (2000) Periodic boundary conditions for dislocation dynamics simulations in three dimensions. MRS Proceedings, 653, Z1.3. https://doi.org/10.1557/PROC-653-Z1.3

Byun T, Farrell K (2004) Plastic instability in polycrystalline metals after low temperature irradiation. Acta Mater 52(6):1597–1608

Byun T, Hashimoto N, Farrell K (2004) Temperature dependence of strain hardening and plastic instability behaviors in austenitic stainless steels. Acta Mater 52(13):3889–3899

Byun T, Hashimoto N, Farrell K (2006a) Deformation mode map of irradiated 316 stainless steel in true stress–dose space. J Nucl Mater 351(1–3):303–315

Byun T, Hashimoto N, Farrell K, Lee E (2006b) Characteristics of microscopic strain localization in irradiated 316 stainless steels and pure vanadium. J Nucl Mater 349(3):251–264

Cai W, Arsenlis A, Weinberger CR, Bulatov VV (2006) A non-singular continuum theory of dislocations. J Mech Phys Solids 54(3):561–587

Calder A, Bacon D (1993) A molecular dynamics study of displacement cascades in α-iron. J Nucl Mater 207:25–45

Cottrell AH, Bilby B (1949) Dislocation theory of yielding and strain ageing of iron. Proc Phys Soc Sect A 62(1):49

Crosby T, Po G, Ghoniem NM (2014) Modeling plastic deformation of post-irradiated copper micro-pillars. J Nucl Mater 455(1–3):126–129

Cui Y, Po G, Ghoniem N (2017) Does irradiation enhance or inhibit strain bursts at the submicron scale? Acta Mater 132:285–297

Cui Y, Po G, Ghoniem NM (2018) A coupled dislocation dynamics-continuum barrier field model with application to irradiated materials. Int J Plast 104:54–67

Deng J, El-Azab A, Larson BC (2008) On the elastic boundary value problem of dislocations in bounded crystals. Philos Mag 88:3527–3548

Devincre B, Madec R, Monnet G, Queyreau S, Gatti R, Kubin L (2011) Modeling crystal plasticity with dislocation dynamics simulations: the micromegas code. In: Thomas O (ed) Mechanics of Nano-objects. Presses des Mines, Paris, pp 81–100

Diaz de La Rubia T, Averback R, Hsieh H, Benedek R (1989) Molecular dynamics simulation of displacement cascades in Cu and Ni: thermal spike behavior. J Mater Res 4(3): 579–586

Diaz de la Rubia T, Zbib HM, Khraishi TA, Wirth BD, Victoria M, Caturla MJ (2000) Multiscale modelling of plastic flow localization in irradiated materials. Nature 406(6798):871 EP. https://doi.org/10.1038/35022544

Drouet J, Dupuy L, Onimus F, Mompiou F, Perusin S, Ambard A (2014) Dislocation dynamics simulations of interactions between gliding dislocations and radiation induced prismatic loops in zirconium. J Nucl Mater 449(1–3):252–262

Dudarev SL, Bullough R, Derlet PM (2008) Effect of the $\alpha - \gamma$ phase transition on the stability of dislocation loops in bcc iron. Phys Rev Lett 100:135503

Eshelby J (1979) Boundary problems. In: Nabarro FRN (ed) Dislocations in solids V1. North-Holland, Amsterdam, pp 167–221

Ferroni F, Tarleton E, Fitzgerald S (2014) Dislocation dynamics modelling of radiation damage in thin films. Model Simul Mater Sci Eng 22(4):045009

Fitzgerald SP, Aubry S (2010) Self-force on dislocation segments in anisotropic crystals. J Phys Condens Matter 22(29):295403

Fitzgerald, SP (2010) Frank–Read sources and the yield of anisotropic cubic crystals. Philos Mag Lett 90:209–218

Fitzgerald SP, Aubry S, Dudarev SL, Cai W (2012) Dislocation dynamics simulation of Frank-Read sources in anisotropic α-Fe. Model Simul Mater Sci Eng 20:045022

Fivel M, Canova G (1999) Developing rigorous boundary conditions to simulations of discrete dislocation dynamics. Model Simul Mater Sci Eng 7(5):753

Geantil P, Devincre B, Kassner ME (2013) Dislocation-induced internal stresses. In: Altenbach H, Kruch S (eds) Advanced materials modelling for structures. Springer, Berlin/Heidelberg, pp 177–187

Ghoniem NM, Amodeo R (1988) Computer simulation of dislocation pattern formation. Solid State Phenomena, 3&4:377–388. https://www.scientific.net/SSP.3-4.377.pdf

Ghoniem NM, Tong SH, Sun L (2000) Parametric dislocation dynamics: a thermodynamics-based approach to investigations of mesoscopic plastic deformation. Phys Rev B 61(2):913

Ghoniem NM, Tong SH, Singh B, Sun L (2001) On dislocation interaction with radiation-induced defect clusters and plastic flow localization in FCC metals. Philos Mag A 81(11):2743–2764

Golubov S, Singh B, Trinkaus H (2000) Defect accumulation in FCC and BCC metals and alloys under cascade damage conditions–towards a generalisation of the production bias model. J Nucl Mater 276(1–3):78–89

Gulluoglu A, Srolovitz DJ, LeSar R, Lomdahl P (1989) Dislocation distributions in two dimensions. Scr Metall 23:1347–1352

Gurrutxaga-Lerma B, Balint DS, Dini D, Eakins DE, Sutton AP (2013) A dynamic discrete dislocation plasticity method for the simulation of plastic relaxation under shock loading. Proc R Soc A 469(2156):20130141

Gururaj K, Robertson C (2011) Plastic deformation in ods ferritic alloys: a 3D dislocation dynamics investigation. Energy Procedia 7:279–285

Gururaj K, Robertson C, Fivel M (2015a) Channel formation and multiplication in irradiated FCC metals: a 3D dislocation dynamics investigation. Philos Mag 95(12):1368–1389

Gururaj K, Robertson C, Fivel M (2015b) Post-irradiation plastic deformation in BCC Fe grains investigated by means of 3D dislocation dynamics simulations. J Nucl Mater 459:194–204

Haghighat SH, Fivel M, Fikar J, Schaeublin R (2009) Dislocation–void interaction in Fe: a comparison between molecular dynamics and dislocation dynamics. J Nucl Mater 386:102–105

Han X, Ghoniem NM, Wang Z (2003) Parametric dislocation dynamics of anisotropic crystals. Philos Mag 83(31–34):3705–3721

Hashimoto N, Byun TS, Farrell K, Zinkle SJ (2004) Deformation microstructure of neutron-irradiated pure polycrystalline metals. J Nucl Mater 329:947–952

Hashimoto N, Byun TS, Farrell K (2006) Microstructural analysis of deformation in neutron-irradiated FCC materials. J Nucl Mater 351(1–3):295–302

Holmes DK (1964) Radiation Damage in Non-Fissionable Metals. In: Strumane R (ed) The Interaction of Radiation with Solids. North-Holland, Amsterdam, 147

Hosemann P (2018) Small-scale mechanical testing on nuclear materials: bridging the experimental length-scale gap. Scr Mater 143:161–168

Keralavarma SM, Cagin T, Arsenlis A, Benzerga AA (2012) Power-law creep from discrete dislocation dynamics. Phys Rev Lett 109:265504. https://doi.org/10.1103/PhysRevLett.109.265504

Khraishi TA, Zbib HM, Diaz de La Rubia T, Victoria M (2002) Localized deformation and hardening in irradiated metals: three-dimensional discrete dislocation dynamics simulations. Metall Mater Trans B 33(2):285–296

Kimura H, Kuhlmann-Wilsdorf D, Maddin R (1963) The growth mechanism of stacking-fault tetrahedra in quenched gold. Appl Phys Lett 3(1):4–5

Kubin LP (2013) Dislocations, mesoscale simulations and plastic flow, vol 5. Oxford University Press, Oxford

Kubin LP, Canova G (1992) The modelling of dislocation patterns. Scr Metall Mater 27(8):957–962

Kubin LP, Canova G, Condat M, Devincre B, Pontikis V, Bréchet Y (1992) Dislocation microstructures and plastic flow: a 3D simulation. Solid State Phenom Trans Tech Publ 23:455–472

Lambrecht M, Meslin E, Malerba L, Hernández-Mayoral M, Bergner F, Pareige P, Radiguet B, Almazouzi A (2010) On the correlation between irradiation-induced microstructural features and the hardening of reactor pressure vessel steels. J Nucl Mater 406(1):84–89

Lehtinen A, Laurson L, Granberg F, Nordlund K, Alava MJ (2018) Effects of precipitates and dislocation loops on the yield stress of irradiated iron. Sci Rep 8(1):6914

Lépinoux J, Kubin LP (1987) The dynamic organization of dislocation structures: a simulation. Scr Metall 21(6):833–838. https://doi.org/10.1016/0036-9748(87)90332-2. http://www.sciencedirect.com/science/article/pii/0036974887903322

Liu XY, Biner S (2008) Molecular dynamics simulations of the interactions between screw dislocations and self-interstitial clusters in body-centered cubic Fe. Scr Mater 59(1):51–54

Madec R, Devincre B, Kubin LP (2004) On the use of periodic boundary conditions in dislocation dynamics simulations. In: Kitagawa H, Shibutani Y (eds), Proceeding of the IUTAM Symposium on Mesoscopic Dynamics of Fracture Process and Materials Strength, held in Osaka, Japan 6–11, pp 35–44. https://www.springer.com/us/book/9781402020377

Marian J, Wirth BD, Schäublin R, Odette G, Perlado JM (2003) MD modeling of defects in Fe and their interactions. J Nucl Mater 323(2–3):181–191

Marian J, Martinez E, Lee HJ, Wirth BD (2009) Micro/meso-scale computational study of dislocation-stacking-fault tetrahedron interactions in copper. J Mater Res 24(12):3628–3635

Martinez E, Marian J, Arsenlis A, Victoria M, Perlado J (2008a) Atomistically informed dislocation dynamics in FCC crystals. J Mech Phys Solids 56(3):869–895

Martinez E, Marian J, Arsenlis A, Victoria M, Perlado J (2008b) A dislocation dynamics study of the strength of stacking fault tetrahedra. Part I: interactions with screw dislocations. Philos Mag 88(6):809–840

Martinez E, Marian J, Perlado J (2008c) A dislocation dynamics study of the strength of stacking fault tetrahedra. Part II: interactions with mixed and edge dislocations. Philos Mag 88(6):841–863

Matzke H (1992) Radiation damage in nuclear materials. Nucl Instrum Methods Phys Res Sect B Beam Int Mater Atoms 65(1–4):30–39

McMurtrey MD, Was GS, Cui B, Robertson I, Smith L, Farkas D (2014) Strain localization at dislocation channel–grain boundary intersections in irradiated stainless steel. Int J Plast 56:219–231

Mohles V (2001) Orowan process controlled dislocation glide in materials containing incoherent particles. Mater Sci Eng A 309:265–269

Mohles V (2004) The critical resolved shear stress of single crystals with long-range ordered precipitates calculated by dislocation dynamics simulations. Mater Sci Eng A 365(1–2):144–150

Monnet G (2006) Investigation of precipitation hardening by dislocation dynamics simulations. Philos Mag 86(36):5927–5941

Monnet G, Devincre B, Kubin L (2004) Dislocation study of prismatic slip systems and their interactions in hexagonal close packed metals: application to zirconium. Acta Mater 52(14):4317–4328

Monnet G, Osetsky YN, Bacon DJ (2010) Mesoscale thermodynamic analysis of atomic-scale dislocation–obstacle interactions simulated by molecular dynamics. Philos Mag 90(7–8):1001–1018

Monnet G, Naamane S, Devincre B (2011) Orowan strengthening at low temperatures in bcc materials studied by dislocation dynamics simulations. Acta Mater 59(2):451–461

Mura T (1987) Micromechanics of Defects in Solids, 2nd edn. Kluwer Academic Publishers, Dordrecht

Nogaret T, Rodney D, Fivel M, Robertson C (2008) Clear band formation simulated by dislocation dynamics: role of helical turns and pile-ups. J Nucl Mater 380(1–3):22–29

Odette G, Lucas G (2001) Embrittlement of nuclear reactor pressure vessels. JOM 53(7):18–22

Osetsky YN, Bacon DJ (2003) Void and precipitate strengthening in α-iron: what can we learn from atomic-level modelling? J Nucl Mater 323(2–3):268–280

Osetsky YN, Bacon DJ (2010) Atomic-scale mechanisms of void hardening in BCC and FCC metals. Philos Mag 90(7–8):945–961

Osetsky YN, Stoller RE, Rodney D, Bacon DJ (2005) Atomic-scale details of dislocation–stacking fault tetrahedra interaction. Mater Sci Eng A 400:370–373

Osetsky YN, Rodney D, Bacon DJ (2006) Atomic-scale study of dislocation–stacking fault tetrahedron interactions. Part I: mechanisms. Philos Mag 86(16):2295–2313

Phythian W, Stoller R, Foreman A, Calder A, Bacon D (1995) A comparison of displacement cascades in copper and iron by molecular dynamics and its application to microstructural evolution. J Nucl Mater 223(3):245–261

Po G, Ghoniem N (2014) A variational formulation of constrained dislocation dynamics coupled with heat and vacancy diffusion. J Mech Phys Solids 66:103–116

Po G, Lazar M, Seif D, Ghoniem N (2014a) Singularity-free dislocation dynamics with strain gradient elasticity. J Mech Phys Solids 68:161–178

Po G, Mohamed MS, Crosby T, Erel C, El-Azab A, Ghoniem N (2014b) Recent progress in discrete dislocation dynamics and its applications to micro plasticity. JOM 66(10):2108–2120

Po G, Lazar M, Admal NC, Ghoniem N (2017) A non-singular theory of dislocations in anisotropic crystals. Int J Plast 103:1–22

Queyreau S, Monnet G, Devincre B (2010) Orowan strengthening and forest hardening superposition examined by dislocation dynamics simulations. Acta Mater 58(17):5586–5595

Queyreau S, Monnet G, Wirth BD, Marian J (2011) Modeling the dislocation-void interaction in a dislocation dynamics simulation. MRS Proceedings, Mrsf10-1297:10-61. https://doi.org/10.1557/opl.2011.462

Rhee M, Stolken JS, Bulatov VV, Diaz de la Rubia T, Zbib HM, Hirth JP (2001) Dislocation stress fields for dynamic codes using anisotropic elasticity: methodology and analysis. Mater Sci Eng A 309:288–293

Ringdalen I, Wenner S, Friis J, Marian J (2017) Dislocation dynamics study of precipitate hardening in Al–Mg–Si alloys with input from experimental characterization. MRS Commun 7(3):626–633

Robertson IM, Beaudoin A, Al-Fadhalah K, Chun-Ming L, Robach J, Wirth B, Arsenlis A, Ahn D, Sofronis P (2005) Dislocation–obstacle interactions: dynamic experiments to continuum modeling. Mater Sci Eng A 400:245–250

Seeger A (1958) On the theory of radiation damage and radiation hardening. In: Proceedings of the second United Nations International Conference on the Peaceful Uses of Atomic Energy, held in Geneva, 6:250. United Nations, New York

Shenoy VB, Kukta RV, Phillips R (2000) Mesoscopic analysis of structure and strength of dislocation junctions in FCC metals. Phys Rev Lett 84(7):1491

Shi XJ, Dupuy L, Devincre B, Terentyev D, Vincent L (2015) Interaction of <1 0 0> dislocation loops with dislocations studied by dislocation dynamics in α-iron. J Nucl Mater 460:37–43

Shin C, Fivel M, Verdier M, Oh K (2003) Dislocation–impenetrable precipitate interaction: a three-dimensional discrete dislocation dynamics analysis. Philos Mag 83(31–34):3691–3704

Sickafus KE, Matzke H, Hartmann T, Yasuda K, Valdez JA, Chodak P III, Nastasi M, Verrall RA (1999) Radiation damage effects in zirconia. J Nucl Mater 274(1–2):66–77

Singh B, Foreman A, Trinkaus H (1997) Radiation hardening revisited: role of intracascade clustering. J Nucl Mater 249(2–3):103–115

Sobie C, Bertin N, Capolungo L (2015) Analysis of obstacle hardening models using dislocation dynamics: application to irradiation-induced defects. Metall Mater Trans A 46(8):3761–3772

Sun LZ, Ghoniem NM, Tong SH, Singh BN (2000) 3D dislocation dynamics study of plastic instability in irradiated copper. J Nucl Mater 283:741–745

Takahashi A, Ghoniem NM (2008) A computational method for dislocation–precipitate interaction. J Mech Phys Solids 56(4):1534–1553

Tang M, Cai W, Xu G, Bulatov VV (2006) A hybrid method for computing forces on curved dislocations intersecting free surfaces in three-dimensional dislocation dynamics. Model Simul Mater Sci Eng 14(7):1139

Terentyev D, Grammatikopoulos P, Bacon D, Osetsky YN (2008) Simulation of the interaction between an edge dislocation and a <1 0 0> interstitial dislocation loop in α-iron. Acta Mater 56(18):5034–5046

Terentyev D, Bacon DJ, Osetsky YN (2010) Reactions between a 1/2<111> screw dislocation and <100> interstitial dislocation loops in alpha-iron modelled at atomic scale. Philos Mag 90(7–8):1019–1033

Terentyev D, Monnet G, Grigorev P (2013) Transfer of molecular dynamics data to dislocation dynamics to assess dislocation–dislocation loop interaction in iron. Scr Mater 69(8):578–581

Van der Giessen E, Needleman A (1995) Discrete dislocation plasticity – a simple planar model. Model Simul Mater Sci 3(5):689–735

Victoria M, Baluc N, Bailat C, Dai Y, Luppo M, Schaublin R, Singh B (2000) The microstructure and associated tensile properties of irradiated FCC and BCC metals. J Nucl Mater 276(1–3):114–122

Weinberger CR, Aubry S, Lee SW, Nix WD, Cai W (2009) Modelling dislocations in a free-standing thin film. Model Simul Mater Sci Eng 17(7):075007

Weygand D, Friedman L, Van der Giessen E, Needleman A (2002) Aspects of boundary-value problem solutions with three-dimensional dislocation dynamics. Model Simul Mater Sci Eng 10(4):437

Wirth B, Bulatov V, Diaz de la Rubia T (2002) Dislocation-stacking fault tetrahedron interactions in Cu. J Eng Mater Tech 124(3):329–334

Woo C, Goesele U (1983) Dislocation bias in an anisotropic diffusive medium and irradiation growth. J Nucl Mater 119(2–3):219–228

Woo C, Singh B (1990) The concept of production bias and its possible role in defect accumulation under cascade damage conditions. Phys Status Solidi B 159(2):609–616

Yao Z, Schäublin R, Victoria M (2004) Tensile properties of irradiated Cu single crystals and their temperature dependence. J Nucl Mater 329:1127–1132

Yin J, Barnett DM, Fitzgerald S, Cai W (2012) Computing dislocation stress fields in anisotropic elastic media using fast multipole expansions. Model Simul Mater Sci Eng 20(4):045015

Yoffe EH (1961) A dislocation at a free surface. Philos Mag 6(69):1147–1155

Young F Jr (1962) Etch pit studies of dislocations in copper crystals deformed by bending: I. Annealed crystals; II. Irradiated crystals. J Appl Phys 33(12):3553–3564

Zinkle SJ (2012) Radiation-induced effects on microstructure. In: Konings RJM (ed) Comprehensive Nuclear Materials, vol 1, pp. 65–98. Elsevier Amsterdam.

Zinkle SJ, Singh BN (2006) Microstructure of neutron-irradiated iron before and after tensile deformation. J Nucl Mater 351(1–3):269–284

Mesoscopic Modelling of Irradiation Damage Processes: Bridging Many-Body Mechanics and Thermodynamics in Rate Processes

92

Chung H. Woo

Contents

Abstract

The integration of atomistic mechanics into the macroscopic modeling of rate-processes in solids should ensure the principles of statistical and thermal physics are not violated. Rate constants are customarily derived based on the transition state theory or its variants such as kinetic Monte Carlo simulation, phase-field model, and classical nucleation theory. In this perspective, we examine the assumptions traditionally made in constructing the mesoscopic connection between molecular mechanics and macroscopic thermodynamics, (1) separability of the reaction path from the noise created by lattice vibrations, (2) validity of classical statistics, and (3) existence of quasi-equilibrium among the reactants, the transition state, and the product. We find that the first assumption excludes rate processes at high temperatures or low reaction barrier, while the second excludes those at low temperatures or low reaction barrier. The third assumption has to reconcile with the nonequilibrium nature of rate processes, such as the neglect of entropy production which is proportional to the reaction rate, and thus

C. H. Woo (✉)
The Hong Kong Polytechnic University, Hong Kong SAR, China
e-mail: chung.woo@polyu.edu.hk

excludes processes proceeding at fast rates. On this basis, the traditional approach in mesoscopic modeling is only reliable for rate processes in irradiation-damage accumulation with a high activation barrier and at not-too-high temperatures. The mitigation of these restrictions and evaluation of consequences due to their neglect can be formulated in a mesoscopic model in which the atomic processes in the many-body system are treated dynamically and energy quantization and quantum uncertainty are taken into account explicitly through the Mori-Zwanzig formalism of nonequilibrium statistical mechanics and the quantum fluctuation-dissipation relation.

Irradiation damage starts with the production of crystalline defects followed by their accumulation and interaction with existing microstructure. To achieve an engineering description, mechanics in the atomistic scales of time and space must be transposed to the macroscopic scales through the laws of thermodynamics. Rate-theory models (Bullough 1985; Ghoniem 1989; Woo and Singh 1992) of irradiation damage, which provide a continuum description of microstructure evolution as a result of the balance of the generation, migration, and annihilation at sinks of crystal defects, have dominated the scene since the early years. This approach is effective in expressing the macroscopic effects due to microstructure changes caused by their reactions with the defects. The integration of atomistic mechanics into rate-theory models is implemented via articulating reaction rate constants derived based on an elastic continuum theory assuming isotropic diffusion-controlled kinetics. Model parameters sometimes may be force fitted to experimental data. Yet, a good physical basis of the model is essential, without which one may just be redistributing errors among different components of the model.

   To overcome the limitations of the continuum model, a mesoscopic formulation of mass transport, which takes into account the discreteness, structure, and interatomic interaction of the static crystal lattice, is incorporated into the reaction constants. The progress of this line of work was reviewed in the context of fundamental mechanisms of irradiation-induced creep and growth (Matthews and Finnis 1988; Eyre and Matthews 1993). To be consistent, the concept of sink bias was generalized to include the effects of diffusion anisotropy difference (DAD) between the vacancies and the SIAs (Woo 1988). Thus, going beyond the isotropic continuum model, effects on the diffusion-controlled kinetics due to interactions with applied fields, implications for the reaction kinetics of low-dimensional diffusers, and stress-induced preferred absorption (SIPA) effects due to elasto-diffusion can be integrated into the rate-theoretic approach. The progress is further updated in the first edition of this *Handbook* (Woo 2005).

   Reaction constants in the foregoing efforts are derived based on quasi-equilibrium classical thermodynamics, neglecting the dynamical and many-body nature of atomistic processes. It is not difficult to imagine the consequence when the activation barrier involved has a magnitude comparable to the energy of the average phonons, a situation often encountered in the kinetics of small interstitial impurity atoms (e.g., helium and hydrogen) and SIAs in the crowdion state (in low-dimension

diffusion controlled reactions). Dynamics of fast-moving dislocations and loops encountered in embrittlement and plasticity constitutes another class of examples where interaction with lattice vibrations plays an important role.

Processes fundamental to the mechanics of irradiation-damage accumulation generally involve the change of the state of crystalline defects during their formation, migration, clustering, and reaction with sinks. The changes can be regarded as mono- or bi-molecular reactions proceeding through a transition state, envisaged as a saddle-point on the potential-energy surface in phase space. The kinetics of the associated rate process may be considered in ways similar to the transition-state theory (TST) (Eyring 1935). Assuming separability of the transition state (activation barrier) from its dynamic environment (lattice vibrations) and as a quasi-equilibrium process, the corresponding reaction rate is expressible in the Arrhenius form as the product of the attempt frequency ν_0 and the probability of formation of the transition state $\exp(-\Delta G_t/k_B T)$ within classical Gibbs-Boltzmann statistics (Glasstone et al. 1941; Laidler and King (1983)). Other methodologies often used in the modeling of rate-processes, including irradiation damage accumulation, such as kinetic Monte Carlo, phase-field model, and classical nucleation theory, are also similarly formulated. Here ΔG_t is the free-energy of formation of the transition state under quasi-equilibrium conditions, which may also be interpreted as the saddle-point energy that must be provided to complete the reaction reversibly, k_B is the Boltzmann constant and T is the absolute temperature.

In the simple example of migration in real space, the point-defect diffusivity D can be written in the Arrhenius form (Vineyard 1957; Mehrer 2007):

$$D = D_0 e^{S_m/k_B} \exp\left(-E_m/k_B T\right), \tag{1}$$

where D_0 is the prefactor that depends on the jump geometry and vibration frequency of the migrating species. S_m and E_m are, respectively, the migration entropy and enthalpy. Since the saddle-point configuration is coupled to the vibrating lattice (and spins) in the neighborhood, the thermodynamics of its formation cannot be isolated from the way the vibration modes are occupied. Proper accounting of the many-body dynamics is required to relate the corresponding atomistic mechanics and thermodynamics.

Indeed, in his critical investigation of the accuracy of models of rate processes such as the TST, Kramers (1940) explicitly showed that the environmental noise derived from many-body dynamics are important controlling factors of the reaction velocity. More recent works also showed that environmental noise has important consequences in a wide scope (Swinburne et al. 2013, 2014, 2017; Pavliotis and Vogiannou 2008; Ala-Nissila and Ying 1992). Accordingly, kinetics of point-defect processes, such as those governing the lifetime of point-defect configurations and the corresponding mobility and diffusional anisotropy, are only sensitive to the static saddle-point configuration when the relative importance of the environmental noise is sufficiently small. This is consistent with the loss of accuracy at high temperatures of the traditional TST which has been shown by Pineda and Schwartz (2006).

In a contemporary study of the many-body effects of lattice vibrations, the statistical occupation of vibration modes has to be accounted for. It is well known that classical statistics is inadequate in explaining the phenomena of the ultraviolet catastrophe, the photoelectric effect, the specific heat, and the associated divergence of the classical entropy. Classical statistics describes lattice dynamics based on a continuous (nonquantized) vibrational energy spectrum, in which all vibration modes are active at any temperature. This is not so where energy levels are discrete (quantized). Indeed, "the granularity of atoms and quantization of energy and how they gives way to apparently continuous and infinitely divisible matter and energy" is how mesoscale science is characterized by Crabtree and Sarrao (2012). In Einstein's model of heat capacity, for example, phonon excitations in bcc iron below 360 K and carbon below 2000 K are basically frozen because the thermal energy needed to excite the average phonon is not available. When the vibration state of the solid becomes dominated by the ground-state phonons, a classical description of the thermodynamics and lattice dynamics can be in significant error if used to interpret experiments conducted at low temperatures, such as resistivity recovery (Dausinger and Schultz 1975; Maury et al. 1978; Fu et al. 2005), electron microscopy (Yi et al. 2015; Sand et al. 2016), and low temperature plasticity (Biget and Saada 1989; Moskalenko et al. 2005).

The situation is even more drastic if the foregoing discussions were to be extended to magnetic metals. An important example is bcc iron, in which the relevant many-body dynamics must also include the spin waves (magnons) (Van Vleck 1945; Van Kranendonk and Van Vleck 1958; Argyle et al. 1963). Based on the experimental spin stiffness of bcc iron from Collins et al. (1969) and Mook and Nicklow (1973), or the magnon DOS calculated from first principles (Halilov et al. 1997), excitation of a magnon mode at low temperatures requires an average energy of ~250–350 meV. The strong energy quantization makes the application of classical statistical distribution to the vibration modes of the spin system untenable for temperatures under ~3000 K.

In addition to energy quantization, effects due to quantum tunneling are also important to reaction rates. Athermal vibrations due to the phonon ground state, the so-called zero-point vibrations, have to be taken into account. The corresponding free energy contributes to rate processes in solids through point-defect formation, migration, and reaction kinetics. Indeed, this athermal component provides the main driving force that (a) effectively lowers the activation barrier via phonon screening and (b) provides the momentum transfer required for the activation jumps, resulting in phonon-assisted tunneling (Flynn and Stoneham 1970; Woo and Wen 2017). More generally, similar athermal field fluctuations have been envisaged as responsible for many physical effects, such as the Lamb shift, the Casimir force, the natural linewidths of energy levels, and the broadening of neutron scattering resonance lines in solids (Milonni 1994).

It is clear that the principles of statistical and thermal physics must be observed in the integration of atomistic mechanics into the macroscopic modeling of irradiation-damage accumulation. TST and its variants, with their well-established physical and

chemical basis, may be ideally suited to provide the rate constants in the formulation of an irradiation-damage model. Yet, according to Pineda and Schwartz (2006) and Miller (1974), traditional TST suffers from the failure of fundamental assumptions pertaining to the separability of the transition state from the environmental noise due to many-body dynamics and the validity of classical mechanics. As we have discussed in the foregoing, the issue of separability sets a high-temperature limit (relative to the activation barrier) beyond which many-body effects cannot be ignored, while the use of classical physics sets a low-temperature limit (relative to the activation barrier) below which quantum effects are important. Further assumption of a quasi-static process has to reconcile with the entropy production, which is proportional to the reaction rate (Kondepudi and Prigogine 1998), and restricts the application to slow reactions. Indeed, weaknesses such as these are also present in traditional irradiation damage modeling, namely, the kinetic Monte Carlo method, the phase-field method, and classical nucleation model. If these fundamental issues are to be addresses, the quantum, many-body, and dynamical nature of the mechanics of the reactions need to be properly introduced. A key concept in such a formulation is the bridge between the quantum mechanical world of discrete energy, space, and configurations and the engineering world of temperatures and thermodynamic functions which is provided by the fluctuation-dissipation relation (see Fig. 1). Progress in the development of this approach will be assessed and issues discussed in Sect. 1. The effects of many-body dynamics and quantum statistics on reaction kinetics will be discussed in Sect. 2. A summary and outlook is provided in Sect. 3.
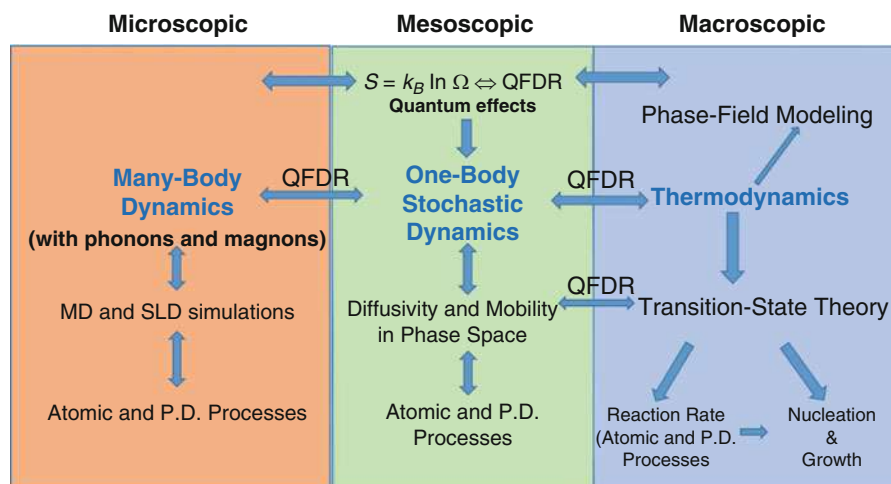
Fig. 1 (color on line) Schematics to illustrate the implementation of quantum many-body dynamics in thermodynamics and reaction kinetics via the Mori-Zwanzig formalism using quantum fluctuation-dissipation relation (QFDR)

1 Fluctuation–Dissipation Relation: The Bridge between Mechanics and Thermodynamics

The importance of environmental noise in the theory of rate processes is well known since the celebrated work of Kramers (1940). That the many-body effects on the mechanics of a many-body system can be represented as the action of stochastic noise is also well established from the works of Mori (1965) and Zwanzig (1965). It is natural that these effects are considered at a fundamental level in the formulation of irradiation-damage models, or at least the conditions of validity of the models be made clear. This is an important concern because the physical link between the rate processes at the atomistic-scale and their manifestation in the macroscopic scale lies in the relation between the mechanical behavior of atoms and the macroscopic laws of thermodynamics. Similar to the famous Boltzmann relation between the thermodynamic concept of entropy and the mechanical microstates, the fluctuation-dissipation relation (FDR) connects the stochastic mechanics of a many-body system and the thermodynamics of the corresponding process (Kubo 1966).

1.1 Many-Body Mechanics in Terms of the Fluctuation–Dissipation Ratio

In principle, many-body dynamics can be treated by directly solving the coupled equations of motion among all particles of the system, such as achieved commonly in a molecular dynamics or spin-lattice dynamics (Ma et al. 2008; Woo et al. 2015) approach. Alternatively, the equation of motion of a state variable in a many-body system can also be formally expressed through a projection operator formalism in the form of a generalized Langevin equation, with the dissipation and fluctuation terms expressed via the Liouville equation (Liouville 1838; Gibbs 1902). This representation of the many-body problem provides a route to an analytic link between the mechanics and thermodynamics of the system of atoms via a pseudo-one-body scheme in which many-body effects are subsumed into a functional relation between the fluctuation-dissipation ratio and the thermodynamic temperature. Indeed, in the earlier works of Einstein (1905) and Smoluchowski (1906), the many-body environment of a diffusing Brownian particle was already modelled using random forces and associated dissipative friction related through the fluctuation-dissipation theorem (Mori 1965; Kubo 1966). The mobility and diffusivity derived this way has a form not bearing any obvious resemblance to the Arrhenius law. Along a similar route, Munakata and Tsurui (1979) and Combs and Kunz (1987) modelled the many-body dynamics of a highly mobile defect in a crystal as a particle under stochastic forces in a periodic force field.

Based on the Mori-Zwanzig formalism (Mori 1965; Zwanzig 1965), Woo and Wen (2017) analyzed the general mechanics of a system of N atoms in an infinite crystal. The mass of the nth atom is denoted by m_n, its momentum $\mathbf{p}_n$ and position $\mathbf{r}_n$. $U(\{\mathbf{r}_n\})$ is the interatomic potential (many-body) corresponding to the lattice

configuration $\{\mathbf{r}_n\}$, neglecting for the time being the interaction between atomic spins. Following Ehrenfest (1927), these authors expressed the equation of motion for each atom in terms of the expectation values which is denoted by $\langle \ldots \rangle_E$. Assuming that the difference between the quantum and classical forces $\left\langle \frac{\partial U}{\partial \mathbf{r}_n} \right\rangle_E - \frac{\partial U(\mathbf{r}_n)}{\partial \mathbf{r}_n}$ that accounts for the quantum uncertainty due to the noncommutative relation between $\mathbf{r}$ and $\mathbf{p}$ may be subsumed into the Mori-Zwanzig random forces that accounts for the fluctuations due to many-body and quantum effects, the equation of motion of each particle can be formally put in the Langevin form. Within this approximation, coherency of the one-particle wave function of the atoms over different sites is neglected, which is justified by noting that the De Broglie wavelength $\lambda \approx \frac{\hbar}{m v_G} \approx 10^{-12}$m of an atom in crystals is much smaller than the interatomic distances under consideration. Assuming a local memory function, the set of pseudo-one-particle Langevin equations of motion can be written. The random forces on the atom n due to its environment (the rest of the crystal) are characterized by the dissipative friction with drag-coefficient g and delta-correlated remnant fluctuations with strength characterized by s, i.e.,

$$\langle \mathbf{f}_n(t) \rangle = 0 \text{ and } \langle f_{in}(t) f_{jl}(t') \rangle = s \delta_{ij} \delta_{nl} \delta(t - t') \tag{2}$$

Here subscripts i and j denote Cartesian components and $\langle \cdots \rangle$ denotes the statistical mean taken over all the possible realizations of the stochastic forces. The corresponding phase-space probability density $W(\{r_{ni}\}, \{p_{ni}\}, t)$ of the crystal is determined by the Fokker-Planck equation. Woo and Wen (2016) showed that the solution of $W(\{r_{ni}\}, \{p_{ni}\})$ at equilibrium is given by $W = C \exp(-\kappa E)$, where C is a normalization constant, E is the total energy of the system, $\kappa^{-1} = s/2g \equiv \eta$ is the fluctuation-dissipation ratio which defines the Langevin equation representing the mechanics of the atoms subjected to the environmental noise. The corresponding phase-space probability density for energy E can then be obtained,

$$W(\{r_{ni}\}, \{p_{ni}\}) = Z \exp[-\kappa E(\{r_{ni}\}, \{p_{ni}\})]. \tag{3}$$

where Z is a normalizing constant.

1.2   Quantum Statistics and Thermodynamics

The probability of the jth state of the $3N$ independent degrees of freedom of the N-atom system from Eq. (3) can be written as $P_j = Z \exp(-\kappa \varepsilon_j)$, where ε_j is the corresponding energy and $Z = \sum_{j=1}^{3N} e^{-\kappa \varepsilon_j}$ is the partition function. P_j maximizes the corresponding mean entropy per atom $\bar{s} = \sum_{j=1}^{3N} P_j \ln P_j$ (in units of k_B) under

the constraints $\sum_{j=1}^{3N} P_j = 1$ and $\sum_{j=1}^{3N} P_j \varepsilon_j = \bar{\varepsilon} \equiv \frac{E}{3N}$. Z is then linked to the mean energy per atom by $\bar{\varepsilon} = -\frac{\partial \ln Z}{\partial \kappa}$ and to the corresponding entropy by $\bar{s} = \kappa \bar{\varepsilon} + \ln Z$. In this regard, $f \equiv -\eta \ln Z = \bar{\varepsilon} - \eta \bar{s}$ is the corresponding Helmholtz free energy per atom. Since both $\bar{\varepsilon}$ and $\bar{s}$ are functions of κ that dictates the phase-space probability density, we may write $d\bar{s} = \kappa d\bar{\varepsilon} + \bar{\varepsilon} d\kappa + \frac{d \ln Z}{d\kappa} d\kappa$, in which the last two terms cancel, yielding $\frac{d\bar{\varepsilon}}{d\bar{s}} = \eta$. Accordingly, if the functional relation between η and the temperature T is known, the thermodynamics (equilibrium) of the system is completely and uniquely defined as a function of the mean energy $\bar{\varepsilon}$. Noting that the effects of fluctuation and dissipation are balanced at equilibrium, Woo and Wen (2017) obtain the relation $\eta = \frac{2}{3N} E_K(T)$, E_K being the kinetic energy of the ensemble which is a function of temperature. In terms of the system energy and entropy, we may write $E = 3N\eta$ and $S = 3N\bar{s}$.

If quantum effects can be neglected, Equipartition theorem gives $E_K = 3Nk_BT/2$, yielding $\eta = k_BT$, the classical fluctuation-dissipation relation (CFDR) for a system of N particles (Mori 1965; Kubo 1966). In this case, the temperature of the system is given by $\frac{dE}{dS} = k_BT$, which is sometimes used as a thermodynamic definition of temperature. The corresponding Helmholtz free energy is then given by $F = E - k_BTS$, a standard result of classical thermodynamics. The Langevin heat bath can also be defined accordingly.

The Equipartition theorem does not hold in quantum mechanics and $\frac{dE}{dS} \neq k_BT$ in general. When the environmental noise is coming from lattice vibrations, standard quantum theory of phonons (Kittel 2004) yields

$$\eta(T) = \frac{E}{3N} = \frac{1}{3N} \sum_{k} \sum_{\alpha=1}^{3} \hbar\omega_\alpha(\mathbf{k}) \left[\frac{1}{e^{\hbar\omega_\alpha(k)/k_BT} - 1} + \frac{1}{2} \right], \tag{4}$$

from which a *nonlinear* relation between η and T can be defined via the phonon dispersion relation, which is governed by a temperature-dependent spring constant. The associated quasi-harmonic approximation (Woo et al. 2015) will be elucidated in the following subsection. It is clear that the right hand side of Eq. (4) reduces to k_BT in the classical limit when $\hbar \to 0$. This relation between η and T is called the *Quantum Fluctuation-Dissipation Relation* (QFDR) for the lattice system. It then follows that all thermodynamic properties of the crystal are functions of temperature via the QFDR $\eta(T)$ in Eq. (4).

Within the Debye model, $\eta(T)$ can be written in close form as

$$\eta(T) = \frac{3}{8} k_B \Theta_D + 3k_B \Theta_D \left(\frac{T}{\Theta_D} \right)^4 \int_0^{\frac{\Theta_D}{T}} \frac{x^3}{e^x - 1} dx, \tag{5}$$

where Θ_D is the Debye temperature of the crystal. The heat capacity, for example, is given by $C_V(T) = \frac{\partial E}{\partial \eta} \frac{d\eta}{dT} = 3Nk_B f_D \left(\frac{\Theta_D}{T} \right)$, which is precisely the well-known expression for the heat capacity that can be derived within the Debye model

directly from quantum statistics. The vibrational entropy $S = \int_{\eta(0)}^{\eta(T)} \frac{1}{\eta} \frac{dE}{d\eta} d\eta' =$ $3N \ln\left(\frac{\eta(T)}{\eta(0)}\right)$ can be derived from the thermodynamic relation $\frac{dE}{dS} = \eta$ obtained earlier. As $T \to 0$, $S \to 3N\left(\frac{8\pi^4}{15}\left(\frac{T}{\Theta_D}\right)^4\right)$, which obeys the third law, vanishing as T^4. The quantum entropy contrasts the classical entropy derived from the CFDR which diverges as $T \to 0$. At high temperatures, $S \to 3N \ln\left(\frac{8T}{3\Theta_D}\right)$ which is the same as the classical result up to a constant.

At low temperatures, $\eta(T)$ in Eq. (5) has an athermal component $\eta_Q = 3k_B \Theta_D/8$ attributable to the phonon ground state, which has a purely quantum origin from the noncommutative relation between $\mathbf{r}$ and $\mathbf{p}$. In this regard, Van Kampen (1992) interpreted that the ensuing quantum uncertainty is effectively an internal noise. Within this interpretation, the athermal term in the fluctuation-dissipation ratio η is simply a reflection of the contribution of quantum noise on the stochastic dynamics of the ensemble. This term corresponds to the difference between the quantum and classical forces $\left\langle \frac{\partial U}{\partial \mathbf{r}_n} \right\rangle_E - \frac{\partial U(\mathbf{r}_n)}{\partial \mathbf{r}_n}$ we mentioned earlier.

At higher temperatures ($T > \Theta_D$), $\eta(T)$ in Eq. (5) is dominated by the thermal component. With decreasing temperature, the dominance wanes and begins to reverse near $T \approx \Theta_D/3$. Thus, at $\Theta_D/3$, the athermal component is already ~50% larger than the thermal one and becomes 20 times greater at $\Theta_D/5$. Therefore, at temperatures below $\Theta_D/3$, it is the athermal ground-state phonons that control the frequency of the activation jumps by (a) lowering the activation barrier by screening and (b) providing the momentum required for the jumps, which is effectively activation by phonon-assisted tunneling (Flynn and Stoneham 1970).

1.3   Thermodynamics of Ferromagnetic Metals

The methodology of large-scale spin-lattice dynamics simulation was first developed under a classical mechanics formulation in the thesis work of Ma et al. (2008) for the mesoscopic treatment of magnetic metals. This scheme was used by Wen and Woo (2014) to calculate the vacancy migration and formation entropies and enthalpies of BCC iron based on classical thermodynamic integration under a reversible thermodynamic path. The development for the quantum treatment of spin vibrations in ferromagnetic metals was based on the quantum fluctuation-dissipation relation developed for lattice vibrations using a Debye theory modified to take into account magnon softening at the ferro-para magnetic transition point (Woo et al. 2015). Atoms in a ferromagnetic metal are modeled as interacting particles with Heisenberg spins embedded in noisy media that represent thermally connected many-body atomic and spin environments (in thermodynamic equilibrium) within the Mori-Zwanzig's formalism.

This theory was applied to the typical case of BCC iron. The dynamics of the many-body system of atoms and spins is solved numerically in a temperature range between 10 and 1400 K. Thermodynamic functions, namely, vibrational energies, heat capacities, and entropies, were calculated from three models, using (1) spin-lattice dynamics with QFDR (quantum and quasi-harmonic); (2) spin-lattice dynamics with CFDR (classical and quasi-harmonic); and (3) analytic QH model based on the 0 K quantum-harmonic phonons and magnons. In Fig. (2), results for the total (spin + lattice) heat capacity from Wen and Woo (2016) are shown together with the experimental data from Wallace et al. (1960). At low temperatures, excellent agreement can be seen among all the quantum results (QFDR and QH) and experimental ones. The nonlinear temperature dependence of the heat capacities that vanish at 0 K is the consequence of Bose-Einstein statistics used in the quantum models. The vanishing of the heat capacity at 0 K is both experimentally observed and theoretically required. The corresponding classical heat capacity based on CFDR remains nonzero at 4 k_B at 0 K, consistent with the equipartition theorem. Historically, Einstein was the first to show that this can be corrected in a way similar to the problem of ultraviolet catastrophe by taking into account Planck's quantization in the excitations of the vibrational modes of the crystal lattice (Donald 2005).

Accuracy of the QH model deteriorates as spin vibrations become increasingly anharmonic at higher temperatures. The excellent agreement between the QFDR and the experimental heat capacity remains intact even beyond 1000 K, successfully reproducing the characteristic sharp rise near the Curie temperature as the spring stiffness for spin vibrations softens. Indeed, harmonic phonons and magnons do
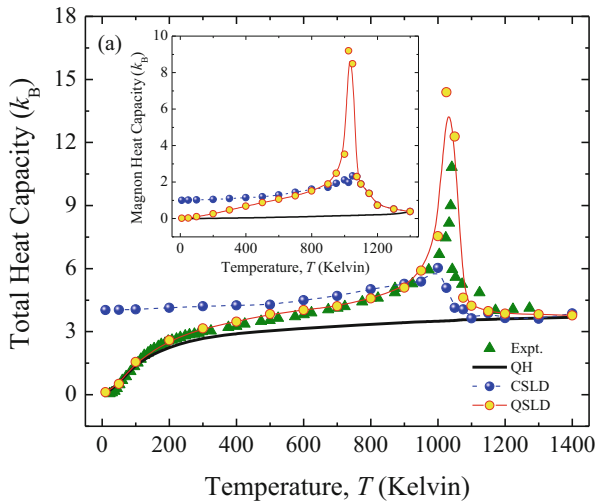
Fig. 2 (color on line) Total heat capacities of ferromagnetic bcc iron (involving interacting phonons and magnons) calculated from QH, QSLD and CSLD models as a function of temperature. Experimental data of Wallace et al. (1960) are plotted for comparison

not interact, which explains the absence of the peak in the harmonic QH model. The interaction between the lattice and spin vibrations is accounted for, although classically, within the CFDR model. This is what gives rise to the suppressed peak in the heat capacity.

The pioneering works of Ma et al. (2008), Wen and Woo (2014, 2016), and Woo et al. (2015) on bcc iron exemplify the complexity and subtlety of the mesoscopic connection between many-body mechanics in the atomistic scale and macroscopic thermodynamics for rate processes in ferromagnetic metals. More importantly, they also show that quantum effects may be important even at temperatures traditionally believed to be safe for the application of the classical thermodynamics.

2   Formulating Reaction Rates in Rate Processes in a Dynamic Scheme

Christiansen (1936) and Kramers (1940) modelled chemical reactions dynamically as a drift-diffusion process in phase space. In this way they avoided having to use the questionable assumption of quasi-equilibrium among the reactants, the transition state, and the product in a chemical process. In this approach, reactants are brought from a state A with equilibrium energy $-E_A$ to a product represented by state B with equilibrium energy $-E_B$ via the saddle point C (see Fig. 3). In this way, rate processes in a many-body environment may be formally considered as "particles" passing from A to B in phase-space in a noisy medium. The corresponding reaction rate may be measured by the particle flux across the saddle point C in terms of the corresponding mobility and diffusivity. The potential-energy gradient $\Delta E_m/a \equiv (E_B - E_A)/2a$ in phase space can be considered as an applied force $\delta \mathbf{F}$ in the effective one-body Langevin dynamics of the particle. Here $E_m = (E_A + E_B)/2$ measures the height of the activation barrier at the saddle-point. These energies (within the adiabatic approximation) are due to the interactions of the "particle" with a *static* environment (no noise) in the two equilibrium configurations before and after the reaction.

Within this framework, the equations of motion of the "particle" in the Mori-Zwanzig formalism can be written in the Langevin form. The probability density $P(\mathbf{r}, t)$ and particle flux are governed by the corresponding Fokker-Planck equation,

Fig. 3 (color on line) Reactants are brought from a state A with equilibrium energy $-E_A$ to a state B with equilibrium energy $-E_B$ via saddle point C
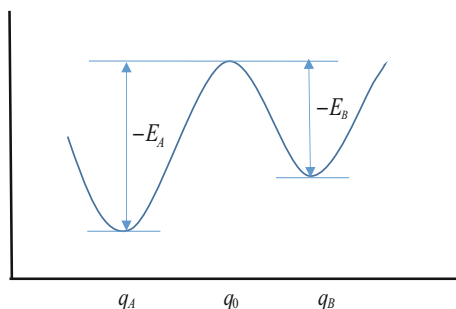
$$\frac{\partial}{\partial t} P\left(\boldsymbol{r}, t\right) = \nabla \cdot \left[\frac{\nabla U}{m\gamma_L} P\left(\boldsymbol{r}, t\right)\right] + D_{\mathrm{E}} \nabla^2 P\left(\boldsymbol{r}, t\right) \equiv -\nabla \cdot \boldsymbol{j}\left(\boldsymbol{r}, t\right). \qquad (6)$$

Here m, $\boldsymbol{r}$, and the velocity $\boldsymbol{v}$ are, respectively, the effective mass, generalized co-ordinate and velocity of the "particle" in phase space. In its trajectory from A to B, the "particle" is subjected to a generalized force $-\nabla U(\boldsymbol{r})$ due to the potential energy gradient, a viscous drag $-m\gamma_L \boldsymbol{v}$ due to the noisy environment, where γ_L is the corresponding frictional coefficient. The environmental noise takes the form of random forces $\sigma_L \boldsymbol{\zeta}(t)$, where $\boldsymbol{\zeta}(t)$ is a Gaussian noise, i.e., $\langle \boldsymbol{\zeta}(t) \rangle = 0$ and $\langle \zeta_i(t) \zeta_j(t') \rangle = \delta_{ij} \delta(t - t')$, with i, j denote the Cartesian components. In this regard, the "particle" described by Eq. (6) can be likened to a Brownian particle embedded in the phonon gas under the action of an external force field (Kramers 1940). Indeed, $D_{\mathrm{E}} = \eta_L / m\gamma_L$ is the *Einstein diffusivity* for a free Brownian particle subjected to the assumption that $|\gamma_L v| >> |\dot{v}|$.

Modelling the potential experienced by the "particle" with a saw-tooth potential in each of the three Cartesian directions, analytic expressions of the phase-space diffusivity D and mobility μ can be derived by solving Eq. (6). The lengthy derivation has been detailed in the Appendix of Woo and Wen (2017). To first order of Δ, the corresponding mean reaction time $\tau_\pm$, which governs the speed of the forward $(+)$ and backward $(-)$ reactions, is given by

$$\tau_\pm = \frac{a^2 m \gamma_L \left(1 + \gamma_m/\gamma_L\right)}{\eta_L} \left(1 \mp \frac{E_B - E_A}{2\eta_L}\right), \qquad (7)$$

where $2a$ is the distance in phase space between the two equilibrium positions. The ratio γ_m/γ_L measures the relative strength of the resistance due to activation barrier versus that due to the lattice.

Physically, the reduced effectiveness of the activation barrier as temperature increases (ξ decreases) is the result of noise-assisted tunneling (Flynn and Stoneham 1970; Woo and Wen 2017).

The reaction rate constant is the inverse of the reaction time in Eq. (7). It only takes the Arrhenius form when the activation barrier is sufficiently high or when the temperature is sufficiently low. When more than one transition states are involved, the reaction rate is the inverse of the sum of reaction times of all processes involved, and the slowest step is the controlling step as expected.

The foregoing general discussion holds independent of whether classical or quantum statistics are used in the derivation. Since classical results can be recovered from the quantum version by taking the limit $\hbar \to 0$, we only need to continue along the quantum route. When the environment is that of a nonmagnetic metal, the quantum fluctuation-dissipation relation (QFDR) can be evaluated from Eq. (4) using the phonon density of state (DOS) $g(\omega) = \frac{4\pi k^2 \Omega}{(2\pi)^3} \frac{1}{\nabla_{\mathbf{k}} \omega}$ where Ω is the atomic volume and $\mathbf{k}$ the phonon wave vector. The DOS can be obtained from experiments or from first-principle calculations, or by simply using the Debye model as Woo et al. (2015).
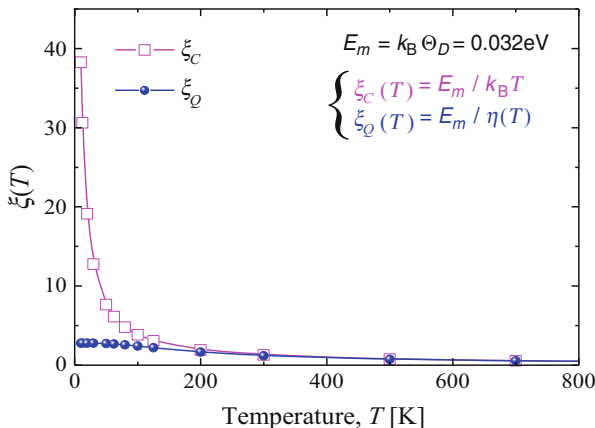
Fig. 4 (color on line) ξ as function of temperature based on classical fluctuation dissipation relation (CFDR) $\xi_C = E_m/k_B T$, and quantum fluctuation dissipation relation (QFDR) $\xi_Q = E_m/\eta(T)$, where $\eta(T)$ is given in Eq. (5) and $E_m = k_B \Theta_D = 0.032$eV (From Woo and Wen 2017)

To visualize the effect of quantum statistics on the many-body dynamics, Woo and Wen (2017) plotted the variable $\xi(T) = \frac{E_m}{\eta_L(T)}$ as a function of temperature T in an example where E_m is hypothetically put equal to the Debye energy of BCC W, i.e., $E_m = k_B \Theta_D = 32$meV. The plot is reproduced here in Fig. 4. It can be seen that above the Debye temperature, i.e., $T > \Theta_D$, quantum effects disappears and $\xi_Q \approx \xi_C$. Below the Debye temperature, i.e., $T < \Theta_D$, CFDR overestimates the phonon density, which is reflected in the large difference between $\xi_C \to \infty$ calculated using CFDR and $\xi_Q = 8/3$ calculated using QFDR as $T \to 0$. If ξ_C is used instead of ξ_Q in Eq. (7), γ_m is greatly over-estimated, causing a serious underestimation of the reaction speed. Thus, referring to Fig. (4), at $T = 10$K, QFDR gives a value of $\xi_Q \approx 8/3$ while CFDR gives a value of $\xi_C \approx 40$. This means that significant reactions between molecules may still occur even if there is insufficient energy to overcome the activation barrier. The same quantum effect on the reaction rate has also been found by Masel (1996).

Within the Debye model, classical thermodynamics is applicable well above the Debye temperature. In this case, $\eta_L = k_B T$ and the QFDR is reduced to the classical fluctuation-dissipation relation (CFDR). The traditional Einstein-Smoluchowski relation between the mobility and diffusivity, namely, $D = \mu k_B T$, can also be recovered (Woo and Wen 2017). For sufficient large activation energy, i.e., $E_m >> k_B T >> k_B \Theta_D$, the classical reaction constant is given by

$$k_{\pm} = \tau_{\pm}^{-1} = \frac{E_m^2}{2a^2 m \gamma_L k_B T} \left(1 \pm \frac{E_B - E_A}{2k_B T} \right) e^{-E_m/k_B T}, \qquad (8)$$

which has the Arrhenius form. Many-body effect is manifested in the dependence of the reaction constant on the high-temperature mobility $\mu_{HT} = (m\gamma_L)^{-1}$, which

measures the viscosity of the medium due to the environmental noise encountered by the reactants. This relation with the lattice vibrations is not considered in the traditional TST derived using a static approach.

When the activation energy is small, i.e., $E_m << k_B T$, $\xi << 1$ and $\frac{\gamma_m}{\gamma_L} \approx \frac{\xi}{3} << 1$ according to Eq. (7). In this case, the reaction time is controlled by the environmental noise, i.e., by γ_L. In comparison, the effectiveness of γ_m is reduced because of phonon-assisted tunneling. In this case, the reaction constant from Eq. (7) is reduced to

$$k_{\pm} = \frac{k_B T}{a^2 m \gamma_L} \left(1 \pm \frac{E_B - E_A}{2 k_B T} \right), \tag{9}$$

which is noticeably independent of the activation barrier and has a temperature dependence that deviates completely from the traditional Arrhenius form. Both are many-body effects that originate from the environmental noise which becomes the controlling factor when the activation barrier E_m is sufficiently low. The deviation of both Eqs. (8) and (9) from the traditional transition state theory is expected from Miller's (1974) observation that the separability assumption used in deriving TST is seldom satisfied. Woo and Wen's (2017) results are also consistent with the observation of Pineda and Schwartz (2006) in the simple exchange reaction between diatomic hydrogen and a hydrogen radical. Indeed, when $E_B - E_A << k_B T$, Eqs. (8) and (9) can be combined to give the modified Arrhenius equation, $k \approx B T^n e^{-E_m / k_B T}$ where $-1 \leq n \leq 1$ in the IUPAC Goldbook (IUPAC 2006).

When quantum effects are taken into account, the frequencies of the vibration modes enter the equation. At low temperatures down to 0 K, the density of the phonon gas is low, but does not vanish due to the athermal phonon ground state. This energy still allows the probable appearance of the particle on the other side of the saddle point without thermal excitation, a phenomenon one may describe as ground-state-phonon-assisted tunneling. This can be clearly seen from the Debye model, within which $\eta_L(T) = \left[\frac{3}{8} + \frac{\pi^4}{5} \left(\frac{T}{\Theta_D} \right)^4 \right] k_B \Theta_D$ from Eq. (5). When the temperature T dips below $\sim \Theta_D / 3$, the second term in the square brackets decreases rapidly as T^4, and η_L tends to a constant value of $\eta_L \approx \frac{3}{8} k_B \Theta_D$. Following Eq. (7), the near-zero reaction constant τ_Z^{-1} is related to the high-temperature mobility μ_{HT}.

Thus, below $\sim \Theta_D / 3$, the reaction constants become temperature independent and is nonzero even at 0 K. The complete departure from both the Arrhenius behavior and the Einstein-Smoluchowski relation is obvious. In particular, when $E_m << \frac{3}{8} k_B \Theta_D$, the reaction constant is also independent of the activation energy and is completely controlled by the drag due to the ground-state phonons, a pure quantum effect. The nonzero reaction rate at 0 K is a consequence of phonon-assisted tunneling resulting from the uncertainty-principle. In addition, below the Debye temperature, the conventional Einstein relation $D = \mu k_B T$ no longer holds due to quantum effects.

The accuracy of the foregoing Langevin-equation approach has been verified by comparing with results from MD simulation (Woo and Wen 2017). In Fig. 5, the
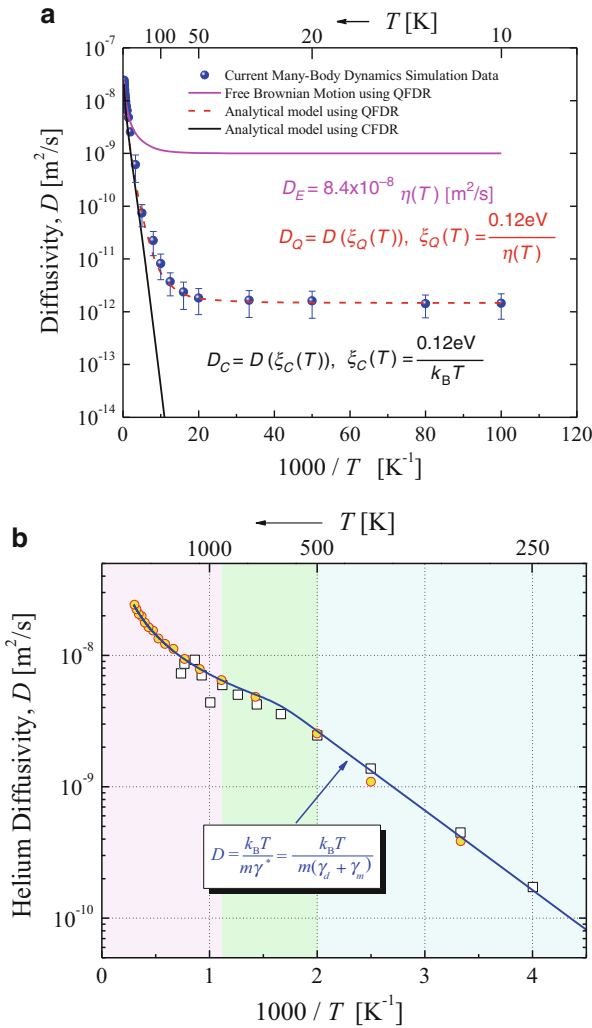
Fig. 5 (color on line) Diffusivities of He in BCC W at temperatures from 10 to 3300 K, calculated from analytical model (Eq. (8)) based on QFDR and CFDR, comparing with corresponding MD results. (**a**) for low temperatures (From Woo and Wen 2017); (**b**) for high temperatures (From Wen et al. 2017)

diffusivity of He in bcc W obtained from MD simulation is shown to only follow the Arrhenius law in a narrow temperature range between ~300 and ~700 K and tends to a constant on the low-temperature side while obeying Einstein's theory and becoming a linear function of temperature on the high-temperature side.

The kinetic description of cluster nucleation and growth may be described as an evaporation-condensation process in which particles are randomly added to or removed from the clusters one at a time. Within the Becker-Döring model (Becker

and Döring 1935), the evolution kinetics can be expressed in terms of a system of hierarchical conservation equations for clusters of increasing sizes. Fluctuations due to lattice vibrations are neglected. If the size of the fluctuations is sufficiently small compared to the size of the cluster, this set of equations may be approximated by the corresponding Fokker-Planck (F-P) equation which is formally a drift-diffusion equation in the cluster-size space (Kelton and Greer 2010). The drift term represents the deterministic (macroscopic) growth rate of the clusters and the diffusion term represents the statistical spread of cluster sizes due to the random fluctuations in the chemical potential. It is on this basis that the well-known Zeldovich-Frenkel equation in classical nucleation theory was derived (Zeldovich 1942; Frenkel 1946). Inherent assumptions made in the Becker-Döring model (1936) from which the Zeldovich-Frenkel equation is derived are the same as that of the TST due to Eyring (1935), namely, separability, quasi-equilibrium, and classical mechanics. The Arrhenius temperature dependence of the nucleation rate so obtained is affected in a similar way when many-body dynamics is taken into account.

3   Summary and Outlook

In a previous contribution to the First Edition of this Handbook, we assessed the efforts seeking to overcome the limitations of modeling irradiation damage based on the continuum theory. The atomistic description of reactions among crystalline defects was integrated into the macroscopic rate-theory model via rate constants derived using a mesoscopic model based on the Gibbs-Boltzmann statistics. The present article continues in this direction but with emphasis on the fundamental principles that can bridge the scales between atomistic mechanics and macroscopic thermodynamics which can then provide a new basis for the various irradiation-damage modeling methodologies. We focus on the shortcomings of the three basic assumptions: (1) separability of the reaction path from the environmental noise, (2) validity of classical statistics, and (3) quasi-equilibrium between the reactants, products, and the transition state. The first assumption prevents rate processes at high temperatures or those with low reaction barrier to be treated properly. The second excludes processes at low temperatures or those with low reaction barrier. The third assumption ignores processes with low activation barrier. As a result, the traditional approach is only reliable for rate processes with a high activation barrier and at intermediate temperatures. To mitigate the limited applicability of the reaction constants so derived, we consider a dynamical approach using a Langevin equation scheme in the Mori-Zwanzig formalism. The many-body dynamics is solved and linked to the thermodynamics and reaction kinetics using the quantum fluctuation-dissipation relation via the Debye theory. Both analytical and MD results are consistent and demonstrates significant effects of lattice and spin dynamics on the reaction rates. One can expect improved accuracy in the descriptions of the kinetics of the general, anisotropic, and low-dimensional diffusion-controlled reactions, clustering of small interstitial atoms (He, H) and self-interstitials, the interpretation of low-temperature experiments, plasticity and dislocation dynamics.

Temperature dependence is clearly central to the theory of rate processes, the issue having been studied for well over a century, as witnessed by the dominant role of the Arrhenius law. Based on similar assumptions, Eyring's equation of TST and the Zeldovich-Frenkel equation of classical nucleation theory also had the Arrhenius form. Given the basic nature of thermodynamics and reaction kinetics, the universality of the environmental noise, quantization and quantum uncertainty, and significance of the effects due to quantum many-body dynamics discussed here, the awareness of a science frontier at the mesoscale (Crabtree and Sarrao 2012; Yip and Short 2013) is a noteworthy development with relevance to radiation damage (Short and Yip 2015). This provides a window to a more fundamental and mechanism-based understanding of reaction and nucleation kinetics (diffusion- or reaction-controlled), all sink biases, production bias, point-defect configurations, diffusivities and mobility, isochronal annealing rates of irradiation damage. The new frontiers tend to be in regimes characterized by extreme environments where temperatures can be very high or very low, activation energies can be low, e.g., self-interstitial migration, and when ferromagnetic metals are involved. While the present discussions are concerned with crystalline materials, the general responses considered, classical versus quantum treatment, low and high activation barriers, and dynamics of many-body interactions have counterparts for amorphous or other states of matter.

References

Ala-Nissila T, Ying SC (1992) Theory of classical surface diffusion. Prog Surf Sci 39:227

Argyle BE, Charap SH, Pugh EW (1963) Deviation from $T^{3/2}$ law for magnetization of ferrometals: Ni, Fe and Fe+3%Si. Phys Rev 132:2051

Becker R, Döring W (1935) Kinetische Behandlung der Keimbildung in ubersattigten Dampfern. Ann d Physik 24:719

Biget MP, Saada G (1989) Low-temperature plasticity of high-purity α-titanium single crystals. Phil Mag A 59:747–757

Bullough R (1985) Dislocations and radiation damage. In: Proceedings of the conference on Dislocations and properties of real materials. The Institute of Metals, London

Christiansen JA (1936) On an extension of Arrhenius's view of the chemical reaction. Z Phys Chem B33 145:1936

Collins MF, Minkiewicz VJ, Nathans R, Passell L, Shirane G (1969) Critical and spin-wave scattering of neutrons from iron. Phys Rev 179:417

Combs JA, Kunz C (1987) Perturbation theory of impurity diffusion. Phys Rev B 36:289

Crabtree GW, Sarrao JL (2012) Opportunities for mesocale science, Mater Soc Bull 37:1079

Dausinger F, Schultz H (1975) Long-range migration of self-interstitial atoms in tungsten. Phys Rev Lett 35:1773

Donald R (2005) Einstein's other theory: the Planck-Bose-Einstein theory of heat capacity. Princeton University Press, Princeton, p 73

Ehrenfest P (1927) Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik. Z Phys 45:455–457

Einstein A (1905) On the motion – required by the molecular kinetic theory of heat – of small particles suspended in a stationary liquid. Ann Phys 322:549

Eyre BL, Matthews JR (1993) Technological impact of microstructural evolution during irradiation. J Nucl Mater 205:1

Eyring H (1935) The activated complex in chemical reactions. J Chem Phys 3(1935):107

Flynn CP, Stoneham AM (1970) Quantum theory of diffusion with application to light interstitials in metals. Phys Rev B 1:3966

Frenkel YI (1946) Kinetic theory of liquids. Oxford University, Oxford

Fu C-C, Dalla TJ, Willaime F, Bocquet J-L, Barbu A (2005) Multiscale modelling of defect kinetics in irradiated iron. Nat Mater 4:68

Ghoniem NM (1989) Stochastic theory of diffusional planar-atomic clustering and its application to dislocation loops. Phys Rev B 39:11810

Gibbs JW (1902) Elementary principles in statistical mechanics. Charles Scribner's Sons, New York

Glasstone S, Laidler KJ, Eyring H (1941) The theory of rate processes. McGraw-Hill, New York

Halilov SV, Perlov AY, Oppeneer PM, Eschrig H (1997) Magnon spectrum and related finite-temperature magnetic properties: a first-principle approach. Europhys Lett 39:91

IUPAC (2006) Compendium of chemical terminology, 2nd edn. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. https://doi.org/10.1351/goldbook. Last update: 2014-02-24; version: 2.3.3. DOI of this term: https://doi.org/10.1351/goldbook.M03963

Kelton KF, Greer AL (2010) The Classical Theory. In: Kelton KF, Greer AL (eds) Nucleation in condensed matter: applications in materials and biology. Pergamon Material Series, vol 15, Pergamon, Oxford, pp. 19–54

Kittel C (2004) Introduction to solid state physics, 8th edn. Wiley, New York

Kondepudi D, Prigogine I (1998) Modern thermodynamics. Wiley, New York

Kramers HA (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 7:284

Kubo R (1966) The fluctuation-dissipation theorem. Rep Prog Phys 29:255

Laidler KJ, King MC (1983) The development of transition-state theory. J Phys Chem 87: 2657–2664

Liouville J (1838) Note on the theory of the variation of arbitrary constants. J Math Pure Appl 3:342–349

Ma PW, Woo CH, Dudarev SL (2008) Large-scale simulation of the spin-lattice dynamics in ferromagnetic iron. Phys Rev B 78:024434

Masel R (1996) Principles of adsorption and reactions on solid surfaces. Wiley, New York

Matthews JR, Finnis MW (1988) Irradiation creep models – an overview. J Nucl Mater 159:257

Maury F, Biget M, Vajda P, Lucasson A, Lucasson P (1978) Frenkel pair creation and stage I recovery in W crystals irradiated near threshold. Radiat Eff 38:53

Mehrer H (2007) Diffusion in solids. Springer, Berlin/Heidelberg

Miller WH (1974) Quantum mechanical transition state theory and a new semiclassical model for reaction rate constants. J Chem Phys 62:1899–1906

Milonni PW (1994) The quantum vacuum: An Introduction to Quantum Electrodynamics. Academic, Boston

Mook HA, Nicklow RM (1973) Neutron scattering investigation of the magnetic excitations in iron. Phys Rev B 7:336

Mori H (1965) Transport, collective motion, and Brownian motion. Prog Theor Phys 33:423

Moskalenko VA, Natsik VD, Kovaleva VN (2005) The role of Peierls relief in the low-temperature plasticity of pure α-Ti. Low Temp Phys 31:907

Munakata T, Tsurui A (1979) Interacting Brownian motion in a periodic potential. Zeitschrift Fur Physik B 34:203

Pavliotis GA, Vogiannou A (2008) Diffusive transport in periodic potentials: underdamped dynamics. Fluct Noise Lett 08:L155

Pineda JRET, Schwartz SD (2006) Protein dynamics and catalysis: the problems of transition state theory and the subtlety of dynamic control. Philos Trans R Soc Lond Ser B Biol Sci 361(1472):1433–1438

Sand AE, Dequeker J, Becquart CS, Domain C, Nordlund K (2016) Non-equilibrium properties of interatomic potentials in cascade simulations in tungsten. J Nucl Mater 470:119

Short MP, Yip S (2015) Materials aging at the mesoscale: kinetics of thermal, stress, radiation activations. Curr Opin Solid State Mater Sci 19:245–252

Smoluchowski M (1906) Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. Ann Phys 326:756

Swinburne TD, Dudarev SL, Fitzgerald SP, Gilbert MR, Sutton AP (2013) Theory and simulation of the diffusion of kinks on dislocations in bcc metals. Phys Rev B 87:064108

Swinburne TD, Dudarev SL, Sutton AP (2014) Classical mobility of highly mobile crystal defects. Phys Rev Lett 113:215501

Swinburne TD, Ma PW, Dudarev SL (2017) Low temperature diffusivity of self-interstitial defects in tungsten. New J Phys 19:073024

Van Kampen NG (1992) Stochastic processes in physics and chemistry. Elsevier Science BV, Amsterdam

Van Kranendonk J, Van Vleck JH (1958) Spin waves. Rev Mod Phys 30:1

Van Vleck JH (1945) A survey of the theory of ferromagnetism. Rev Mod Phys 17:27

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3:121

Wallace DC, Sidles PH, Danielson GC (1960) Specific heat of high purity iron by a pulse heating method. J Appl Phys 31:168

Wen H, Woo CH (2014) Temperature dependence of enthalpies and entropies of formation and migration of mono-vacancy in BCC iron. J Nucl Mater 455:31

Wen H, Woo CH (2016) Quantum statistics and anharmonicity in the thermodynamics of spin waves in ferromagnetic metals. Phys Rev E 94:032104

Woo CH (1988) Theory of irradiation deformation in non-cubic metals: effects of anisotropic diffusion. J Nucl Mater 159:237

Woo CH (2005) Modeling irradiation damage accumulation in crystals. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht, pp 959–986

Woo CH, Singh BN (1992) Production bias due to clustering of point defects in irradiation-induced cascades. Phil Mag A 65:889

Woo CH, Wen H (2017) Quantum statistical effects in the mass transport of interstitial-solutes in a crystalline solid. Phys Rev E 96:032133

Woo CH, Wen H, Semenov AA, Dudarev SL, Ma PW (2015) Quantum heat bath for spin-lattice dynamics. Phys Rev B 91:104306

Yi X, Sand AE, Mason DR, Kirk MA, Roberts SG, Nordlund K, Dudarev SL (2015) Direct observation of size scaling and elastic interaction between nano-scale defects in collision cascade. Europhys Lett 110:36001

Yip S, Short MP (2013) Multiscale materials modeling at the mesoscale. Nat Mater 12:774–777

Zeldovich YB (1942) Theory of nucleation and condensation. Sov J Exp Theor Phys 12:525

Zwanzig R (1965) Time-correlation functions and transport coefficients in statistical mechanics. Annu Rev Phys Chem 16:67

Multiphysics Modeling of Nuclear Materials

93

Benjamin Spencer, Daniel Schwen, and Jason Hales

Contents

Abstract

The materials used in nuclear fuel elements and other structural components within a nuclear reactor are subjected to an aggressive operating environment that subjects them to challenging physical conditions and leads to changes to their microstructure and engineering properties over time. Numerical simulation is increasingly used to characterize the response of these materials at both engineering and material scales. Many problems of interest in nuclear systems inherently require a capability to solve for multiple physics systems in a coupled manner. This chapter provides a summary of the governing equations for some of the

B. Spencer (✉) · D. Schwen · J. Hales
Fuel Modeling and Simulation Department, Idaho National Laboratory, Idaho Falls, ID, USA
e-mail: benjamin.spencer@inl.gov; daniel.schwen@inl.gov; jason.hales@inl.gov

physics that are typically of interest for these problems, numerical discretization techniques, and methods for solution of the resulting systems of equations. In addition, it provides examples of three important nuclear material simulation applications: engineering-scale nuclear fuel performance, hydride formation in zirconium alloy cladding material, and UO_2 microstructure evolution.

The operating environment of nuclear reactors presents many unique challenges both for the materials comprising fuel elements (Olander 1976) and other structural components within and surrounding the reactor (Zinkle and Was 2013). These environmental challenges are due to the simultaneous exposure to conditions from multiple physical systems, including high temperatures, high radiation fluxes, aggressive chemical environments, and mechanical loading.

Numerical simulation has been used for multiple decades as a tool for understanding the behavior of various components in this environment. As these tools have been developed to increasing levels of sophistication, they have been used increasingly both for engineering analysis of components and for simulating the processes of microstructure evolution, which leads to changes in the engineering properties of interest in the materials used in these components.

In some cases, the behavior of a single physics can be considered independently, simply because the response of one system does not change the conditions a component or material is subject to in other physical systems. For example, a component may be subjected to mechanical loading and thermal expansion due to elevated temperatures, but often, the mechanical response has a negligible effect on the thermal environment, so the temperature field can simply be imposed as a boundary condition to the mechanical model. However, there are also many scenarios in which the response to one physics has a strong effect on other physics. In the thermal/mechanical example, if the mechanical response of that component significantly changes its configuration, that could alter paths for heat transfer and have a dramatic effect on the thermal field. In such a case, the simulation should account for two-way feedback between the models of these physical systems to accurately represent its response.

There are many cases in which significant feedback occurs between physical systems within the nuclear reactor environment. These occur at length scales ranging from that of the full reactor system to the material level.

At the reactor scale, there is two-way feedback between the neutron transport and the thermal conditions within the nuclear fuel, which is volumetrically heated by the process of fissioning atoms under neutron flux. The nuclear cross sections, or probabilities of interaction between a nucleus and an incident neutron, are dependent on the temperature. Because of the temperature dependence of the cross sections, the fission rate is dependent on the fuel temperature.

Even if the fission rate is taken as a prescribed condition, computing the temperature in nuclear fuel is a complex multiphysics problem. In light water reactor (LWR) fuel, cylindrical uranium dioxide fuel pellets are stacked within long metallic fuel rods. Heat transfer from the fuel to the coolant outside the fuel rod is highly dependent on the size of the gap between the fuel and cladding, which can evolve

due to the mechanical response of the system. Computing the thermal response of a fuel rod thus necessarily involves solving coupled equations of mechanics and heat conduction.

At the material scale, within ceramic nuclear fuel, exposure to irradiation and elevated temperatures causes significant changes in the grain structure of the fuel. Gaseous fission products diffuse through the solid material and form bubbles. These bubbles, along with larger pores, have a tendency to migrate up the temperature gradient. At high temperatures, grain growth occurs either in an equiaxial or columnar fashion, depending on the local conditions.

Similarly, phenomena affecting the material performance of the zirconium alloys used for cladding in LWRs are driven by multiple interacting physics. Exposure to coolant water causes atomic hydrogen to diffuse through the cladding. Some of this hydrogen precipitates in the form of hydrides, which can form, dissolve, reprecipitate, and change their orientation depending on the temperature and stress state. Under a constant temperature and far-field stress, the process of hydride formation is itself a multiphysics problem, as hydride formation induces local stresses due to the volumetric expansion that occurs during the phase change, and the process of that phase change is affected by the stress state.

Away from the fuel elements, the integrity of the structural elements comprising the reactor system is challenged by exposure to the reactor environment. These elements include core supporting structures, pressure vessels, control rods, piping, steam generators, and concrete structures such as biological shield walls and containment vessels.

Material evolution can be considered at many length and time scales. When considering phenomena of interest at the engineering scale, material behavior can often be homogenized and represented as continuous behavior. When understanding of the processes of microstructure evolution is needed, simulations can be performed at the mesoscopic scale and explicitly include features of the material structure, such as individual grains or constituents of a composite material. At lower scales, other modeling techniques can explicitly include features in crystal structures such as dislocations, voids, self-interstitial atoms, or other matrix defects.

The discussion here is focused on modeling techniques for aspects of material behavior that vary spatially in a relatively continuous manner, so that the problem to be solved is to find spatially resolved field variables whose behavior is governed by partial differential equations (PDEs). The techniques for solving such problems are applicable at a wide variety of length scales, but are not applicable for modeling every aspect of multiphysics material response.

This chapter will first present a brief overview of the characteristics of the major physics affecting fuels and structural components in the nuclear reactor environment. This is followed by a discussion on the techniques used to numerically simulate these coupled systems. This includes both the discretization techniques and techniques for solving the resulting coupled systems of equations. Finally, several representative examples of coupled physics problems relevant to the performance of engineered systems and materials in a nuclear reactor environment are presented, along with details on the numerical simulation of those systems.

1   Characteristics of Physics

Multiple physical systems can potentially have a significant impact on the performance of nuclear reactor components and materials. These include mechanical deformation, heat transfer, species diffusion and reactions, phase change, fluid dynamics, and neutron transport. The characteristics of the physical systems that are typically most relevant to simulation of materials in nuclear reactors are summarized here.

1.1   Solid Mechanics

Materials and structures can experience mechanical deformation due to either external loading or intrinsically generated strains, such as those due to thermal expansion. The primary solution variable is the displacement vector $\boldsymbol{u}$, which varies spatially over the domain and contains components for its values in the directions of the coordinate system employed.

Deformation is governed by stress equilibrium within the volume of the domain:

$$\nabla \cdot \boldsymbol{\sigma} + \boldsymbol{b} = \boldsymbol{0} \tag{1}$$

where $\boldsymbol{\sigma}$ is the stress tensor and $\boldsymbol{b}$ is a vector of body forces (such as gravity loads). These tensors and vectors consist of components in the coordinate system of the model.

The strain tensor, $\boldsymbol{\varepsilon}$, is related to the gradient of the displacement field through a kinematic relationship, shown here for the small-strain case:

$$\boldsymbol{\varepsilon} = \frac{1}{2}\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right) \tag{2}$$

The treatment for finite strains is considerably more complex (Rashid 1993) and not covered here.

The stress is computed from the strain through a constitutive relationship, the details of which are material-specific. Under low stresses, many materials exhibit elastic behavior, in which the stress is linearly related to the applied strain through the elasticity tensor, $\boldsymbol{E}$:

$$\boldsymbol{\sigma} = \boldsymbol{E} : \boldsymbol{\varepsilon} \tag{3}$$

Intrinsically generated strains (also referred to as eigenstrains) are subtracted from $\boldsymbol{\varepsilon}$ before applying Eq. 3. Under higher stresses, materials often exhibit nonlinear inelastic behavior in the form of creep or plasticity. This is usually represented by using an incremental form for stress computation, where stress is a function of the increment in the strain between the current and previous time steps (Simo and Hughes 1998).

The equations shown here are for the quasistatic case, where the deformation occurs slowly or, in other words, when the time derivatives of $\boldsymbol{u}$ are very small, so

that inertial effects are assumed to be negligible. This is often the case for steady-state conditions in nuclear reactor components and materials. Additional terms for the effects of inertia and damping should be included in Eq. 1 if these are important.

1.2   Diffusion Processes

Many physical systems are governed by diffusion processes. Heat transfer and species transport in solids are two such systems of interest in nuclear reactors. These are commonly represented by Fickian diffusion (Fick 1855), in which the change in local concentration of a given species, ϕ over time, t, is governed by the gradient of that concentration:

$$\frac{\partial \phi}{\partial t} = D\nabla^2 \phi \tag{4}$$

where D is the diffusion coefficient. This PDE can be solved for the concentration ϕ, which is the primary solution variable. The flux vector J of a system following this process is expressed as:

$$\boldsymbol{J} = -D\nabla\phi \tag{5}$$

Heat transfer within a solid (in the absence of advection) is governed by Fickian diffusion. In this case, the concentration variable is the temperature, T. The energy balance equation governing heat transfer within a volume, including a source term due to an applied volumetric heating rate $\dot{Q}$, is expressed as:

$$\rho C_p \frac{\partial T}{\partial t} = k\nabla^2 T + \dot{Q} \tag{6}$$

where ρ is the density, C_p is the specific heat at constant pressure, and k is the thermal conductivity. The coefficients C_p and k can depend on temperature, making this a nonlinear equation.

Many diffusion systems of interest are also affected by the Soret effect (Rahman and Saghir 2014), where the concentration of a species is a function of both the concentration gradient and the temperature. An example of such a system in a nuclear environment is hydrogen concentration in LWR fuel cladding (Stafford 2015). Heat and moisture transport in concrete are affected by both the Soret effect and its counterpart, the Dufour effect, in which temperature is affected by the concentration gradient (Bažant et al. 1981).

1.3   Phase-Field Microstructure

In ideal solutions, the diffusive driving force is represented by the concentration gradient, and species concentrations can be computed using diffusion models as described in the previous section. However, in systems other than ideal solutions, where the chemical potential is not a linear function of the concentration, the

diffusive driving force is given by the gradient of the chemical potential rather than the concentration gradient. The modeling of species transport and phase changes on the mesoscale in these systems thus requires knowledge of the system's thermodynamic potential and its kinetic coefficients.

The phase-field method is a well-established tool for simulating the coevolution of microstructure and physical properties at the mesoscale. In the phase-field method, the microstructure is described by a system of continuous variables, also called order parameters. Microstructure interfaces are approximated using a finite width, and the order parameters vary smoothly over those interfaces. The evolution of non-conserved order parameters η_j (e.g., phase regions and grains) is governed by the Allen–Cahn (1972) equation (7), and the evolution of conserved order parameters c_i (e.g., concentrations) is governed by the Cahn–Hilliard (1958) equation (8):

$$\frac{\partial \eta_j}{\partial t} = -L_j \frac{\delta F}{\delta \eta_j}, \qquad\qquad j = 1, \ldots, N_\eta \qquad (7)$$

$$\frac{\partial c_i}{\partial t} = \nabla \cdot \left(M_i \nabla \frac{\delta F}{\delta c_i} \right), \qquad\qquad i = 1, \ldots, N_c \qquad (8)$$

Here, F is the total free energy of the system, which can be formulated as a volume integral

$$F = \int_\Omega \left(f_{\text{loc}} + f_{\text{gr}} + E_{\text{d}} \right) \, \mathrm{d}V, \qquad (9)$$

where Ω is the simulation domain,

$$f_{\text{loc}} \equiv f_{\text{loc}}(\eta_1, \eta_2, \ldots, c_1, c_2, \ldots) \qquad (10)$$

is the local free energy density, and

$$f_{\text{gr}} \equiv f_{\text{gr}}(\nabla \eta_1, \nabla \eta_2, \ldots, \nabla c_1, \nabla c_2, \ldots) \qquad (11)$$

is the gradient energy contribution.

In isolated systems, the evolution of these variables leads to a monotonically decreasing free energy as a function of time. The phase-field method has been used to model a large range of physical phenomena, including solidification, phase transformation, and grain growth.

2   Discretization Techniques

The behavior of physical systems is continuous in both space and time. To make the solution amenable to numerical solution, it is necessary to approximate the continuous response of the actual system using a finite number of variables. This process, known as discretization, allows the model to be solved using an algebraic system of equations. Discretization is used both in space and time – spatial

discretization is used to describe the spatial variation of field quantities, while time discretization is used to solve for the system's response at a fixed number of times.

2.1 Finite Element Method

The finite element method (FEM) is widely used as a discretization technique for solving a variety of PDEs (Zienkiewicz et al. 2013). In FEM, the solution domain is represented by a mesh, which is a collection of connected primitive geometric elements. Associated with these elements is a set of nodes, which are located at the vertices of elements, and can also be located on edges or faces, or within the volume of elements. Continuity of the solution between neighboring elements is enforced when those elements share common nodes.

FEM is used to solve a discretized form of the governing PDEs, which are enforced in a weak, or integral, form at the finite element degrees of freedom. The solution is locally interpolated over individual elements using a set of shape functions, which are associated with the degrees of freedom in the discrete system. A number of different types of shape functions can be used. Lagrangian shape functions are among the most commonly used and interpolate between degrees of freedom directly associated with the nodes of a given element.

The solution field, denoted here as $u(x)$, is interpolated at any point x within the domain by a set of interpolation or shape functions ϕ_i. Lagrangian shape functions are associated with the finite element nodes, so that for node i, $\phi_i = 1$, and at all other nodes, $\phi_i = 0$. At any location within an element, the sum of the shape functions is 1. The solution and its gradient can be expressed at any location in the domain as a function of the shape functions and their gradients:

$$u(x) = \sum_i \phi_i(x) u_i \tag{12}$$

$$\nabla u(x) = \sum_i \nabla \phi_i(x) u_i \tag{13}$$

where u_i are the values of u at the nodes of the element containing point x.

To solve a PDE using FEM, it must be cast in a variational or weak form. This involves pre-multiplying each term of the PDE by a test function and expressing the PDE as a domain integral. In the Galerkin method, these test functions correspond to the shape functions. The divergence theorem is then applied to integrate terms with divergence operators into two separate terms involving a volume and a surface integral and which can be represented using gradients of the shape functions.

Finally, the weak form of the governing equation is integrated numerically. The integrals over the domain are performed by taking a summation of the integrals over individual elements, and the individual elements are integrated using numerical quadrature, typically Gaussian quadrature. An arbitrary function $f(x)$ is integrated numerically over the domain of an element Ω_e as:

$$\int_{\Omega_e} f(\boldsymbol{x}) \approx \sum_q f(\boldsymbol{x}_q) w_q \qquad (14)$$

where q are the quadrature points, $\boldsymbol{x}_q$ are the coordinates of the quadrature points, and w_q are the weights of the quadrature points for an element.

An important side effect of the use of quadrature is that while solution variables can be interpolated anywhere in the domain, coefficients in the PDE that are computed at quadrature points are only available at those locations, and extrapolation to other locations is somewhat problematic.

One of the advantages of FEM is that it is very flexible. It has become widely used for solving PDEs associated with a variety of physics. Finite element meshes can be constructed to conform to the shapes of very general domain geometries, so it can be used for modeling a variety of engineering components. The flexibility in mesh generation also readily permits local refinement in regions where increased fidelity is needed.

2.2 Other Methods

FEM is the most widely adopted technique for multiphysics PDE solvers and is the basis for the methods discussed here. However, a number of other PDE solution techniques can be used for solving multiphysics problems, and a few of them are briefly discussed here.

2.2.1 Finite Difference Method

The finite difference method is a technique for discretizing PDEs in either space or time and was developed independently of FEM. FDM also uses a mesh to represent the solution domain, but the mesh must be topologically regular. It can be adjusted to fit domains with curved surfaces, but it does not have nearly the flexibility that FEM meshes do to conform to parts with complex geometry.

FDM is much simpler than FEM. FDM approximates the spatial derivatives of solution field variables by computing the differences between values of those variables on opposing sides of a given node in each coordinate direction. It does not rely on quadrature or interpolation functions in the way that FEM does.

FDM can be very numerically efficient, and although it is limited in the geometries that it can represent, it is very well-suited to material-scale modeling where the solution domain commonly consists of periodic representative volume elements.

2.2.2 Spectral Methods

Rather than discretizing the domain using degrees of freedom tied to specific material points, as is done in the FEM or FDM, the spatial variation of the solution field can be represented using a Fourier series expansion (Chen and Shen 1998). This technique is widely used in phase-field simulations of periodic representative volume elements and is very efficient.

2.2.3 Peridynamics

Peridynamics is another method for solving governing equations. Peridynamics is a nonlocal meshfree method that solves integral rather than differential equations (Silling 2000). Its development was driven by a need to better handle fracture in solid mechanics simulations. Using PDEs, along with the solution methods developed to solve PDEs, to represent mechanical deformation is problematic when material fractures, because the response is no longer continuous. Peridynamics overcomes some of the shortcomings of FEM in this area by allowing connections between material points to naturally break. It has since been generalized to other physics, and has been applied to coupled physics models, including nuclear fuel applications (Oterkus and Madenci 2017).

3 Solution Techniques

These discretization methods for PDEs all result in the representation of the system as a set of coupled algebraic equations. The integration of these equations in time can be done using explicit methods, in which no equation solution is needed, or implicit methods, which require the solution of a system of linear or nonlinear equations at each time step. Many of the materials challenges in nuclear reactors involve slowly evolving processes, which are more amenable to solution using implicit techniques, so those are the focus of the present discussion.

Typically, Newton's method or a variant thereof is used to solve the nonlinear system of equations for a given physics. The goal of the nonlinear solver is to obtain the solution vector x for which the associated residual r is zero:

$$r(x) = 0 \tag{15}$$

The Jacobian matrix J is:

$$J(x) = \frac{\partial r(x)}{\partial x}. \tag{16}$$

In Newton's method, the following procedure is iteratively repeated to update the solution until Eq. 15 is met within an acceptable tolerance:

$$\text{Compute } J(x_k), r(x_k) \tag{17}$$

$$\text{Solve } J(x_k)s = -r(x_k) \text{ for } s \tag{18}$$

$$x_{k+1} = x_k + s \tag{19}$$

where s is the iterative update to the solution vector and k is the nonlinear iteration index.

Solving coupled physics systems introduces many mathematical and software engineering challenges that are not encountered in solving single physics equations.

For multiphysics simulations, the complexity of solving the system is increased significantly because the solution vector x contains the unknown degrees of freedom for all of the physical systems modeled. To solve the system for all those physics using Newton's method requires forming the full Jacobian matrix for that coupled system. This can be quite challenging, because computing the terms related to the interactions between the physics can be very complex.

To avoid this difficulty, it is common to solve coupled physics problems in a loosely coupled fashion, where the systems of equations for the individual physics are solved independently. The results from the solutions of individual physics models are sequentially transferred to the other models. Fixed point iterations can be performed to iterate between these models until they reach a converged solution. Alternatively, if there is weak coupling between the systems, it may be acceptable to simply transfer the solutions from the other physics from the previous time step and skip these iterations entirely.

This loose coupling approach allows for the individual physics models to be based on completely different solution and discretization schemes. The individual simulation codes and discretizations can be optimized for specific physics, and all that is required is a means to transfer results between the models. This approach is widely used in practice to couple physics codes that were developed independently.

For coupled physical systems with significant feedback, however, loose coupling approaches often suffer from decreased performance relative to tight coupling. A study performed by Novascone et al. (2015) compared the performance of loose and tight coupling approaches for thermal/mechanical systems. It was found that for weakly coupled systems, there were no significant benefits to using a tightly coupled solution strategy, but using tight coupling significantly improved performance for problems with strong two-way feedback between the physics.

The Jacobian-free Newton-Krylov (JFNK) method, which is a variant of the Newton method that does not explicitly require the Jacobian (Knoll and Keyes 2004), has great utility for solving multiphysics problems as a single coupled system of equations since it alleviates some of the burden of forming the coupling terms in the Jacobian matrix. This algorithm employs a Krylov iterative method, such as GMRES (Saad and Schultz 1986), to compute the Newton iterative update in Eq. 18:

$$s = a_0 r_0 + a_1 J r_0 + a_2 J^2 r_0 + \cdots + a_l J^l r_0 \qquad (20)$$

where r_0 is the initial residual and a_i are the series of coefficients computed in the individual Krylov iterations.

The Krylov algorithm only requires the action of the Jacobian on a vector v, which can be approximated with the finite difference expression:

$$J(x_k)v \approx \frac{r(x_k + \epsilon v) - r(x_k)}{\epsilon}. \qquad (21)$$

where ϵ is a small number that reflects a perturbation from the last iterative solution. As long as the Krylov iterations in Eq. 20 are sufficiently converged, this technique

has the same nonlinear convergence characteristics as Newton's method. There are two primary advantages to not explicitly forming a Jacobian: it can require significant computational resources to store and solve using the Jacobian, and it can be challenging to compute the off-diagonal coupling terms in the Jacobian.

Although the full Jacobian matrix is not strictly required in the JFNK algorithm, The Krylov iterations converge much more rapidly if the system is well-conditioned, so preconditioning is typically used with this approach. Using the exact Jacobian matrix for preconditioning will yield best convergence rates for the Krylov iterations, but an approximation of it often performs quite well.

In multiphysics simulations, it is often fairly straightforward to compute the residual for physics as a function of the state of another physics, but computing the derivatives with respect to that other physics can be much more difficult. Hence the JFNK algorithm can be very attractive in that those derivatives are not strictly required. A working initial capability can be rapidly developed without those derivatives, and then they can be developed later as needed to optimize the algorithm.

An alternative approach to deal with the complexity of forming the Jacobian is to use automatic differentiation (AD) techniques to automatically compute the derivatives needed for the Jacobian from the code that computes the residual (Pawlowski et al. 2012a). The residual evaluation code can be broken down into basic mathematical operations with known rules for computing derivatives. AD can be used to compute an exact Jacobian matrix for an entire equation system, which can be applied very effectively to obtain the solution. The main drawback to AD is that it requires sophisticated software engineering. For the derivatives to be evaluable from the code that computes residuals typically requires quite invasive changes to be made to that code.

A number of commercial and open-source codes and frameworks supporting multiphysics simulations are available, based on a variety of designs. A few notable examples, all based on FEM, are mentioned here.

Many commercial codes such as Abaqus (Dassault Systemes 2016), which started out as single-physics solid mechanics codes, now support coupling with multiple physics. COMSOL (COMSOL, Inc 2017) is a commercial product specifically targeted at multiphysics simulation and which permits adding user-defined physics. It has been used as the basis of a nuclear fuel simulation tool (Liu et al. 2016).

Multiple frameworks have been developed with an explicit goal of being the basis for developing multiphysics simulation capabilities, with very different philosophies. A few examples of such frameworks are mentioned here.

The SIERRA framework (Stewart and Edwards 2004) is an example of a software framework based on the use of loose coupling to handle the solution of multiphysics models. It does this by providing common finite element data structures and services to transfer solutions between applications, which solve their individual physics equations independently.

Other frameworks solve multiphysics models using a single coupled system of equations. One such framework is MOOSE (Gaston et al. 2009), which provides a pluggable set of interfaces to define individual physics models, and manages the

FEM assembly of the residuals from those models, using preconditioned JFNK and other techniques to solve the physics models as a single, tightly coupled system. MOOSE is notable in the context of this chapter because its development was motivated specifically for the solution of simulation problems in nuclear energy. Multiple applications based on that framework have been developed for various aspects of nuclear energy simulations and are used as the basis for the results shown here. Other open-source frameworks that solve a single system of equations are Trilinos (Pawlowski et al. 2012b), FEniCS (Logg et al. 2012), and Life (Prud'homme 2006). These also provide pluggable interfaces to define physics models and solve the coupled physics equations in a single tightly coupled system, but they focus on using AD techniques to automate the construction of the Jacobian matrix.

4   Applications

4.1   Fuel Performance

A wide variety of nuclear reactors, with specialized nuclear fuels, have been designed or proposed. These include light water reactors (LWRs), research reactors employing plate fuel, fast reactors, high temperature gas reactors, and others. The discussion here focuses on the fuel performance of UO_2/zirconium alloy fuel rods for LWRs. These rods consist of hundreds of UO_2 pellets atop one another encased in a zirconium alloy cladding with a gas-filled plenum at the top of the cladding.

The fundamental quantity of interest in LWR fuel performance analysis is the fuel center-line temperature. This quantity is critical due to its effect on neutronics and other fuel performance behaviors.

Considering Eq. 6, for steady operation, the important parameters are the volumetric heating rate and the thermal conductivity. The volumetric heating rate directly depends on the fission rate, which is given as in input for fuel performance analysis. The thermal conductivity of the ceramic fuel is quite low, whereas the thermal conductivity of the zirconium alloy cladding is fairly high given that the cladding is a metal. Thus the more important quantity is the fuel thermal conductivity.

Fuel thermal conductivity is temperature dependent (Fink 2000) as well as dependent on irradiation effects (Lucuta et al. 1996). This being the case, the thermal conductivity of the fuel varies across the fuel pellet due to the temperature gradient in the pellet and also varies in time. The heat transfer equation is therefore nonlinear.

The heat in the fuel passes through the gap between the fuel and the cladding before passing through the cladding and then to the coolant. The gap between the fuel and cladding is filled with helium gas when the rod is fabricated. The ease at which heat passes through the gap is governed by the size of the gap; the thermal conductivity of the gas in the gap; pressure due to mechanical contact, if any, between the fuel and the cladding; and radiative heat transfer effects. The size of the gap depends on the initial geometry and the deformation of the fuel

and cladding, which ties the mechanical response of the fuel rod to the thermal response. In addition, one product of nuclear fission is gas (primarily xenon and krypton) that will accumulate in the gap over time. These gaseous fission products affect the internal pressure of the fuel rod (the mechanical response) and the thermal conductivity of the gas mixture (the thermal response).

Conductance across the gap due to the gas follows Eq. 22:

$$h_{\text{gap}} = \frac{k_g(T_g)}{r_c - r_f + d} \tag{22}$$

where h_{gap} is the gap conductance; k_g is the gas thermal conductivity; T_g is the temperature of the gas; r_c and r_f are the radial positions of the cladding and fuel, respectively; and d is an additional distance that accounts for roughness of the surfaces, for example. The need to calculate the distance between the cladding and the fuel clearly shows a dependence on the mechanical solution.

Other important fuel behaviors include densification, swelling, cracking, and creep. Early in life in the reactor, pellets densify and shrink slightly. Fuel swelling happens continuously under irradiation and is of two types. The first is solid fission product swelling and occurs in a linear relationship with irradiation. The second is due to gaseous fission products and is much more difficult to predict. Gaseous swelling depends on the diffusion of gas atoms and bubbles within grains, the accumulation and interconnection of gas bubbles at grain boundaries, and the eventual release of fission gas to the gap and plenum.

When a fuel rod is initially brought to power in the reactor, the temperature gradient causes the pellet to crack due to thermal stresses. This cracking relieves stress and allows the pellet to move outward, slightly, in the radial direction. (This also affects temperature predictions because of the effect on the gap size.) Cracking also greatly influences the creep of the fuel pellet due to its effect on the stress state.

Thermal and irradiation creep, along with irradiation growth, of the zirconium alloy cladding are other behaviors that must be modeled. The coolant pressure in LWRs engages creep mechanisms in cladding, causing the cladding to move inward toward the pellet, affecting internal rod pressure, gap size, and, once again, temperature.

The effects of several sources of eigenstrains must be accounted for in the computation of the stress. For an example of how this is done, the elastic stress computation takes these into account as:

$$\sigma = E(\varepsilon - \varepsilon_T - \varepsilon_{sw} - \varepsilon_d - \ldots) \tag{23}$$

where σ is the stress, E is the elasticity tensor, ε is the total mechanical strain, ε_T is the thermal strain, ε_{sw} is the strain due to swelling, and ε_d is the swelling due to densification. Given that stress is dependent on the thermal strain, the displacement solution has a direct dependence on the temperature solution.

Other behaviors of interest in fuel performance modeling include the shift of power toward the outer rim of the fuel pellet, the thermal boundary condition on the

exterior of the cladding, corrosion of the cladding, thermal expansion, the evolving internal rod pressure, mechanical contact between the fuel and cladding, and mass diffusion.

Most, though not all, fuel performance codes resolve the thermal and mechanical physics with a loose coupling approach. In this approach, the temperature at the grid points is computed based on a fixed set of displacements, and then the displacements and stresses are computed based on those temperatures. The new stress state and displacements are then used to update the temperature, and the process is repeated until a satisfactory condition is met. It is worth noting that this approach can be nested. For example, the displacements may be computed using a fixed internal rod pressure. With new displacements, the internal rod pressure can be recomputed and the two steps repeated to reach a final converged displacement solution.

Fuel performance codes are used to compute fuel center-line temperatures when comparing against experimental data, for postulated conditions, or when coupling with other (e.g., neutronics) codes. There is particular interest in using fuel performance codes to simulate abnormal, or accident, conditions. Because the fuel properties and behavior are outside normal regimes and can exhibit highly nonlinear behavior, this can be particularly challenging for fuel performance analysis.

For example, in a loss of coolant accident, the supply of coolant to the reactor is interrupted. This results in increased fuel rod and coolant temperatures. If coolant flow is not restored, the coolant will change to steam, and the fuel rod temperatures will continue to rise. A lack of external pressure in conjunction with increased temperatures can cause the cladding to creep outward. This outward motion may result in a ballooned section of the fuel rod and in cladding burst.

This situation requires more advanced thermal boundary conditions for the fuel performance code (often taken from the output of a dedicated thermal-hydraulics package). In addition, creep models for the cladding appropriate for this temperature regime are needed. Fission gas release and other fuel-specific models must also be appropriate for these conditions.

Other challenging conditions to analyze, all of which require a multiphysics approach, include reactivity insertion accidents, so-called missing pellet surface defects (a chipped pellet that creates high localized stresses), and stress corrosion cracking of cladding.

Examples of fuel performance codes in use today include TRANSURANUS (Lassmann 1992), ENIGMA (Rossiter 2011), DIONISIO (Soba and Denis 2008), ALCYONE (Michel et al. 2013), FRAPCON (Berna et al. 1997), FALCON (Yagnik et al. 2004), and Bison (Williamson et al. 2012). Bison is based on the MOOSE framework discussed in Sect. 3.

Example calculation results for layered 1D (stacked radial slices of fuel and cladding), 2D axisymmetric, and 3D fuel analysis are shown in Fig. 1. Low dimensionality models are useful for rapid evaluation of the global fuel rod response, while higher dimensionality models can be used to study local behavior in detail. These models were all run using the Bison code, which uses the same set of computational models with arbitrary dimensionality.
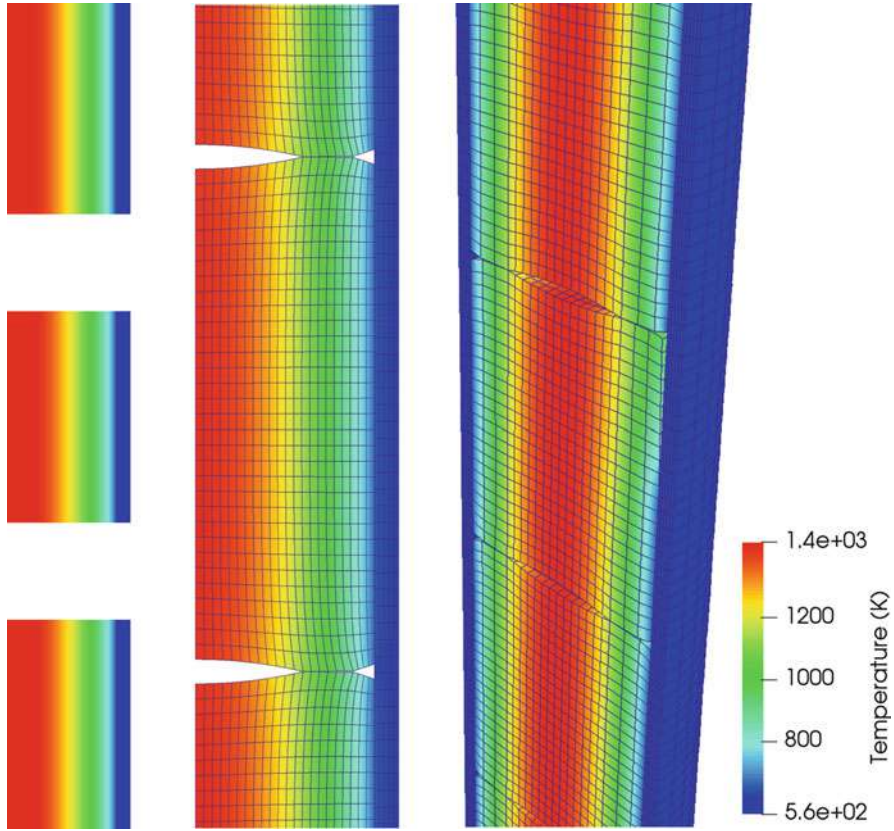
Fig. 1 1D (left), 2D (center), and 3D (right) nuclear fuel models. The 1D model is a set of slices of fuel with radial 1D elements representing fuel and cladding at a given elevation. The 2D model uses an axisymmetric representation and can include individual fuel pellets as shown. The 3D model uses a symmetry plane through the center of the rod and can be used to study local effects such as the behavior around a defective pellet on upper right side of center pellet

4.2   Hydride Formation

Zirconium (Zr) alloys are a well-established light water reactor fuel cladding material due to the low neutron capture cross section of zirconium, its mechanical strength, and good corrosion resistance up to high temperatures.

Fuel elements are in constant contact with the water coolant medium during both reactor operation and wet storage, and one of the main challenges facing these alloys is that Zr has an undesirable reaction with water. At the clad surface, H_2O molecules are split and zirconium oxides are formed. A portion of the remaining hydrogen (H) atoms diffuse through the oxide layer forming on that surface into the bulk of the

cladding. Continued H uptake pushes the material past the H solubility limit, at which point a phase change to a thermodynamically more stable hydride phases is initiated.

Hydride formation has a substantial impact on the engineering-scale mechanical properties of the clad material. The low fracture toughness of the ZrH phases causes clad embrittlement and can lead to clad failure due to fracture.

This solid-state diffusion and phase change reaction lends itself to modeling using the phase-field method. Required parameterization includes the bulk free energies of the various ZrH phases, the interfacial free energy between phases, and the kinetic coefficients, i.e., interfacial and species mobilities. Data for this can be found in the literature.

The multiphysics aspects of this problem arise from the fact that the hydride phase transformations result in lattice changes that exhibit volume changes and thus misfit strains with respect to the Zr matrix. These strains result in elastic stress fields, which in turn affect the stored mechanical energy of the system. This mechanical energy contributes to the total free energy of the system, the chemical potentials of the diffusing species, and as a result, the microstructure evolution. Further complications can arise from local plastic deformations due to dislocations emitted at points of stress concentrations at developing hydride precipitates.

The mechanism of delayed hydride cracking is a diffusion-limited time-dependent crack formation mechanism in which ZrH platelets are constantly formed at the tip of a crack. The crack propagates in steps through the newly formed hydride as soon as a critical condition, relating to hydride size and stress intensity factors, is met. Inherent to this mechanism is the multiphysics coupling of diffusive species transport, phase transformation, and mechanical state changes.

A comprehensive model for ZrH formation needs to combine the chemical free energy contributions and the mechanical interactions of the various phases. The resulting coupling in practice is twofold. On the one hand, the evolving phase-field microstructure variables will affect the mechanical properties of the system by means of variable-dependent eigenstrains as well as variable-dependent elasticity tensors. On the other hand, the elastic stress state gives rise to an elastic energy $\frac{1}{2}\sigma\epsilon$ which needs to be added to the thermodynamic free energy density used in the phase-field evolution equations. The elastic free energy is a function of a subset of the phase-field order parameters though the mechanical properties, which depend on the microstructural state. As a consequence, driving forces for these phase-field order parameters emerge from the variational derivative of this free energy contribution.

The phase-field and mechanics equations can be solved implicitly in a tightly coupled system using the finite element method. Figure 2 shows a representative coupled phase-field and mechanical finite element simulation of hydride precipitation. In periodic systems under the small-strain assumption, spectral methods can solve this system effectively as well.

Fig. 2 Stress field and concentration contours of three differently oriented γ hydride precipitates
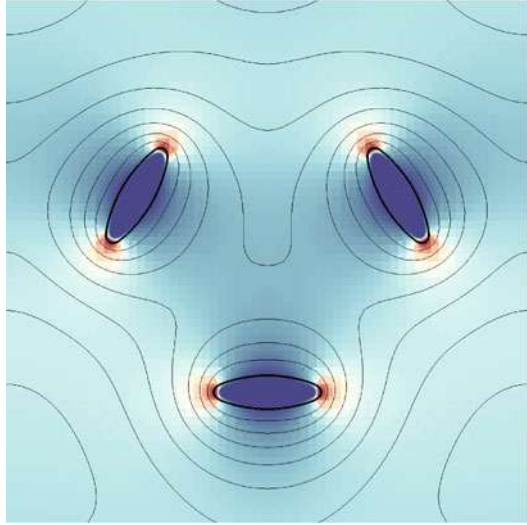
4.3   UO$_2$ Microstructure Evolution

Uranium dioxide (UO$_2$), a common light water reactor fuel, is a polycrystalline ceramic material. For reactor applications, the fuel is synthesized in powder form and pressed into cylindrical pellets with a diameter and height of about a centimeter. These pellets are then sintered under a reducing atmosphere and loaded into zirconium alloy cladding tubes. Under operation conditions in a critical nuclear reactor, fission of the ^{235}U nuclei generates fission fragments, neutrons (which keep up the fission reaction), prompt γ-rays, and, through decaying fission products, β^--particles, delayed γ-rays, and heat due to neutron capture. As a consequence the nuclear fuel heats up, incurs lattice damage from collision cascades, and is doped with a broad variety of new chemical species – the fission fragments. About 80% of the heat is deposited during the stopping process of the fission fragments, and roughly 15% of the fission products deposited in the fuel are insoluble gases.

These harsh conditions result in driving forces that strongly impact the microstructural evolution of the fuel. The techniques described in Sect. 4.1 are necessary to compute the temperature and stress conditions in the fuel. Thermal and stress gradients affect the migration of both chemical species and irradiation-generated point defects and defect clusters in the fuel. This requires solving a diffusional or phase-field problem.

Beyond these basic mechanisms, which already form a connected multiphysics problem, are the advanced mechanisms of irradiation effects, added chemical species, creep, and fracture. The constant internal irradiation generates a defect population far above the thermal equilibrium. Defects interact with each other to
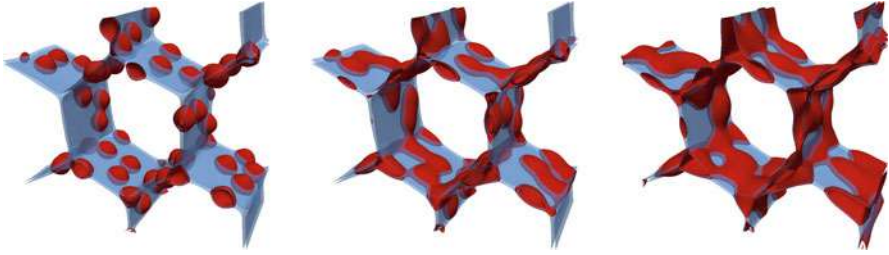
Fig. 3 3D simulation of increasing fission gas grain boundary coverage with progressing burnup in UO_2 (From Aagesen et al. 2017)

form complex defect structures such as loops or voids, and they interact with solute species such as fission gas atoms to form trap sites or gas bubbles. The irradiation also forces chemical mixing through ballistic displacements of atoms leading, for example, to resolution of fission gas out of the precipitated bubbles. Chemical species intentionally added to the fuel affect its composition and must be taken into account. Figure 3 shows a simulation of the evolution of fission gas bubbles at grain boundaries.

At the mesoscale level, radiation effects can be simulated with varying levels of fidelity. If the primary concern is the non-equilibrium defect population, phase-field simulations can include source terms for the vacancy and interstitial fields. These source terms can be homogeneous, to represent mean field defect production, empirical distributions fitted to typical collision cascade dimensions, or data obtained from a coupled damage calculation code. The most established code for ion transport in matter and the generation of point defects is the binary collision Monte Carlo (BCMC) code TRIM/SRIM (Ziegler et al. 2010), which is based on a scattering integral approximation of the Ziegler-Biersack-Littmark potential. BCMC data can also be used to introduce ballistic mixing contributions to a phase-field microstructure evolution simulation.

At the grain level, recrystallization can occur in nuclear fuel as the existing grains acquire a free energy contribution from stored defect populations, such as dislocations. In absence of a fast mechanism to relieve the grains of the stored defects by annihilation or diffusion into sinks, new, defect-free grains with a lower free energy may spontaneously nucleate and form small-grained microstructures. This process can be captured by phase-field simulations.

4.4   Summary

Many aspects of the behavior of components and materials in the nuclear reactor environment are fundamentally affected by phenomena involving the coupled response of multiple physical systems. Including this coupled multiphysics behavior in simulation models is essential for accurately representing many problems of engineering significance.

This chapter provides an overview of the main governing equations involved, the techniques for discretization and solution of these equations, and the software algorithms, frameworks, and codes used for solving multiphysics problems. It also provides an overview of important aspects of a few areas of interest in multiphysics simulation of nuclear reactor components and materials. This is a broad and constantly developing field. The algorithms, codes, and applications described here are not intended to be a comprehensive listing, but are representative of the current state of the art.

References

Aagesen L, Schwen D, Zhang Y (2017) Microstructure-level modeling of stage 3 fission gas release in UO$_2$ fuel. Technical report INL/EXT-17-43374 Rev. 0, Idaho National Laboratory

Allen SM, Cahn JW (1972) Ground state structures in ordered binary alloys with second neighbor interactions. Acta Metall 20(3):423–433. https://doi.org/10.1016/0001-6160(72)90037-5

Bažant ZP, Chern JC, Thonguthai W (1981) Finite element program for moisture and heat transfer in heated concrete. Nucl Eng Des 68(1):61–70. https://doi.org/10.1016/0029-5493(82)90040-1

Berna GA, Beyer GA, Davis KL, Lanning DD (1997) FRAPCON-3: a computer code for the calculation of steady-state, thermal-mechanical behavior of oxide fuel rods for high burnup. Technical report NUREG/CR–6534-Vol.2; PNNL–11513-Vol.2, Nuclear Regulatory Commission, Washington, DC

Cahn JW, Hilliard JE (1958) Free energy of a nonuniform system. I. Interfacial free energy. J Chem Phys 28(2):258–267. https://doi.org/10.1063/1.1744102

Chen L, Shen J (1998) Applications of semi-implicit Fourier-spectral method to phase field equations. Comput Phys Commun 108(2–3):147–158. https://doi.org/10.1016/S0010-4655(97)00115-X

COMSOL, Inc (2017) COMSOL multiphysics reference manual, version 5.3. www.comsol.com

Dassault Systemes D (2016) Abaqus unified FEA user documentation. www.3ds.com

Fick A (1855) On liquid diffusion. Poggendorffs Annalen 94:59. https://doi.org/10.1016/0376-7388(94)00230-V, reprinted in "On liquid diffusion". J Membr Sci 100:33–38, 1995

Fink JK (2000) Thermophysical properties of uranium dioxide. J Nucl Mater 279(1):1–18

Gaston D, Newman C, Hansen G, Lebrun-Grandié D (2009) MOOSE: a parallel computational framework for coupled systems of nonlinear equations. Nucl Eng Des 239(10):1768–1778. https://doi.org/10.1016/j.nucengdes.2009.05.021

Knoll D, Keyes D (2004) Jacobian-free Newton–Krylov methods: a survey of approaches and applications. J Comput Phys 193(2):357–397. https://doi.org/10.1016/j.jcp.2003.08.010

Lassmann K (1992) TRANSURANUS: a fuel rod analysis code ready for use. J Nulc Mater 188:295–302

Liu R, Prudil A, Zhou W, Chan PK (2016) Multiphysics coupled modeling of light water reactor fuel performance. Prog Nucl Energy 91:38–48. https://doi.org/10.1016/j.pnucene.2016.03.030

Logg A, Mardal KA, Wells G (eds) (2012) Automated solution of differential equations by the finite element method. Lecture notes in computational science and engineering, vol 84. Springer, Berlin/Heidelberg. https://doi.org/10.1007/978-3-642-23099-8

Lucuta PG, Matzke HJ, Hastings IJ (1996) A pragmatic approach to modelling thermal conductiv-
ity of irradiated UO$_2$ fuel: review and recommendations. J Nucl Mater 232:166–180

Michel B, Nonon C, Sercombe J, Michel F, Marelle V (2013) Simulation of pellet-cladding
interaction with the PLEIADES fuel performance software environment. Nucl Technol 182(2):
124–137. https://doi.org/10.13182/NT13-A16424

Novascone S, Spencer B, Hales J, Williamson R (2015) Evaluation of coupling approaches
for thermomechanical simulations. Nucl Eng Design 295:910–921. https://doi.org/10.1016/j.
nucengdes.2015.07.005

Olander D (1976) Fundamental aspects of nuclear reactor fuel elements. Technical report TID-
26711-P1, Technical Information Center, Office of Public Affairs, Energy Research and
Development Administration. https://doi.org/10.2172/7343826

Oterkus S, Madenci E (2017) Peridynamic modeling of fuel pellet cracking. Eng Fract Mech.
https://doi.org/10.1016/j.engfracmech.2017.02.014

Pawlowski RP, Phipps ET, Salinger AG (2012a) Automating embedded analysis capabilities
and managing software complexity in multiphysics simulation, part I: template-based generic
programming. Sci Program (2):197–219

Pawlowski RP, Phipps ET, Salinger AG, Owen SJ, Siefert CM, Staten ML (2012b) Automating
embedded analysis capabilities and managing software complexity in multiphysics simulation,
part II: application to partial differential equations. Sci Program (3):327–345. https://doi.org/10.
3233/SPR-2012-0351

Prud'homme C (2006) A domain specific embedded language in C++ for automatic differentiation,
projection, integration and variational formulations. Sci Program 14(2):81–110. https://doi.org/
10.1155/2006/150736

Rahman M, Saghir M (2014) Thermodiffusion or Soret effect: historical review. Int J Heat Mass
Trans 73:693–705. https://doi.org/10.1016/j.ijheatmasstransfer.2014.02.057

Rashid MM (1993) Incremental kinematics for finite element applications. Int J Numer Methods
Eng 36(23):3937–3956. https://doi.org/10.1002/nme.1620362302

Rossiter G (2011) Development of the ENIGMA fuel performance code for whole core analysis
and dry storage assessments. Nucl Eng Technol 43(6):489–498

Saad Y, Schultz MH (1986) GMRES: a generalized minimal residual algorithm for solving linear
systems. SIAM J Sci Statist Comput 7:856

Silling S (2000) Reformulation of elasticity theory for discontinuities and long-range forces.
J Mech Phys Solids 48(1):175–209. https://doi.org/10.1016/S0022-5096(99)00029-0

Simo JC, Hughes TJR (1998) Computational inelasticity. Interdisciplinary applied mathematics,
vol. 7. Springer, New York

Soba A, Denis A (2008) Simulation with DIONISIO 1.0 of thermal and mechanical pellet-cladding
interaction in nuclear fuel rods. J Nucl Mater 374(1):32–43. https://doi.org/10.1016/j.jnucmat.
2007.06.020

Stafford D (2015) Multidimensional simulations of hydrides during fuel rod lifecycle. J Nucl Mater
466:362–372. https://doi.org/10.1016/j.jnucmat.2015.06.037

Stewart JR, Edwards H (2004) A framework approach for developing parallel adaptive multi-
physics applications. Finite Elem Anal Des 40(12):1599–1617. https://doi.org/10.1016/j.finel.
2003.10.006

Williamson RL, Hales JD, Novascone SR, Tonks MR, Gaston DR, Permann CJ, Andrs D,
Martineau RC (2012) Multidimensional multiphysics simulation of nuclear fuel behavior.
J Nuclear Mater 423:149–163. https://doi.org/10.1016/j.jnucmat.2012.01.012

Yagnik S, Rashid Y, Dunham R, Montgomery R (2004) Fuel analysis and licensing code: FALCON
MOD01 – volume 1: theoretical and numerical bases. Technical report EPRI-1011307

Ziegler JF, Ziegler M, Biersack J (2010) SRIM: the stopping and range of ions in matter
(2010). Nucl Instrum Methods Phys Res Sect B Beam Interactions Mater Atoms 268(11–12):
1818–1823. https://doi.org/10.1016/j.nimb.2010.02.091

Zienkiewicz OC, Taylor RL, Zhu JZ (2013) The finite element method: its basis and fundamentals,
7th edn. Elsevier/Butterworth-Heinemann, Amsterdam

Zinkle SJ, Was GS (2013) Materials challenges in nuclear energy. Acta Mater 61(3):735–758.
https://doi.org/10.1016/j.actamat.2012.11.004

Phase-Field Modeling of Microstructure Evolution in Nuclear Materials

94

Karim Ahmed and Anter El-Azab

Contents

Abstract

Irradiation is a unique driver for microstructure evolution in nuclear materials. It produces high concentrations of point defects and extended defects that trigger the formation of microstructural features such as voids, bubbles, precipitates and dislocations agglomerates. The evolution of such features significantly impacts the properties of materials, thus requiring careful modeling. The phase-field method has demonstrated its capability of simulating microstructure evolution in nuclear materials. Although the method does not yield as of yet quantitative results in various aspects of microstructure evolution, the community is working with the premise that the method will be a standard predictive tool in the future. The objective of this chapter is thus to present a concise summary of the status of

K. Ahmed (✉)
Department of Nuclear Engineering, Texas A&M University, College Station, TX, USA
e-mail: karim.ahmed@tamu.edu

A. El-Azab
School of Materials Engineering, Purdue University, West Lafayette, IN, USA
e-mail: aelazab@purdue.edu

development of the phase field approach for nuclear materials applications, with a special focus on the quantitative results obtained by following this approach. The strengths and limitations of the application of phase-field modeling of microstructure evolution in nuclear materials are discussed, and a summary of possible future research directions is presented.

1   Introduction

Nuclear reactor materials are often exposed to harsh conditions during service. In addition to high temperatures and stresses in corrosive environments, these materials are constantly bombarded with different types of energetic particles. The interactions of such particles with materials produce high concentrations of defects including vacancies, interstitials, and small defect clusters. These defects eventually nucleate extended defects such as dislocation loops, voids, bubbles, and solid-phase precipitates and lead to evolution of grain boundaries. All of these microstructural features drastically change the thermal and mechanical properties of materials and hence limit their performance (Olander 1976; Short and Yip 2015; Was 2017; Olander and Motta 2017). Hence, understanding how microstructure evolution takes place in nuclear materials is of paramount importance.

Due to the inherent multi-physics and multi-scale nature of radiation effects in materials, several computational modeling techniques were employed to simulate the resulting microstructure processes (Was 2017; Short and Yip 2015; Olander and Motta 2017). These techniques include density functional theory and molecular dynamics at atomic scale, kinetic Monte Carlo and cluster dynamics at intermediate mesoscale, and phase-field and sharp-interface models at the microstructure scale, to homogenization models such as rate theory. All of these models were utilized to investigate different aspects of the complex problem.

While rate theory and cluster dynamics models have a long history in modeling irradiation effects in materials at mesoscale (Brailsford and Bullough 1972; Was 2017; Olander and Motta 2017), the phase-field method is a relatively newcomer to the field (Millet and Tonks 2011; El-Azab et al. 2014; Y Li et al. 2017). Nonetheless, its unparalleled success in simulating microstructure evolution in all types of materials in other environments made its use in the field of radiation effects inevitable. The phase-field method has clear major advantages that render it uniquely suitable for simulating irradiation-induced microstructure evolution. It can account for several species and microstructural features simultaneously, it has both spatial and temporal resolution, it can account for different driving forces concurrently, it can account for all stages of phase transition from separation/nucleation to growth to coarsening, and it is relatively easy to implement numerically even for complex microstructures in 3D. Indeed, the method was able to successfully simulate different microstructural evolution processes in irradiated materials and their effect on the degradation of physical properties of such materials (Millet and Tonks 2011; El-Azab et al. 2014; Y Li et al. 2017). This success resulted in the sponsorship of the Nuclear Energy Advanced Modeling and Simulation

(NEAMS) program funded by the Department of Energy (DOE) to develop the phase-field-based mesoscale simulator MARMOT (Tonks et al. 2012) to facilitate the implementation of new models for existing and advanced fuel and cladding materials. MARMOT results will be integrated into the fuel performance code BISON (Williamson et al. 2012) to enhance its predictive capability by providing physics-based constitutive laws instead of frequently used empirical relations.

In this communication, we give an overview of the application of the phase-field method to simulate irradiation effects in nuclear materials at the mesoscale. We discuss its advantages and shortcomings and propose ways to mitigate its limitations. In addition, we highlight possible future progress areas. The organization of this manuscript is as follows. Following a brief introduction of the phase-field method in Sect. 1, we summarize in Sect. 2 the current commercial and open-source frameworks available for solving the phase-field kinetic evolution equations. In Sect. 3, we discuss the application of the phase-field modeling technique to nuclear materials, focusing on problems where the method yields quantitative results. Lastly, concluding remarks and suggestions for future research directions are presented in Sect. 4.

2   Background to the Phase-Field Modeling Approach

Phase-field modeling is now recognized to be the approach of choice for modeling microstructure evolution in materials (Boettinger et al. 2002; Chen 2002; Moelans et al. 2008; Emmerich 2008; Provatas and Elder 2010; Ahmed 2011; Ahmed 2015; Biner 2017). Its unique feature is the treatment of interfaces in heterogeneous systems as diffuse. The material properties of different phases, which are represented in terms of phase-field variables, then change continuously across the interface. Such treatment obviates the necessity of tracking the interfaces, as opposed to the classical sharp-interface models that give rise to moving/free boundary problems in which the interface tracking is necessary. Not having to track interfaces in the phase-field approach facilitates the numerical implementation even in cases where complex 3D geometries are involved. This clear advantage of the method is the primary reason for its widespread adaptation in the field of materials science and engineering. Moreover, the technique is now finding ever-expanding applications in other fields such as earth sciences, biology, and medicine (Provatas and Elder 2010; Biner 2017).

Historical developments of the diffuse-interface concepts can be traced back to the works of Van der Waals on gas condensation (Rowlinson 1979), Landau and Ginzburg on phase transitions (Ginzburg and Landau 1950), and Cahn and Hilliard on the thermodynamics of heterogeneous systems (Cahn and Hilliard 1958). In all of these treatments, suitable phase-field variables (or order parameters) are identified to simulate the material and phase transition of interest. Thus, the phase-field variables can be thought of as an extension of the state variables concept in classical thermodynamics.

A typical procedure for constructing a phase-field model is as follows. First, the free energy of the heterogeneous system is expressed in terms of the phase-field variables and their gradients. If one considers only one conserved variable (c) such as concentration and one non-conserved variable (η) such as spin direction or grain orientation, such model usually referred to in literature as model C in analogy to the stochastic models of dynamic critical phenomena (Hohenberg and Halperin 1977), the total isothermal free energy functional becomes:

$$F = \int \left[f\left(c(r), \eta(r)\right) + \frac{\kappa_c}{2}|\nabla c(r)|^2 + \frac{\kappa_\eta}{2}|\nabla\eta(r)|^2 + E\left(c(r), \eta(r)\right) \right] d^3r. \tag{1}$$

Here, $f(c, \eta)$ is the short-range chemical free energy density, and $E(c, \eta)$ is a long-range energy contribution from elastic or electrostatic forces. The chemical free energy can be obtained from thermodynamic databases such as CALPHAD (Moelans et al. 2008). The elastic or electrostatic energies have their regular forms where here the material properties such elastic/dielectric constant or magnetization may be phase-dependent. The gradient terms account for the interfacial energies. The evolution equations for the phase-field variables are then derived following the standard procedure of irreversible thermodynamics (De Groot and Mazur 1962), i.e., the evolution equations are constructed so as to monotonically decrease the free energy while ensuring nonnegative entropy production. A non-conserved order parameter evolves according to the Allen-Cahn/Ginzburg-Landau equation (Allen and Cahn 1979):

$$\partial_t \eta = -L \left[\partial_\eta f + \partial_\eta E - \kappa_\eta \nabla^2 \eta \right], \tag{2}$$

where L is Allen-Cahn mobility. The evolution of the conserved order parameter is governed by Cahn-Hilliard equation (Cahn 1961):

$$\partial_t c = \nabla \cdot M \nabla \mu, \tag{3}$$

$$\mu = \partial_c f + \partial_c E - \kappa_c \nabla^2 c. \tag{4}$$

In Eq. (3), M is the chemical mobility and μ is the chemical potential. If only non-conserved or conserved order parameters and kinetics are present, the phase-field models are called models of type A or B, respectively. If elastic and electrostatic energies are taken into account, their balance laws must be solved simultaneously with Eqs. (2) and (3). The kinetic evolution equations can also be derived from integrating some approximate master equations. In such scenario, the phase-field models are considered the coarse-grained/continuum version of the discrete kinetic Ising models (Glauber 1963, Kawasaki 1966).

In this general phase-field model C type (sometimes called WBM model after Wheeler, Boettinger, and McFadden work on solidification (1992)), the interfacial

energy and interface width can be estimated from the model parameters as (Cahn and Hilliard 1958, Wheeler et al. 1992, Moelans et al. 2008, Ahmed et al. 2013):

$$\gamma \approx \sqrt{f^* \left(\overline{c}, \overline{\eta} \right)} \sqrt{g \left(\kappa_c, \kappa_\eta \right)}, \tag{5}$$

$$\ell \approx \sqrt{\frac{g \left(\kappa_c, \kappa_\eta \right)}{f^* \left(\overline{c}, \overline{\eta} \right)}}. \tag{6}$$

Here is $f^* \left(\overline{c}, \overline{\eta} \right)$ the value of the bulk free energy calculated at the middle of the interface where the phase-field variables take on intermediate values between the two equilibrium values representing the two phases, and $g(\kappa_c, \kappa_\eta)$is some linear function of the gradient coefficients. These simple relations, however, highlight a major limitation of this formulation, which is the coupling between interfacial and bulk properties. The two forces cannot be independently specified. Also, the numerical value of the interface width for realistic values of bulk chemical energy is on the order of nanometer rendering large-scale simulations unfeasible. Nonetheless, other phase-field formulations that overcome this problem were proposed. The most notable ones are the Kim-Kim-Suzuki (KKS) model (Kim et al. 1999) and the grand potential-based phase-field model by Plapp (2011). These two versions are now widely utilized to conduct quantitative phase-field simulations. The phase-field model summarized above is for a simple binary two-phase system. Nevertheless, multiphase and multicomponent phase-field models have also appeared in literature (Moelans 2011; Steinbach 2009).

The diffuse-interface (phase-field) models can also be directly compared to the classical sharp-interface models (Pego 1989; Elder et al. 2001; Emmerich 2008; Provatas and Elder 2010; Ahmed and El-Azab 2018). Such connection is important for two main reasons. First, it demonstrates the thermodynamic and kinetic consistency of new phase-field models. Second, it establishes a direct relation between the phase-field model parameters and the regular thermodynamic and kinetic material properties that appear in the sharp-interface models, which is crucial for enabling quantitative phase-field simulations. The sharp-interface limits of phase-field models can be deduced using formal asymptotic analyses based on singular perturbation theory (Fife 1992).

3   Numerical Implementation of Phase-Field Models

In order to track the microstructure evolution using the phase-field approach, the phase-field kinetic equations, which are partial differential equations (PDEs), must be solved numerically. Most of the well-established methods for solving PDEs such as finite-difference, finite-volume, finite-element, and spectral methods were used for obtaining solutions of such models (Provatas and Elder 2010; Biner 2017). Each method has its strengths and limitations. Initially, the community opted for using their in-house codes written in Fortran90, C++, Python, etc., but recently the

utilization of commercial and open-source software for solving PDEs has gained momentum. This new direction has several advantages such as minimizing code duplication, reducing time and effort of implementing new models, decreasing the time required for obtaining solutions, and increasing the precision and accuracy of these solutions, allowing modelers to focus more on developing physics-based models rather than numerical algorithms and facilitating multi-physics and multi-scale coupling.

Examples of popular codes adopted by the community are the following. For commercial codes, the Microstructure Evolution Simulation Software (MICRESS 2017) and COMSOL (COMSOL 2017) are the most used. For open-source codes, the list is growing rapidly; the ones readily available for download are MOOSE (MOOSE 2017), FEniCS (FEniCS 2017), PRISMS (PRISMS 2017), FiPY (FiPy 2017), and OpenPhase (OpenPhase 2017).

The phase-field community seems more inclined to utilize the open-source codes for two main reasons. First, of course, they are free to use. Second, in contrast to commercial software that are usually black box, they provide complete access to the source code. Such access makes these codes extensible and encourages collaborative work. Moreover, this facilitates code verification and validation. In fact, the community has recently established a set of benchmark phase-field problems to test existing and emerging codes against. This was part of an effort by the Center for Hierarchical Materials Design (CHiMaD 2017) that aims to accelerate materials discovery through integration of theory and computational methods. Representatives of most of the developers of the abovementioned codes were invited to participate in workshops and hackathons to solve these problems and upload the solutions. The benchmark problems and their solutions are hosted on the Phase Field Community Hub website (PF Hub 2017) and summarized in the publication by Jokisaari et al. (2017).

Out of the open-source codes listed before, the one utilized the most by the nuclear materials community is the Multi-Physics Object-Oriented Simulation Environment (MOOSE) developed by Idaho National Laboratory (MOOSE 2017; Gaston et al. 2009). This is due to the fact that MOOSE framework is the building block of the new nuclear fuel performance codes MARMOT (Tonks et al. 2012) and BISON (Williamson et al. 2012). These codes are part of a toolkit that is being developed by the Nuclear Energy Advanced Modeling and Simulation (NEAMS) program funded by the Department of Energy (DOE) to accelerate the development and deployment of new nuclear technologies (NEAMS 2017). In addition, MOOSE has several features that make it appealing to different types of researchers. It facilitates solving multi-physics and multi-scale problems in a fully coupled or weakly coupled approach. Explicit and fully implicit schemes can be used. It provides built-in mesh and time adaptivity. It is dimension independent and readily parallel.

MOOSE has a convenient modular structure that leverages existing codes and simplifies implementation of new models. A schematic of the main structure of MOOSE is shown in Fig. 1. It provides a high-level interface to the finite-element library LibMesh developed by the University of Texas at Austin (LibMesh 2017)
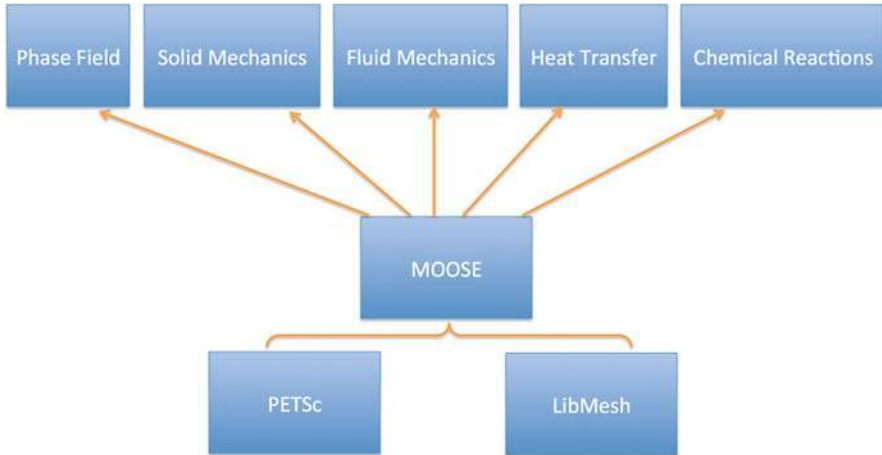
Fig. 1 A schematic illustration of the main structure of the Multi-Physics Object-Oriented Simulation Environment (MOOSE) developed by Idaho National Laboratory (MOOSE 2017). MOOSE is built on top of the finite-element library LibMesh (LibMesh 2017) and the scientific computing toolkit PETSc (PETSc 2017). MOOSE contains modules designed for solving different types of physics, where each module includes several kernels representing the most frequently used models, boundary conditions, initial conditions, etc. in that field. Note that we only highlighted the main modules here; MOOSE has more modules with potential for even further extension. There is also a combined module where all the kernels from different modules can be accessed and multi-physics simulations can be conducted

and the nonlinear solvers in the scientific computing toolkit PETSc developed by Argonne National Laboratory (PETSc 2017). As can be seen from Fig. 1, at its highest level, it consists of different modules targeting different types of physics such as solid mechanics, heat transfer, fluid mechanics, chemical reactions, and phase-field. Each module contains several kernels representing terms that appear in the frequently used PDEs in that subject area. For example, the phase-field module includes the weak forms of Allen-Cahn (Allen and Cahn 1979) and Cahn-Hilliard (Cahn 1961) equations:

$$(\partial \eta / \partial t, \phi) + L\,(\partial f / \partial \eta, \phi) + L\left(\kappa_\eta \nabla \eta, \nabla \phi\right) - <\kappa_\eta \nabla \eta \cdot \mathbf{n}, \phi> = 0, \qquad (7)$$

$$(\mu, \phi) - (\partial f / \partial c, \phi) - (\kappa_c \nabla c, \nabla \phi) + <\kappa_c \nabla c \cdot \mathbf{n}, \phi> = 0, \qquad (8)$$

$$(\partial c / \partial t, \phi) + (M \nabla \mu, \nabla \phi)\ - <M \nabla \mu \cdot \mathbf{n}, \phi> = 0. \qquad (9)$$

In the above, ϕ is a test function, $(\cdot, \cdot)$ stands for interior integration, and $< \cdot, \cdot >$ stands for boundary integration (Tonks et al. 2012; Ahmed et al. 2017). Here, the fourth-order Cahn-Hilliard equation is transformed into two coupled second-order equations which was shown to be a good compromise between accuracy and efficiency (Zhang et al. 2013). The user can then simply code in a specific free

energy form and its derivatives or even take advantage of the newly developed automatic differentiation capabilities Schwen et al. 2017). Of course the user can also call different modules at the same time (from the same input file) to enable multi-physics simulations.

In addition to being a finite-element framework, MOOSE is also an application programming interface (API). It assists developers in launching their own applications that can take advantage of all capabilities present in MOOSE. These applications could be designed to solve advanced specific engineering or physics problems. This unique feature of MOOSE was added by design to facilitate the development of various codes for simulating all aspects of nuclear reactors operation, which is the ultimate goal of the NEAMS program (NEAMS 2017). Based on that philosophy, the MARMOT-MOOSE-BISON framework is currently being utilized to investigate fuel performance under both normal and accidents conditions (Gaston et al. 2015; Tonks et al. 2017). A schematic illustration of that framework is depicted in Fig. 2. In this framework, both MARMOT and BISON are MOOSE-based applications used to simulate fuel and cladding materials response to radiation, stresses, and temperature gradients in the entire fuel rod with spatial resolution down to the size of nanometer bubbles. MARMOT is a phase-field-based code with optional coupling to other thermal, mechanical, and irradiation
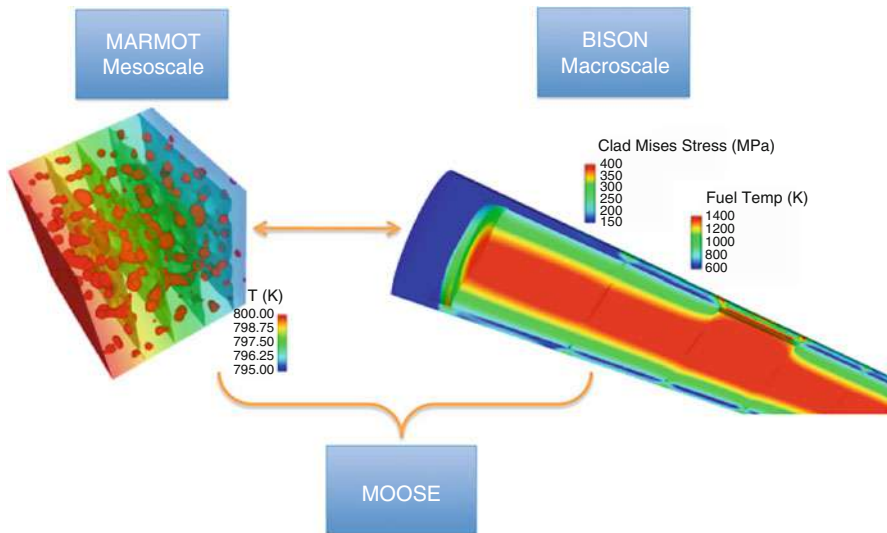
Fig. 2 An illustration of the concept of multi-scale modeling of nuclear fuel performance using MOOSE. The phase-field-based mesoscale simulator MARMOT (MARMOT 2012) and the finite-element-based engineering/macroscale simulator BISON (BISON 2012) are used to study fuel behavior under normal and transient conditions across multiple length and time scales. MARMOT is used to evolve the microstructure and quantify its features such as grains and bubbles sizes and distributions, and BISON calculates the global stress and temperature in the fuel rod. Since both codes are MOOSE-based applications, transferring data back and forth can be done in an efficient and a straightforward manner (Gaston et al. 2015). Furthermore, multi-scale simulations can be conducted in a hierarchal or concurrent setting

driving forces (Tonks et al. 2012). BISON is a finite-element-based code that solves the regular mass, momentum, and heat balance equations to find the global stress and temperature profiles (Williamson et al. 2012). Therefore, the coupling between MARMOT and BISON helps in improving the understanding of the unique phenomena that take place in nuclear materials that usually span multiple length and time scales. Furthermore, there are other MOOSE-based applications concerned with other types of physics and aspects of nuclear power plants, the mention of which is beyond the scope of this chapter.

4   Applications of Phase-Field Models to Nuclear Materials

Over the last decade, the phase-field modeling approach has been used to investigate irradiation effects in materials (Millet and Tonks 2011; El-Azab et al. 2014; Y Li et al. 2017). Studies of voids and bubbles formation and growth (Yu and Lu 2005; Hu et al. 2009; Hu and Henager 2009; Semenov and Woo 2012; Rokkam et al. 2009; Millett et al. 2009, 2011a, b; Li et al. 2013; Xiao et al. 2013), voids migration under temperature gradient (Hu and Henager 2010; Li et al. 2010; Zhang et al. 2012), bubble/void superlattice development (Yu and Lu 2005; Hu et al. 2016a, 2016b), growth of dislocation loops (Li et al. 2012; Rouchette et al. 2014), grain growth and recrystallization (Ahmed et al. 2014, 2016a, 2016b; Tonks et al. 2014; Liang et al. 2016; Mei et al. 2016), and segregation and precipitation (Badillo et al. 2015; Biner et al. 2016, Li et al. 2014) have appeared in literature. Moreover, studies of the effect of the formation and evolution of these microstructural features on the degradation of nuclear materials were conducted. For example, phase-field simulations of the effect of irradiation-induced microstructure evolution on swelling (Chang et al. 2016), thermal conductivity (Millet et al. 2012), and fracture (Chakraborty et al. 2016) were carried out.

In the phase-field formulation of irradiation effects, all types of equilibrium and non-equilibrium microstructural features such as grains, precipitates, voids, bubbles, cracks, and dislocation loops are treated as second phases that nucleate from the metastable solid phase under irradiation. The irradiation conditions are simulated by the production, diffusion, and reaction of point defects and gas atoms. A general phase-field model formulation of irradiation effects in solids can be described as follows. First, sets of order parameters representing defects concentrations $\{x_1, x_2, ..., x_k\}$, species concentrations $\{c_1, c_2, ..., c_m\}$, and microstructural features $\{\eta_1, \eta_2, ..., \eta_n\}$ of interest are selected. Second, the total free energy functional of the system, postulated to be of the Ginzburg-Landau/Cahn-Hilliard type, is expressed as (recall Eq. (1)):

$$F = \int f_{\text{bulk}} + f_{\text{int}} \ dV, \tag{10}$$

where f_{bulk} is the thermodynamic bulk free energy containing both the short-range chemical free energy and the long-range contribution due to elastic, electrostatic, or magnetic interactions. It is in general a function of all order parameters. f_{int} is

the interfacial free energy of the system. It is usually expressed as the sum of the square gradients of all order parameters (as in Eq. (1)). Lastly, the kinetic evolution equations are derived following the principles of irreversible thermodynamics. The effect of irradiation conditions is directly accounted for by adding terms that represent the defects production and loss rates into the kinetic eqs. A typical phase-field model of irradiation effects in nuclear materials includes the following kinetic equations for the order parameters (phase-field variables):

$$\partial_t x_i = P_i - \nabla \cdot \mathbf{J}_i - L_i \quad \forall i, \tag{11}$$

$$\partial_t c_i = -\nabla \cdot \mathbf{J}_i \quad \forall i, \tag{12}$$

$$\partial_t \eta_i = -M_{\eta_i} \left[\partial_{\eta_i} f_{\text{bulk}} - \kappa_{\eta_i} \nabla^2 \eta_i \right] \quad \forall i. \tag{13}$$

In the above, P_i is the production rate of defects that can be estimated from dose/fission rate, $\mathbf{J}_i$ is the corresponding flux of species/defects, and L_i is the loss due to defects recombination and defects losses to sinks. The fluxes are related to the gradients of the corresponding chemical potentials. Note that due to the coupling between defects and species through the bulk free energy, coupling of their fluxes naturally occurs. Hence, these models can directly take into consideration defects/species segregation or desegregation at interfaces. The strain energy is usually included in these models to account for the stress effect of point defects, gas atoms, bubble pressure, and microstructural features on the overall kinetics. In such case, the stress equilibrium equation is solved simultaneously with the phase-field kinetic equations (Aagesen et al. 2017).

In spite of the success of phase-field models in simulating microstructure evolution in nuclear materials, only a few quantitative investigations were carried out. Most of existing investigations only present qualitative results that cannot be compared with experiments. However, this is not a limitation of the model since, as discussed earlier, the model parameters can be uniquely identified in terms of thermodynamic and kinetic material properties using a formal asymptotic analysis. Such procedure for the case of nuclear materials was conducted recently by Ahmed et al. (2016a, 2016b; Ahmed and El-Azab 2018). In the following, we highlight some of the quantitative results obtained from phase-field simulations of microstructure evolution in nuclear materials.

4.1 General Void Growth Kinetics

The problem of voids/bubbles formation and growth was the first example of the usage of the phase-field method in simulating irradiation effects in materials (Yu and Lu 2005; Hu et al. 2009; Rokkam et al. 2009; Millett et al. 2009). Several physical characteristics of the process such as the existence of incubation period before nucleation, existence of void denuded zones near grain boundaries, dependence

of growth and coarsening rates on dose rate, and the reduction of void number density during coarsening were correctly captured, though only qualitatively, by the phase-field model. However, two different types of the phase-field model for void growth were utilized without rigorous analysis of their differences. Models of type B only used the concentrations of point defects as the phase-field variables/order parameters, while models of type C combined the defects concentrations with a non-conserved order parameter representing the void as a microstructural feature. Moreover, most of the simulations conducted in these studies used unrealistically high defects concentrations on the order of 0.1 (10%).

Furthermore, the results of the void/bubble dynamics are usually presented qualitatively using non-dimensionalized time and length scales. All these limitations clearly prevent model verification and validation.

Ahmed and El-Azab (2018) have recently demonstrated that quantitative phase-field models of void growth in irradiated solids can be constructed in a systematic way using a formal asymptotic analysis. The analysis showed that phase-field models of type B can only capture diffusion-controlled kinetics and are equivalent to the well-known rate theory description of the problem (Olander 1976; Was 2017). On the other hand, phase-field models of type C are capable of describing general void growth kinetics from which diffusion-controlled and surface-controlled arise as limiting cases. The generality of model C was found to be the result of the extra Allen-Cahn equation that acts as an interfacial constitutive law that guarantees nonnegative interfacial entropy production. Therefore, it was concluded that the Allen-Cahn mobility is directly connected to the surface mobility/kinetic barrier through a regular Arrhenius-type relation. The analysis also matched the phase-field model of type C to the recently developed sharp-interface model of void growth (Hochrainer and El-Azab 2015). This matching facilitates fixing the phase-field model parameters in terms of the regular thermodynamic and kinetic data. This was demonstrated by investigating void growth in copper. The quantitative predictions of the model are captured in Figs. 3, 4, and 5. Figure 3 depicts the evolution of vacancy concentration during void growth from a slightly supersaturated copper matrix (with realistically low vacancy concentration). The different types of void growth kinetics are captured in Figs. 5 and 6. As the surface kinetic barrier (Δg) increases relative to the bulk migration energy (E_{m}),
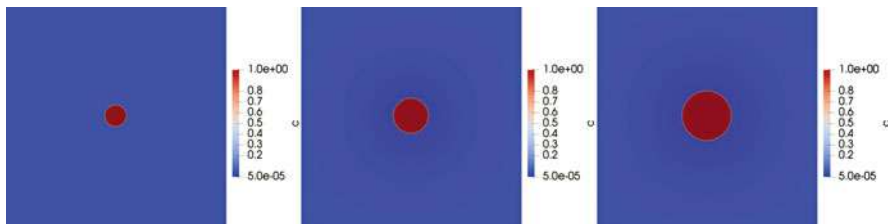
Fig. 3 Snapshots of the evolution of the vacancy concentration during the growth of a void from a supersaturated copper matrix. Note that the vacancy concentration here does not have to be artificially high as in most of existing phase-field models

Fig. 4 The chemical potential across the centerline of the domain in Fig. 3. Large gradient develops across the domain for the case of diffusion-controlled growth, while the chemical potential is almost constant in the entire domain for the interface-controlled case
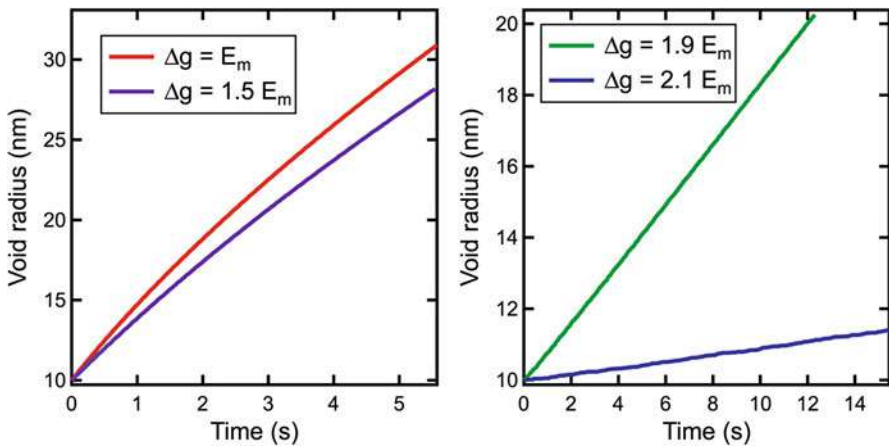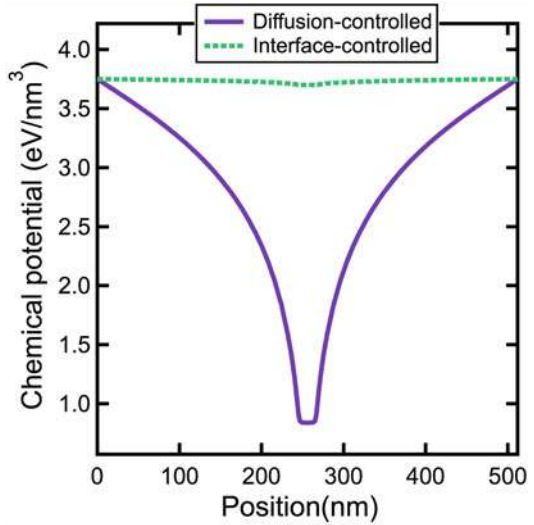
Fig. 5 Effect of the surface kinetic barrier (Δg), assumed here to be a multiple of the bulk migration barrier (E_m), on the overall void growth kinetics in a supersaturated copper at 800 K. The void growth kinetic diminishes as the surface migration barrier increases. Moreover, the kinetics of the process transforms from diffusion-controlled to surface-controlled with increasing the surface barrier

the overall kinetics changes from diffusion-controlled to surface-controlled. This is evident from Fig. 4, where the chemical potential as function of position is plotted. High gradients of the chemical potential are established for the case of diffusion-controlled, while negligible gradients of chemical potential exist during surface-controlled growth. Moreover, the overall kinetics of the void growth process slows down with increasing the kinetic surface barrier as quantitatively represented in Fig. 5.
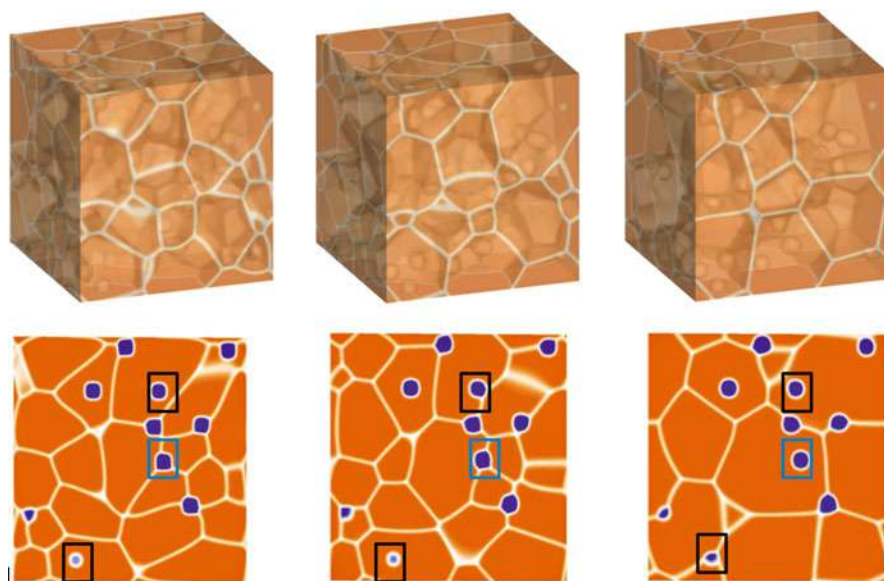
Fig. 6 Snapshots showing the different pore-grain boundary interactions during grain growth of a porous polycrystalline UO_2 at 2000 K. Pore breakaway is evident from the 3D snapshots (upper row). The model is clearly capable of capturing all possible pore-boundary interactions simultaneously as obvious from the 2D cross-sectional images (lower row). Pores can move along with or detach from a migrating boundary. Moreover, initially isolated pores can get attached to a moving boundary. Furthermore, a specific pore can go through a series of attachments and detachments

4.2   Grain Growth in Porous Uranium Dioxide

It is well known that most of the physical properties of polycrystalline solids are strongly dependent on the grain size (Rahaman 2003). Moreover, the behavior of such materials under harsh conditions such as irradiation, high temperature, and high stresses is also highly influenced by grain size. For instance, it was shown that fission gas swelling and release rates decrease with increasing grain size (Turnbull 1974). This has motivated several experimental, theoretical, and computational works for investigating the kinetics of grain growth in the main nuclear fuel.

For the case of porous ceramics such as UO_2, the pore-grain boundary interactions must be taken into account to appropriately model grain growth in such materials (Ahmed 2015). Second-phase particles such as precipitates, bubbles, inclusions, and pores exert a drag force on the migrating grain boundary that retards the grain growth process and may completely stop it as was predicted by the famous Zener's pinning model (Rahaman 2003). However, mobile particles such as pores can move along with the migrating boundary if any of the mass transport mechanisms such as bulk diffusion, surface diffusion, or vapor transport is active. It was shown that most of existing data on grain growth of UO_2 is consistent pore migration via surface diffusion (Bourgeois et al. 2001).

A detailed phase-field model of pore-grain boundary interactions and its effect on grain growth in ceramics was developed by Ahmed et al. (2013, 2014, 2016a, 2016b; 2017). The model takes into consideration the coupled curvature-driven motion of grain boundaries and pore migration by surface diffusion. The phase-field model relaxes the restrictive assumptions on the pore and boundary geometries and motion in the classical models (Nichols 1968, Brook 1969, and Carpay 1977) and obviates the numerical difficulties of the sharp-interface models (Hsueh et al. 1982; Spears and Evans 1982; Svoboda and Riedel 1992; Riedel and Svoboda 1993). Using a formal asymptotic analysis (Ahmed et al. 2016a, 2016b), it was shown that the phase-field model recovers the sharp-interface description of the problem, which facilitates quantitative simulations of grain growth in different porous ceramics (Ahmed et al. 2013, 2014).

The phase-field model of grain growth in porous ceramics was used to simulate grain growth in UO_2 (Ahmed et al. 2014). The model was successfully able to account for all possible pore-grain boundary interactions in such material as obvious from Fig. 6 that captures these different scenarios. A migrating grain boundary can drag mobile pores along with it. It could also break away from less mobile pores. The presence of large number of immobile pores may lead to the cease of grain growth process. The prevailing type of interaction depends on the grain boundary and pore/surface mobilities and the grains and pores sizes and morphologies. However, in a heterogeneous microstructure, all different types of interactions may occur simultaneously as demonstrated in Fig. 6. Moreover, the 3D simulations revealed that boundary breakaway is artificially suppressed in 2D simulations due to the implied assumption of cylindrical pore shape that exaggerates pore drag. An excellent agreement between model predictions and experimental data was found as shown in Fig. 7.

Fig. 7 Comparison of phase-field model predictions and experimental data of grain growth porous polycrystalline UO_2 at 2000 K. The direct comparison was carried out by extrapolating the model results and fitting the experimental data as discussed in Ahmed et al. (2014). Note that such high-fidelity model verification is only possible for quantitative phase-field models (with no arbitrary model parameters)
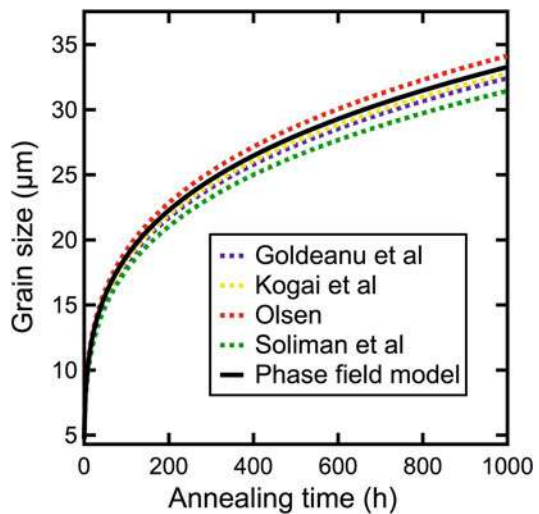
4.3   Hydride Precipitation in Zircaloys

Zirconium alloys are widely used as cladding materials for water-cooled reactors. However, it is known that zircaloys tend to pick up hydrogen during service which, due to the low hydrogen solubility in zirconium, leads to the formation of zirconium hydrides (Carpenter 1973). The precipitation of these hydrides is detrimental to the mechanical integrity of the cladding because hydrides are brittle. Accumulation of these hydrides facilitates crack initiation and propagation that eventually leads to the rupture of the cladding. A few stable and metastable hydrides that precipitate from the hexagonal $\alpha - $ Zr were identified (Carpenter 1973). The most investigated ones are the FCC δ hydride and the FCT γ hydride. The former is the stable configuration usually observed in cladding, and the latter is a metastable phase that forms under high cooling rates. The misfit strains for both hydrides were determined by Carpenter (1973).

A few phase-field models of hydrides formation and growth were proposed in literature (Ma et al. 2002a, 2002b, 2006; Guo et al. 2008a, b). Most of them focus on either δ or γ hydride, but a recent multiphase-field model (Bair et al. 2017) was introduced to simulate all phases of hydrides and the possibility of γ to δ transition. In these models, the misfit strain tensors are usually expressed in terms of phase-field variables to capture their variation with phase. The stress equilibrium equation is then solved concurrently with the phase-field evolution equations.

For the case of γ hydride, phase-field simulations demonstrated, in agreement with experiments, the equilibrium needle shape of these hydrides. Since they have anisotropic misfit strain in the basal plane, their variants with 120° angles between each other can form during an HCP to FCT transition. In the absence of external stress, all variants are likely to precipitate with equal probability. However, under applied stress, the variant normal to the direction of applied stress has lower nucleation barrier. This results in a preferential precipitation of that variant. These results are demonstrated in Fig. 8, which captures the precipitation of these hydrides in the absence and presence of applied stress. The preferential precipitation of hydrides was suggested as possible mechanism of hydride reorientation that takes place under cyclic temperature and stress conditions (Ma et al. 2002). A threshold stress value of 200 MPa was found to be necessary for hydride reorientation. This value lies well within the range reported in literature (Ma et al. 2002a, 2002b).

On the other hand, for the case of δ hydride, the misfit strain is isotropic in the basal plane with the elastically hard direction being normal to that plane. Therefore, the equilibrium shape of these hydrides is a platelet. A 3D phase-field simulation showing the development of that equilibrium shape is presented in Fig. 9. However, their reorientation cannot simply be explained by preferential precipitation as in the case of γ hydrides. It is believed that the 3D stacking of microscale hydrides and plastic effects may hold the key for their reorientation. Nonetheless, a detailed theoretical or computational model that accounts for these effects is still lacking.
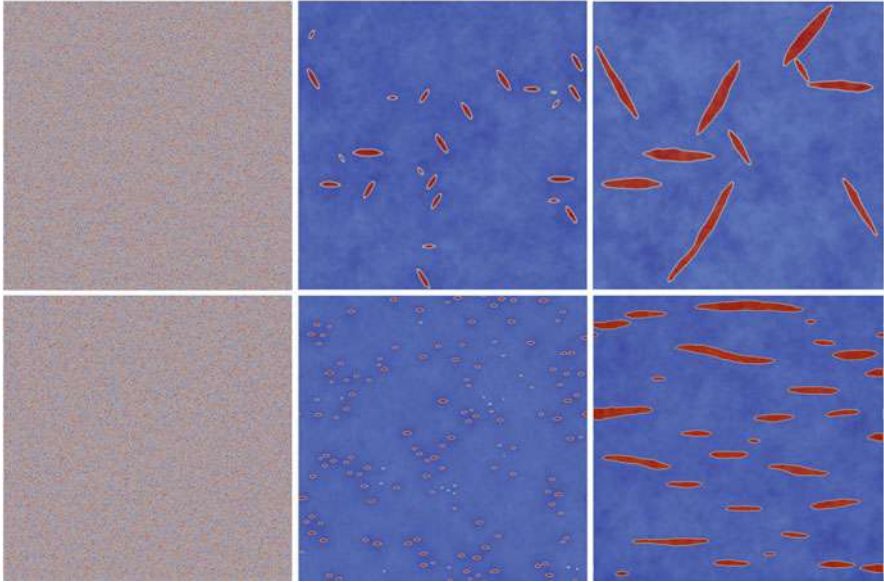
Fig. 8 Snapshots of the precipitation process of γ hydrides under no external load (upper row) and 200 MPa vertical tensile load (lower row). In the absence of applied stresses, all variants of hydrides precipitate out with equal probability. In the presence of an applied load, preferential precipitation of the variant perpendicular to the direction of the applied load takes place
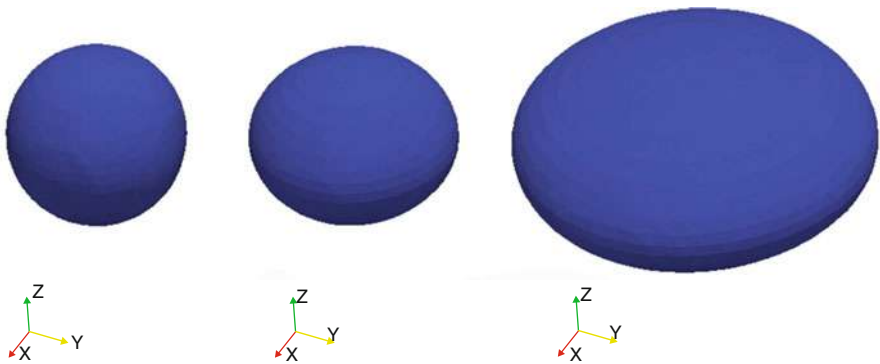
Fig. 9 Snapshots of the development of the equilibrium platelet shape of δ hydrides as observed in experiments. The z-axis here is assumed to coincide with the elastically hard eigen/misfit strain direction [0001]

5   Conclusions and Outlook

Based on its success in simulating microstructure evolution in different materials, the phase-field method was adapted by the nuclear materials community to investigate irradiation effects on materials. The phase-field models of nuclear materials account directly for the production, diffusion, recombination, and reaction

of point defects. Moreover, other extended defects and microstructural features are also included and naturally represented by phase-field variables/order parameters. Furthermore, thanks to its thermodynamic formulation, different types of driving forces such as irradiation, thermal, and mechanical can be considered and coupled in a straightforward manner. In addition, thanks to their spatiotemporal nature, phase-field models inherently account for the effect of microstructure and species heterogeneity on the kinetics of evolution. All these features of the phase-field method enable it to give a more complete picture of microstructure evolution in irradiated materials than other competitive techniques such as rate theory or Monte Carlo. This has led the NEAMS program under DOE to foster the establishment and development of the phase-field-based mesoscale simulator MARMOT to model microstructure evolution in fuel and cladding materials (NEAMS 2017).

Phase-field models were successfully utilized to simulate irradiation-induced microstructure evolution in materials. Phenomena such as irradiation-induced segregation and precipitation; irradiation-induced voids and bubbles formation; growth, coarsening, and dislocation loops; and grain growth were investigated using the phase-field method (Millet and Tonks 2011; El-Azab et al. 2014; Y Li et al. 2017). However, most of these studies only presented qualitative results that cannot be used for predicting material behavior. Nonetheless, as discussed here, this is not an inherent limitation of the method, and a few quantitative investigations were carried out. The construction of such quantitative models requires careful thermodynamic and mathematical formulation of the free energy functional (Moelans 2011; El-Azab et al. 2014). In addition, the sharp-interface limits of these diffuse-interface phase-field models must be obtained in order to directly connect the model parameters to the regular thermodynamic and kinetic material properties (El-Azab et al. 2014; Hochrainer and El-Azab 2015; Ahmed et al. 2016a, 2016b; Ahmed and El-Azab 2018).

Of course, as with any modeling technique, phase-field methods have their limitations. Similar to any coarse-grained modeling approach, their predictive capability is limited by the availability of thermodynamic and kinetic data. There is a clear shortage of such data for nuclear materials. Nonetheless, such limitation can be overcome by conducting lower-scale simulations and/or experiments to obtain these data. Also, parametric studies and uncertainty quantification techniques may be used to assess the effect of missing data on the model predictions. The second major limitation is the high numerical cost of these models. Phase-field simulations scale linearly with number of variables and domain dimension, which make them difficult to implement in 3D for multiphase and multicomponent systems. However, this limitation can be mitigated by using efficient multigrid solver methods and adaptive mesh and time-step techniques. The advancement of sophisticated frameworks such as MOOSE/MARMOT will definitely help in that regard.

There are three main future directions that can be pursued to advance the construction and implementation of phase-field models of irradiation effects in nuclear materials. First, model verification and validation must be a high priority. In order to accomplish this goal, only quantitative phase-field models (having no arbitrary model parameters) should be used. Developing such models is possible through matching with thermodynamically consistent sharp-interface models. The

model predictions can then be directly compared with experiments. Some simplified separate-effects experiments may be designed to facilitate model validation. For instance, experiments that only demonstrate the effects of dose/fission rate, temperature, and total dose/burnup will be invaluable not just to evaluate model predictions but also to enhance model design by pinning down which parts of the model need further improvements. Only then we can start validating more advanced models with coupled driving forces. However, developing and validating such generalized models that account for different aspects of irradiation effects simultaneously must be the ultimate goal. Second, modelers should try to mitigate the limitations of the phase-field method discussed above. Such task requires collaborative community work to expand the thermodynamic and kinetic data of nuclear materials, preferably also to be organized into an accessible online library and improve/develop robust and efficient numerical algorithms. Third, the phase-field models of irradiation effects should be coupled to lower-scale models such as MD, DFT, and KMC to account for both stages of irradiation damage and irradiation effects. Since those stages usually span order of magnitudes with different length and time scales, only a true multi-scale modeling approach can elucidate material behavior in such challenging environment. Innovative theoretical and numerical schemes are required to achieve this goal.

References

Aagesen LK, Schwen D, Ahmed K, Tonks M (2017) Quantifying elastic energy effects on interfacial energy in the Kim-Kim-Suzuki phase-field model with different interpolation schemes. Comput Mater Sci 140:10

Ahmed K (2011) Phase field modeling of microstructure evolution in thermal barrier coating systems. MSc Thesis, Florida State University, Florida

Ahmed K (2015) Phase field modeling of grain growth in porous polycrystalline solids. PhD, Purdue University, Indiana

Ahmed K, El-Azab A (2018) An analysis of two classes of phase field models for void growth and coarsening in irradiated crystalline solids. Mater Theory 2:1

Ahmed K, Yablinsky C, Schulte A, Allen T, El-Azab A (2013) Phase field modeling of the effect of porosity on grain growth kinetics in polycrystalline ceramics. Model Simul Mat Sci Eng 21:065005

Ahmed K, Pakarinen J, Allen T, El-Azab A (2014) Phase field simulation of grain growth in porous uranium dioxide. J Nucl Mater 446:90

Ahmed K, Allen T, El-Azab A (2016a) Phase field modeling for grain growth in porous solids. J Mater Sci 51:1261

Ahmed K, Bai X, Zhang Y, and Biner B (2016b) High-Burnup-Structure (HBS): model development in MARMOT for HBS formation and stability under radiation and high temperature. Technical Report, Idaho National Laboratory (INL), Idaho Falls

Ahmed K, Tonks M, Zhang Y, Biner B, El-Azab A (2017) Particle-grain boundary interactions: a phase field study. Comput Mater Sci 134:25

Allen SM, Cahn JW (1979) A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. Acta Metall 27:1085

Badillo A, Bellon P, Averback RS (2015) A phase field model for segregation and precipitation induced by irradiation in alloys. Model Simul Mater Sci Eng 23:035008

Bair J, Zaeem MA, Schwen D (2017) Acta Mater 123:235

Biner SB (2017) Programming phase-field modeling. Springer, Cham

Biner SB, Rao WF, Zhang YF (2016) The stability of precepitates and the role of lattice defects in Fe-1at% Cu-1at%Ni-1at%Mn alloy: a phase-field model study. J Nucl Mater 468:9

Boettinger WJ, Warren JA, Beckermann C, Karma A (2002) Phase-field simulation of solidification. Ann Rev Mater Res 23163

Bourgeois L, Dehaudt P, Lemaignan C, Fredric JP (2001) J Nucl Mater 295:73

Brailsford AD, Bullough R (1972) The rate theory of swelling due to void growth in irradiated metals. J Nucl Mater 44:121

Brook RJ (1969) Pore-Grain Boundary Interactions and Grain Growth. J Am Ceram Soc 52:56

Cahn JW (1961) On spinodal decomposition. Acta Metall 9:795

Cahn JW, Hilliard JEJ (1958) Free Energy of a Nonuniform System. I. Interfacial Free Energy. Chem Phys 28:258

Carpay FMA (1977) Discontinuous Grain Growth and Pore Drag. J Am Ceram Soc 60:82

Carpenter GJC (1973) The dilatational misfit of zirconium hydrides precipitated in zirconium. J Nucl Mater 48:264

Chakraborty P, Zhang Y, Tonks MR (2016) Multi-scale modeling of microstructure dependent intergranular brittle fracture using a quantitative phase-field based method. Comput Mater Sci 113:38

Chang K, Lee GG, Kwon J (2016) A phase-field modeling of void swelling in the austenitic stainless steel. Radiat Eff Def Solids 171:242–251

Chen LQ (2002) Phase-field models for microstructure evolution. Ann Rev Mater Res 32:113

CHiMaD (2017) http://chimad.northwestern.edu/. Accessed 15 Oct 2017

COMSOL (2017) https://www.comsol.com/. Accessed 15 Oct 2017

De Groot S, Mazur P (1962) Non-equilibrium thermodynamics. North-Holland Publishing Company, Amsterdam

El-Azab A, Ahmed K, Rokkam S, Hochrainer T (2014) Diffuse interface modeling of void growth in irradiated materials. Mathematical, thermodynamic and atomistic perspectives. Curr Opinion Solid State Mat Sci 18:90

Elder KR, Grant M, Provatas N, Kosterlitz JM (2001) Sharp interface limits of phase-field models. Phys Rev E 64:021604

Emmerich H (2008) Advances of and by phase-field modelling in condensed-matter physics. Adv Phys 57:1

FEniCS (2017) https://fenicsproject.org/. Accessed 15 Oct 2017

Fife PC (1992) Dynamics of internal layers and diffuse interfaces. SIAM, Philadelphia

FiPy (2017) https://www.ctcms.nist.gov/fipy/. Accessed 15 Oct 2017

Gaston D, Newman C, Hansen G, Lebrun-Grandie D (2009) MOOSE: A Parallel Computational Framework for Coupled Systems of Nonlinear Equations. Nucl Eng Des 239:1768

Gaston D et al (2015) Physics-based multiscale coupling for full core nuclear reactor simulation. Ann Nucl Energy 84:45

Ginzburg V, Landau L (1950) On the theory of superconductivity. Sov Phys JETP 20:1064

Glauber RJ (1963) The quantum theory of optical coherence. J Math Phys 4:294

Guo XH, Shi SQ, Zhang QM, Ma XQ (2008a) An elastoplastic phase-field model for the evolution of hydride precipitation in zirconium. Part I: smooth specimen. J Nucl Mater 378:110–119

Guo XH, Shi SQ, Zhang QM, Ma XQ (2008b) An elastoplastic phase-field model for the evolution of hydride precipitation in zirconium. Part II: specimen with flaws. J Nucl Mater 378: 120–125

Hochrainer T, El-Azab A (2015) A sharp interface model for void growth in irradiated materials. Phil Mag 95:948

Hohenberg PC, Halperin BI (1977) Theory of dynamic critical phenomena. Rev Mod Phys 49:435

Hsueh CH, Evans AG, Coble RL (1982) Microstructural Development During Final/Intermediate Stage Sintering - I. Pore/Grain Boundary Separation. Acta Metall 30:1269

Hu SY, Henager CH Jr (2009) Phase-field modeling of void lattice formation under irradiation. J Nucl Mater 394:155

Hu SY, Henager CH Jr (2010) Phase-field simulation of void migration in a temperature gradient. Acta Mater 58:3230

Hu SY, Henager CH Jr, Heinisch HL, Stan M, Baskes MI, Valone SM (2009) Phase-field modeling of gas bubbles and thermal conductivity evolution in nuclear fuels. J Nucl Mater 392:292

Hu SY et al (2016a) Formation mechanism of gas bubble superlattice in UMo metal fuels: phase-field modeling investigation. J Nucl Mater 479:202–215

Hu SY, Burkes D, Lavender CA, Joshi V (2016b) Effect of grain morphology on gas bubble swelling in UMo fuels—a 3D microstructure dependent booth model. J Nucl Mater 480:323

Jokisaari A, Vorhees P, Guyer J, Warren J, Heinonen O (2017) Benchmark Problems for Numerical Implementations of Phase Field Models. Comput Mater Sci 126:139–151

Kawasaki K (1966) Diffusion Constants near the Critical Point for Time-Dependent Ising Models. Phys Rev 145:224

Kim SG, Kim WT, Suzuki T (1999) Phase-field model for binary alloys. Phys Rev E 60:7186

Li Y, Hu S, Sun X, Gao F, Henager H, Khaleel M (2010) Phase-field Modeling of Void Migration and Growth Kinetics in Materials under Irradiation and Temperature Field. J Nucl Mater 407:119

Li Y et al (2012) Computer simulations of interstitial loop growth kinetics in irradiated bcc Fe. J Nucl Mater 427:259–267

Li Y, Hu S, Montgomer R, Gao F, Sun X (2013) Phase-field simulations of intragranular fission gas bubble evolution in UO2 under post-irradiation thermal annealing. Nucl Instr Meth Res B 303:62

Li Y, Hu S, Zhang L, Sun X (2014) Non-classical nuclei and growth kinetics of Cr precipitates in FeCr alloys during aging. Modelling Simul. Mater. Sci. Eng. 22:025002

Li Y et al (2017) A review: applications of the phase field method in predicting microstructure and property evolution of irradiated nuclear materials. npj Computational Materials 3:16

Liang L et al (2016) Mesoscale model for fission-induced recrystallization in U-7Mo alloy. Comput Mater Sci 124:228–237

LibMesh (2017) http://libmesh.github.io/. Accessed 15 Oct 2017

Ma XQ, Shi SQ, Woo CH, Chen LQ (2002a) Phase-field simulation of hydride precipitation in bi-crystalline zirconium. Scr Mater 47:237–241

Ma XQ, Shi SQ, Woo CH, Chen LQ (2002b) Effect of applied load on nucleation and growth of gamma-hydrides in zirconium. Comput Mater Sci 23:283–290

Ma XQ, Shi SQ, Woo CH, Chen LQ (2006) The phase field model for hydrogen diffusion and gamma-hydride precipitation in zirconium under non-uniformly applied stress. Mech Mater 38:3

Mei ZG et al (2016) Grain growth in U-7Mo alloy: a combined first-principles and phase field study. J Nucl Mater 473:300–308

MICRESS (2017) http://web.micress.de/. Accessed 15 Oct 2017

Millett P, Tonks M (2011) Application of phase-field modeling to irradiation effects in materials. Curr Opin Solid State Mater Sci 15:125

Millett P, Rokkam S, El-Azab A, Tonks M, Wolf D (2009) Void nucleation and growth in irradiated polycrystalline metals: A phase-field model. Mod Sim Mater Sci Eng 17:064003

Millett P, El-Azab A, Rokkam S, Tonks M, Wolf D (2011a) Phase-field simulation of irradiated metals: Part I: Void kinetics. Comput Mater Sci 50:949

Millett P, El-Azab A, Wolf D (2011b) Phase-field simulation of irradiated metals: Part II: Gas bubble kinetics. Comput Mater Sci 50:960

Millett PC et al (2012) Phase-field simulation of intergranular bubble growth and percolation in bicrystals. J Nucl Mater 425:130–135

Moelans N (2011) A quantitative and thermodynamically consistent phase- field interpolation function for multi-phase systems. Acta Mater 59(3):1077

Moelans N, Blanpain B, Wollants P (2008) An introduction to phase-field modeling of microstructure evolution. Calphad 32:268

MOOSE (2017 http://mooseframework.org/. Accessed 15 Oct 2017

NEAMS (2017) https://neams.inl.gov/. Accessed 15 Oct 2017

Nichols FA (1968) Further Comments on the Theory of Grain Growth in Porous Compacts. J Am Ceram Soc 51:468

Olander D (1976) Fundamental aspects of nuclear reactor fuel element. Technical Information Service, Springfield

Olander D, Motta A (2017) Light water reactor materials volume I: fundamentals. American Nuclear Society, LaGrange Park

OpenPhase (2017) http://www.openphase.de/. Accessed 15 Oct 2017

Pego RL (1989) Proc Roy Soc London Ser A 422:261

PETSc (2017) http://www.mcs.anl.gov/petsc/. Accessed 15 Oct 2017

PF Hub (2017) https://pages.nist.gov/chimad-phase-field/. Accessed 15 Oct 2017

Plapp M (2011) Unified derivation of phase-field models for alloy solidification from a grand-potential functional. Phys Rev E 84:031601

PRISMS (2017) http://www.prisms-center.org/. Accessed 15 Oct 2017

Provatas N, Elder K (2010) Phase-field methods in materials science and engineering. Wiley, Weinheim

Rahaman MN (2003) Ceramic processing and sintering. Marcel Dekker, New York

Riedel H, Svoboda J (1993) A theoretical study of grain growth of porous solids during sintering. Acta Metall Mater 41:1929

Rokkam S, El-Azab A, Millett P, Wolf D (2009) Modelling and simulation in materials science and engineering. Mod Sim Mater Sci Eng 17:064002

Rouchette H, Thuinet L, Legris A, Ambard A, Domain C (2014) Quantitative phase field model for dislocation sink strength calculations. Comput Mater Sci 88:50–60

Rowlinson JS (1979) Translation of J. D. van der Waals' "The thermodynamic theory of capillarity under the hypothesis of a continuous variation of density". J Stat Phys 20:197

Schwen D, Aagesen LK, Peterson JW, Tonks M (2017) Rapid multiphase-field model development using a modular free energy based approach with automatic differentiation in MOOSE/MARMOT. Comput Mater Sci 132:36–45

Semenov AA, Woo CH (2012) Phase-field modeling of void formation and growth under irradiation. Acta Mater 60:6112

Short MP, Yip S (2015) Materials aging at the mesoscale: kinetics of thermal, stress, radiation activations. Curr Opin Solid State Mater Sci 19:245–252

Spears MA, Evans AG (1982) Microstructural development during final/intermediate stage sintering, II. grain and pore coarsening. Acta Metall 30:1281

Steinbach I (2009) Phase-field models in materials science. Model Simul Mater Sci Eng 17: 073001

Svoboda J, Riedel H (1992) Pore-boundary interactions and evolution equations for the porosity and the grain size during sintering. Acta Metall Mater 40:2829

Tonks MR, Gaston D, Millett PC, Andrs D, Talbot P (2012) An object-oriented finite element framework for multiphysics phase field simulations. Comput Mater Sci 51:20

Tonks MR, Zhang Y, Bai X, Millett PC (2014) Demonstrating the temperature gradient impact on grain growth in UO_2 using the phase field method. Mater Res Lett 2:23

Tonks MR et al (2017) Mechanistic materials modeling for nuclear fuel performance. Annals of Nuclear Energy 105, 11–24

Turnbull JA (1974) The effect of grain size on the swelling and gas release properties of UO_2 during irradiation. J Nucl Mater 50:62

Was GS (2017) Fundamentals of radiation materials science-metals and alloys. Springer, Berlin

Wheeler AA, Boettinger WJ, McFadden GB (1992) Phase-field model for isothermal phase-transitions in binary-alloys. Phys Rev A 45:7424–7439

Williamson RL, Hales JD, Novascone SR, Tonks MR, Gaston DR, Permann CJ, Andrs D, Martineau RC (2012) Multidimensional multiphysics simulation of nuclear fuel behavior. J Nucl Mater 423:149–163

Xiao ZH, Semenov AA, Woo CH, Shi SQ (2013) Single void dynamics in phase field modeling. J Nucl Mater 439:25

Yu HC, Lu W (2005) Dynamics of the self-assembly of nanovoids and nanobubbles in solids. Acta Mater 53:1799

Zhang L et al (2012) Phase-field modeling of temperature gradient driven pore migration coupling with thermal conduction. Comput Mater Sci 56:161–165

Zhang L, Tonks MR, Gaston D, Peterson JW, Andrs D, Millett PC, Biner BS (2013) A quantitative comparison between C0 and C1 elements for solving the Cahn–Hilliard equation. J Comp Phys 236:74

Thermodynamic Modeling of Nuclear Fuel Materials

95

Jacob W. McMurray and Theodore M. Besmann

Contents

J. W. McMurray (✉)
Materials Science and Technology, Oak Ridge National Laboratory, Oak Ridge, TN, USA
e-mail: mcmurrayjw1@ornl.gov

T. M. Besmann
College of Engineering and Computing, University of South Carolina, Columbia, SC, USA
e-mail: besmann@cec.sc.edu

Abstract

Thermodynamics is essential, fundamental information for understanding the behavior of any system. For nuclear materials, the situation is even more critical. The fuel in particular becomes highly multicomponent with burnup and is constantly evolving from beginning to end of life. Calculation of Phase Diagram (CALPHAD) modeling gives a continuous thermodynamic description of a system and allows coupling of the information to other phenomena facilitating multiphysics simulations.

1   Introduction

The combined first and second laws of thermodynamics give us the extremum principle. Stated simply, an isolated system will maximize its entropy and minimize its free energy. As will be discussed later, the Gibbs energy has temperature and pressure as natural variables and is therefore generally the most convenient thermodynamic function to use for treating chemical equilibria. Mathematically, the equilibrium state of a multicomponent isolated system at constant temperature and pressure is the assemblage of phases and their respective compositions that correspond to the global constrained minimum of the total Gibbs free energy.

A nuclear fuel system is continuously evolving due to transmutations and fission product (FP) generation. Because of the high temperatures, there is a rapid approach to equilibrium; therefore, thermodynamic arguments are often sufficient to predict many phenomena and are, at a minimum, necessary to understand more complex, dynamic physics such as corrosion, heat transfer, and mass transport. While the structural materials for reactors can benefit from thermodynamic insight, the emphasis here is reserved for the fuel since it is the fuel that generates the heat and experiences the highest temperatures and evolves to a multicomponent system of up to 60 elements.

A major issue for fresh light water reactor (LWR) fuel is it can have variable initial composition. It is therefore a complex system even before irradiation. All fuels become even more so with burnup as fission and transmutation products are generated. Some of these dissolve into the parent phase while others form secondary ones, for example, the five-metal white phase (molybdenum, rhodium, palladium, ruthenium, and technetium) and perovskites in oxide fuels. Therefore, thermodynamic representations of the system must capture the important complex behavior, for example, nonstoichiometry of the crystalline structures, ordering in ionic melts, and dissolution of elements into the original and secondary phases. Fortunately, there are models available that incorporate these characteristics; they are discussed in Sect. 5.

Finally, while thermodynamics does not explicitly include the physics of time evolution to equilibrium, the tendency of a system to evolve to its lowest energy state and to balance potentials does provide driving forces for many kinetic phenomena within certain formalisms, for example, the *theory of irreversible processes*

(De Groot and Mazur 2013). As mentioned, fuel experiences high temperatures. Thus, the assumption of local thermodynamic equilibrium (LTE) is useful for predicting the chemical state of the elements and the material properties of the fuel; these are important inputs for representing many kinetically driven processes like phase transformations, microstructural evolution, and transport phenomenon.

2   Classical Thermodynamics

Thermodynamics can be viewed as the study of the manifestations of energy and its transformations. The fundamental mathematical relations were empirically derived from the necessity of describing the operation of steam engines (J.M Smith et al. 1996), and the theory was essentially complete at the end of the 19th century (DeHoff 2006).

The first law captures the essence of the science through the following statement:

Although energy assumes many forms, the total quantity of energy is constant, and when energy disappears in one form it appears simultaneously in other forms. (J.M Smith et al. 1996)

Expressed mathematically:

$$dU = \delta Q - PdV \tag{1}$$

where U is internal energy, Q is heat, P is pressure, and V is volume. The δ symbol signifies that Q is a process functions, i.e., the integral of δQ depends not only on the initial and final state, but also on the path.

The second law asserts there is a proper direction for transformation of energy from one form to another. It is the statement of the principle of entropy increase given as:

$$dS \geq 0 \tag{2}$$

A consequence of Eq. 2 is that an isolated system, if left alone, may never decrease in entropy. The combined first and second laws, formulated in Eq. 3, give the fundamental relation of thermodynamics (Willard 1961).

$$dU \leq TdS - PdV \tag{3}$$

where the S and V are the natural variables of U. It is often desirable to represent Eq. 3 as a different thermodynamic potential dependent upon another set of more experimentally useful natural variables; this can be accomplished by performing a partial Legendre transform on Eq. 3 to obtain the Helmholtz (F) and the Gibbs (G) energies.

$$G = U + PV - TS \tag{4}$$

$$F = U - TS \tag{5}$$

Equations 4 and 5 yield all the information contained in Eq. 3, but with natural variables (T,P) and (T,V), respectively. Upon differentiation, Eqs. 4 and 5 become:

$$dG = dU + VdP + PdV - TdS - SdT \tag{6}$$

$$dF = dU - TdS - SdT \tag{7}$$

Applying Eq. 3 means that the G and F must never increase at constant (T,P) and (T,V), respectively.

$$dG \leq 0 \tag{8}$$

$$dF \leq 0 \tag{9}$$

Representing any real irreversible process with a reversible path underpins calculations made within the framework of *Classical Thermodynamics*; therefore, the equality will strictly apply and Eq. 8 will be used from here on to define equilibrium, i.e., $dG = 0$.

Despite significant advances in science since the advent of *Classical Thermodynamics*, only a minor modification to the first law has been introduced and no experiment has ever directly or indirectly provided evidence that contradicts the second law (Van Wylen et al. 1994). The famous Einstein equation, $E = mc^2$, effectively combines the laws of conservation of mass and energy such that mass is simply yet another manifestation of energy and the first law remains soundly intact. There is a zeroth law that states that two bodies in thermal equilibrium with a third are themselves in thermal equilibrium (Van Wylen et al. 1994). The third law will be discussed later on as it applies to modeling, but the fundamentals of *Classical Thermodynamics* can be applied based on the first and second laws alone.

3   The Ideal Gas

A natural didactic and utilitarian companion to the *Classical Thermodynamics* framework is the equation of state (EOS) for an ideal gas derived from combination of Charles', Boyle's, and Avogadro's laws (Chang 1998); it is mathematically expressed as:

$$PV = nRT \tag{10}$$

where n is number of moles and R is the ideal gas constant. Dalton's law states that ideal gases are noninteracting and every gas behaves as a vacuum unto every other gas (Chang 1998). This leads to an important simplification for low pressures and/or high temperatures with very convenient and useful consequences for thermodynamic calculations; the sum of partial pressures equals the total pressure (Eq. 11):

$$P \equiv \sum_i p_i = \sum_i x_i P = \sum_i \frac{n_i}{N} P \tag{11}$$

where n_i is the number of moles of component i, N is the total number of moles of the system, and x_i is the mole fraction of component i. The change in the Gibbs energy at constant temperature can be shown to be:

$$dG = V dP \tag{12}$$

Employing Eq. 10 as the EOS, the familiar equation for the pressure effect on the Gibbs energy of an ideal gas is recovered.

$$G = n^\circ G + nRT \ln \frac{P}{1bar} \tag{13}$$

Here, $^\circ G$ is the molar Gibbs energy of the ideal gas at the temperature of the system referenced to a standard pressure of 1 bar. Furthermore, upon application of Dalton's law of partial pressures, the entropic effect on the Gibbs energy from the mixing of ideal gases is included in Eq. 14.

$$G = \sum_i n_i G^\circ_i + \sum_i n_i \ln x_i + NRT \ln \frac{P}{1bar} \tag{14}$$

where the change in entropy, or the entropy of mixing, S^{mix}, is:

$$S^{mix} = R \left(\sum_i n_i \ln x_i \right) \tag{15}$$

4   Ideal Mixing: The Statistical Mechanics Approach to Entropy

Pioneered by Ludwig Boltzmann and J. Willard Gibbs, statistical mechanics provides a link between the energy states described by quantum and classical treatments and the macroscopic observable properties of thermodynamics described earlier. Entropy is best defined using statistical mechanics and, according to the Boltzmann equation, is:

$$S = k_B \ln \Omega(U) \tag{16}$$

where k_B is the Boltzmann constant, and $\Omega(U)$ is the number of microstates with energy U. Assuming no interactions between solution components, Eq. 16 is used to derive the *ideal entropy of mixing* (S^{id}) for condensed phases resulting in exactly the same expression given by Eq. 15 for an ideal gas mixture.

$$S^{id} = S^{mix} = R \left(\sum_i n_i \ln x_i \right) \tag{17}$$

Although approximate and generally an overestimate, Eq. 17 is widely applied in the field of thermodynamic modeling and is a convenient, utilitarian way of representing the entropy of mixing resulting from solution formation. It is used directly in the regular solution model and with modification in the sublattice model for ionic crystalline material discussed later. It is worth noting that S^{id} is used to distinguish from S^{mix} since S^{mix} provides an exact description of the configurational entropy from the mixing of ideal gases and is an exceptionally accurate approximation for *actual* gases within the limit that Dalton's law may be applied. The word *actual* is purposely used here in lieu of *real* since the term *real gas* is the conventional way of referring to gases that deviate from ideal behavior, i.e., Dalton's law is inapplicable and significant interactions are present.

It is well established that the phenomenological relations of *Classical Thermodynamics*, such as the pressure effect on the Gibbs energy and the entropy of mixing of an ideal gas, can be derived from quantum and classical statistics using many of the same simplifying assumptions that led to the development of the empirical laws that underpin Eqs. 10 and 11. This fact validates the statistical mechanics approach to entropy and helps to establish a link between the micro- and macroscopic quantum and classical treatments of thermodynamics.

5   Thermodynamic Models

A useful form of the Gibbs energy equation is obtained by defining a new potential, the enthalpy, H. The definition of H is given as:

$$H = U + PV \tag{18}$$

Enthalpy has natural variables S and P as can be seen from the differential form:

$$dH = TdS + VdP \tag{19}$$

It is one of the four fundamental thermodynamic potentials along with U, G, and F. Here, H is introduced since a very convenient and widely applied from of G is given by:

$$G = H - TS \tag{20}$$

While the third law of thermodynamics establishes the concept of absolute entropy, there is no such assertion for enthalpy (H); therefore, Gibbs energy functions rely on energy differences between references taken at a standard temperature and pressure (STP) generally 298 K and 1 bar. The convention is to use the elements as references with zero H in their stable form at STP. This gives rise to the concept of a standard enthalpy of formation ΔH_f, alternately referred to as the stable element reference and denoted H^{SER}. Using Maxwell's relations and assigning $H_{298} = \Delta H_f$, the functions for H and S with temperature can be defined as (E Brian Smith 2013):

$$H = H_{298} + \int_{298}^{T} c_p dT \tag{21}$$

$$S = S_{298} + \int_{298}^{T} \frac{c_p}{T} dT \tag{22}$$

where c_p is the heat capacity at constant pressure. The Gibbs energy is expressed as a temperature dependent polynomial of the general form:

$$G = a + b + cT \ln T + dT^2 + eT^3 + fT^{-1} \ldots \tag{23}$$

The ellipsis denotes that additional terms can be added to achieve a better representation when compared to experimentally determined values, but these are rarely needed.

The heat capacity can be found from the G by utilizing the following relation (Prausnitz et al. 1998):

$$c_p = -T \left(\frac{\partial^2 G}{\partial T^2} \right) \tag{24}$$

resulting in:

$$c_p = -c - 2dT - 6eT^2 - 2fT^{-2} \ldots \tag{25}$$

While c_p can always be calculated from the Gibbs function, the reverse is not true due to the creation of integration constants thus underscoring the importance of knowing a standard enthalpy and absolute (or standard) entropy. It is worth noting that differences in standard states can have small but serious computational consequences and this can arise when utilizing data from different sources.

5.1   Pure Elements and Stoichiometric Compounds

For a pure element or a stoichiometric compound, Eqs. 23 and 24 are sufficient for describing the Gibbs energy and heat capacity. The deconvolution of Eq. 23 gives S_{298} and H_{298} and therefore contains all of the thermodynamic information illustrating the importance of its potential in computational thermodynamic applications.

5.2   The Regular Solution Model

The ideal solution model is given by Eq. 14 for a gas but can be extended to condensed phases by assuming negligible pressure dependence, which is generally true in practice; therefore, Eq. 14, in terms of mole fractions, reduces to:

$$G^{total} = N \left(\sum_i x_i \, ^\circ G_i + \sum_i x_i \ln x_i \right) \tag{26}$$

Eq. 26 can be expressed in terms of the molar Gibbs energy, G, as:

$$G = \sum_i x_i \, ^\circ G_i + \sum_i x_i \ln x_i \tag{27}$$

The ideal solution model assumes no interaction between end-members that constitute the phase. The Gibbs energy of many liquid and solid mixtures can be sufficiently represented with Eq. 27, particularly within the limit of dilute solutions. For solutions that exhibit strong interactions, an additional term is added and Eq. 27 becomes the well-known regular solution model given by:

$$G = \sum_i x_i \, ^\circ G_i + \sum_i x_i \ln x_i + {}^{xs}G \tag{28}$$

where ${}^{xs}G$ is termed the excess Gibbs energy. A common form of ${}^{xs}G$ is the Redlich-Kister polynomial (Redlich and Kister 1948) that yields interaction energies as a function of composition (Prausnitz et al. 1998). To illustrate, for a binary system, a Redlich-Kister excess Gibbs energy is given as:

$$^{xs}G = x_1 x_2 \sum {}^i L (x_1 - x_2)^i \tag{29}$$

where the ${}^i L$ terms are the ith order interaction energies and are commonly represented as a function of temperature following the form:

$$^i L = {}^i A + {}^i BT + {}^i CT \ln T + {}^i DT^2 \ldots \tag{30}$$

Gases that deviate from ideality are treated differently, due primarily to pressure effects, but the idea that interactions between unlike components generate an excess Gibbs energy contribution is the same. Eq. 14 can still be used for a *real gas* by correcting the pressure term in Eq. 10 using the fugacity coefficient, φ, given by:

$$\phi = e^{\int_0^P \left(\frac{z-1}{P}\right)dP} \tag{31}$$

where z is known as the compressibility factor given by:

$$z \equiv \frac{PV}{NRT} \tag{32}$$

When a *real gas mixture* description is warranted, φ is modified to $\widehat{\phi_i}$, or the fugacity coefficient of species i in solution, which is a function of the interaction between components that reduces to φ for pure i as the composition of i approaches unity. Finally, the excess function for a real gas can be characterized by Eq. 33 below.

$$^{xs}G = \sum RT \ln \widehat{\phi_i} \tag{33}$$

5.3   The Associate Species Model

To represent the thermodynamics of nonstoichiometric solution phases or the short range order (SRO) in some liquids, the associate species model uses fictive or fictitious molecule-like species as end-members in a regular solution model. As an example, to represent the oxygen potential of the $U_{1-y}Gd_yO_{2 \pm x}$ phase, Lindemer and Sutton (Lindemer and Sutton Jr. 1988) used the U_2O_5 and $UGdO_4$ associate compounds in a regular solution model framework. To illustrate the SRO in a generic liquid mixture of components A and B with a tendency to form AB associates, the appropriate model is given by:

$$G^{total} = n_A G_A + n_B G_B + n_{AB} G_{AB} + NRT \left(x_A \ln x_A + x_B \ln x_B + x_{AB} \ln x_{AB}\right)$$
$$+ G^{xs} \tag{34}$$

The major drawback to Eq. 34 is that it does not reduce to the regular solution model when there is no tendency for association, i.e., $\Delta G_{AB} = 0$ for the so called pair exchange reaction given by Eq. 35 below.

$$(A - A) + (B - B) = 2(A - B) \tag{35}$$
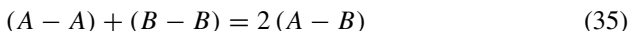
5.4   Sublattice Models

The modeling of complex interstitial and substitutional solutions of crystalline phases can best be accomplished with a sublattice approach using either a reciprocal solid solution (RSS) or compound energy formalism (CEF) model. The two are not easily distinguishable as the RSS is a special case of the CEF. The major difference lies in the fact that the CEF is a broadly applied formalism approach for which the RSS, depending on the stoichiometry of the crystal, could be a model within the CEF framework.

Since the CEF assumes the general form of the sublattice model, it will be used from here on to generate a description. The long range ordering (LRO) of a crystal with multiple sites, or sublattices, on which mixing takes place gives rise to a modification of the entropic term and, according to the CEF, it becomes:

$$S^{CEF} = -R \left(\sum_s \sum_J n^s y_J^s \ln y_J^s \right) \tag{36}$$

where y_J^s is the site fraction of constituent J on sublattice s and n^s is the stoichiometric coefficient of sublattice s. There is a mechanical mixture of all the end-member *compounds*, called a surface of reference ($G^{s.r.}$), that is, all possible combinations of the members from different sublattices. Here, the term *compound* is italicized and loosely applied since some combinations may be unphysical or fictive, meaning that they only exist in solution and not as a pure compound. The surface of reference is described in general mathematical terms as:

$$G^{s.r.} = \sum_k {}^{\circ}G_k \prod_s y_J^s \tag{37}$$

where ${}^{\circ}G_k$ is the Gibbs energy for the end-members and the $\prod_s y_J^s$ term accounts for all the sublattices that contain components of end-member k, the end-member being operated on by the summation. Combining Eqs. 37 and 38 and adding ${}^{xs}G$ gives the common form of the molar Gibbs energy CEF expression:

$$G^{CEF} = \sum_k {}^{\circ}G_k \prod_s y_J^s + RT \left(\sum_s \sum_J n^s y_J^s \ln y_J^s \right) + {}^{xs}G \tag{38}$$

Liquid mixtures can be represented with a regular solution model, but a tendency for ordering necessitates a more sophisticated approach that directly treats interactions among constituents. While the technique of including associates discussed above was originally developed for this purpose, an alternative is the two-sublattice liquid model (TSLM) that can be used to describe the ordering in ionic liquids (Hillert et al. 1985). The TSLM is based on the concept first proposed by Temkin (Temkin 1945) for salt melts where each atom is practically surrounded by unlike

atoms and cations mix at random only with each other while anions randomly reposition exclusively with other anions. The situation is conveniently represented with two sublattices, one for the cations and one for anions, and the melt is described with the same approach used for crystalline phases with LRO (Hillert et al. 1985).

Hillert et al. (Hillert et al. 1985) made modifications to extend the application of the TSLM to include components with multiple valences by allowing the stoichiometry to vary. Vacancies and associates, both neutral and charged, were also introduced to describe nonstoichiometric melts. While this may not always be a physical reality, it is an expedient way of representing the Gibbs energy of liquids ranging from a purely metallic-like solution to one with short range ordering from ionic interactions. This approach has since come to be known as the partially ionic TSLM and within its framework electroneutrality is defined as:

$$P = \sum_j v_j y_{A_j} + Q y_{Va} \tag{39}$$

$$Q = \sum_i v_i y_{C_i} \tag{40}$$

where v_j and y_{A_j} are the charge and site fraction of the anion species and v_i and y_{C_i} are the charge and site fraction of the cation C, respectively; y_{Va} is the site fraction for a vacancy of charge Q-, thus.

$$G = \sum_C \sum_A y_C y_A {}^{\circ}G_{C:A} + Q y_{Va} \sum_C y_C {}^{\circ}G_C + Q \sum_B y_B {}^{\circ}G_B - TS + {}^{xs}G \tag{41}$$

where ${}^{\circ}G_{C:A}$ is the Gibbs energy of the liquid constituent corresponding to the formula $C_{v_A} A_{v_C}$ and ${}^{\circ}G_C$ is the Gibbs energy of element C. The symbols y_B and ${}^{\circ}G_B$ correspond to the site fraction and the Gibbs energies of neutral B associates.

The configurational entropy term S is given by:

$$S = -R \left[P \sum_C y_C \ln y_C + Q \left(\sum_A y_A \ln y_A + y_{Va} \ln y_{Va} + \sum_B y_B \ln y_B \right) \right] \tag{42}$$

5.5   Quasichemical and Modified Quasichemical Models

A quasichemical treatment of SRO in liquids is an alternative to the associate species and the TSLM approaches. The model was developed by Fowler and Guggenheim who based the configurational entropy on a one-dimensional lattice. Pelton and coworkers introduced modifications to the model in order to permit the composition

of maximum SRO to vary and to extend the application to multicomponent systems (AD Pelton et al. 2000; Arthur D Pelton and Chartrand 2001).

The modified quasichemical model (MQM) is best illustrated first with a generic binary example consisting of components A and B which are distributed over a one-dimensional quasilattice. The pair approximation is used, that is, only nearest neighbor pairs are considered to interact and this interaction is independent of the other pairs. As such, the driving force for ordering is the pair exchange reaction given by Eq. 35:

This can be viewed as a *quasi*-chemical reaction with an energy, ΔG_{AB}. The Gibbs energy of the solution is modified by ΔG_{AB} and becomes:

$$G^{total} = n_A \, ^\circ G_A + n_B \, ^\circ G_B - T S^{config} + \left(\frac{n_{AB}}{2} \right) \Delta G_{AB} \tag{43}$$

where n_A, n_B, and n_{AB} are the number of moles of component A, B, and AB pairs, respectively, and these quantities are bound by the relations:

$$Z_A n_A = 2 n_{AA} + n_{AB} \tag{44}$$

$$Z_B n_B = 2 n_{BB} + n_{AB} \tag{45}$$

Here Z_A and Z_B are the coordination numbers of A and B and $n_{AB} = n_{BA}$. In the quasichemical model after Guggenheim and Fowler (Fowler and Guggenheim 1949), $Z_A = Z_B = Z$; however, by allowing these to be chosen independently, the composition corresponding to maximum ordering is determined by the ratio Z_B/Z_A and can therefore be adjusted to fit the behavior of the system.

The pair fractions are defined as:

$$X_{ij} = \frac{n_{ij}}{n_{AA} + n_{BB} + n_{AB}} \tag{46}$$

while the coordination equivalent fractions are given as:

$$Y_A = \frac{Z_A n_A}{Z_A n_A + Z_B n_B} \tag{47}$$

$$Y_B = \frac{Z_B n_B}{Z_A n_A + Z_B n_B} \tag{48}$$

A correction is added to configurational entropy since it is no longer given by the random mixing assumption from Eq. 17; therefore, S^{mix} becomes:

$$
\begin{aligned}
S^{mix} = -R \Bigg[&(n_A \ln X_A + n_B \ln X_B) + \left(n_{AA} \ln \frac{X_{AA}}{Y_A^2} + n_{BB} \ln \frac{X_{BB}}{Y_B^2} \right. \\
&\left. -n_{AB} \ln \frac{X_{AB}}{2Y_A Y_B} \right) \Bigg]
\end{aligned}
\tag{49}
$$

The quantities Y_A^2, Y_B^2, and $2Y_A Y_B$ represent the probability of finding AA, BB, an AB, respectively, and this corresponds to pair fractions from random mixing. The pair fractions, X_{ij}, are bound by the law of mass action relation:

$$
\frac{X_{AB}^2}{X_{AA} X_{BB}} = 4e^{\left(\frac{-\Delta G_{AB}}{RT} \right)}
\tag{50}
$$

If $\Delta G_{AB} = 0$, each pair is associated with the same energy and there is no tendency for ordering. When this happens $X_{AA} = Y_A^2$, $X_{BB} = Y_B^2$, and $X_{AB} = 2Y_A Y_B$ and Eq. 49 reduces to Eq. 17 for S^{id}.

Finally, ΔG_{AB} can be expanded in terms of coordinate equivalent fractions, for example, using a Redlich-Kister polynomial, in order to achieve a good fit to experimental data.

For multicomponent quasichemical models, assessed binary subsystems are used to construct higher order representations. The coordination number is compositionally dependent via:

$$
\frac{1}{Z_i} = \frac{1}{2n_{ii} + \sum\limits_{j \neq i} n_{ij}} \left(\frac{2n_{ii}}{Z_{ii}^i} + \sum\limits_{j \neq i} \frac{n_{ij}}{Z_{ij}^i} \right)
\tag{51}
$$

The composition of maximum SRO for each binary subsystem is determined by the ratio $\left(Z_{m,n}^m / Z_{n,m}^n \right)$. The multicomponent analogue to Eqs. 44 and 45 can be written as:

$$
Z_i n_i = 2n_{ii} + \sum\limits_{j \neq i} n_{ij}
\tag{52}
$$

Combining Eqs. 51 and 52 results in the relation:

$$
n_i = \frac{2n_{ii}}{Z_{ii}^i} + \sum\limits_{j \neq i} \frac{n_{ij}}{Z_{ij}^i}
\tag{53}
$$

Eqs. 46–48 can be generalized as

$$
X_{ij} = \frac{n_{ij}}{\sum\limits_{k} n_{kk} + \sum\limits_{k} \sum\limits_{l=k+1} n_{kl}}
\tag{54}
$$

$$X_i = \frac{n_i}{\sum_k n_k} \tag{55}$$

$$Y_i = \frac{Z_i n_i}{\sum_k Z_k n_k} \tag{56}$$

The Gibbs energy of solution and the configurational entropy terms now become:

$$G = \sum n_i {}^\circ G_i - T S^{config} + \sum_i \sum_j \left(\frac{n_{ij}}{2}\right) \Delta G_{ij} \tag{57}$$

$$S^{config} = -R \left[\sum n_i \ln X_i + \sum_i n_{ii} \ln \frac{X_{ii}}{Y_i^2} + \sum_i \sum_{j=i+1} n_{ij} \ln \frac{X_{ij}}{2Y_i Y_j} \right] \tag{58}$$

The technique of predicting multicomponent thermodynamic behavior by using assessed binary and/or ternary subsystems can be applied since ΔG_{ij} is a function of two body interactions only.

5.6 Excess Functions

The random mixing assumptions form the base for Eq. 28, 34, 38, 41, and 57. In order to model observed thermodynamic behavior, an ^{xs}G term is included. An important part of modeling efforts is the expedient choice of an excess function that facilitates extrapolation from binary and ternary subsystems to multicomponent mixtures. For this reason, ^{xs}G is generally chosen as a function of binary interactions only. There are more sophisticated representations for ^{xs}G, such as the three-suffix Margules and Whol's equations that are applied to liquid regular solution models, which likely better reproduce observed experimental behavior. However, these expressions suffer from the fact that interaction parameters between all components starting from binary up to the formal order of the system are included. In general, for practical purposes, the ^{xs}G equation should be chosen such that it is defined in terms of binary components or truncated to include only ternary interactions. This allows higher order systems to be extrapolated from binary and/or ternaries and therefore minimizes the experimental effort required to fit the adjustable parameters and describe the system. Higher order interactions are included as needed. This is known as the Muggianu method.

6 Data Sources

Tabulated data have been developed based on experimental measurements spanning several decades. For pure substances, that is the elements, stoichiometric compounds and their compositionally equivalent liquids, and molecular gases, good resources exist such as *NIST-JANAF Thermochemical Tables* (1998) and *Thermochemical Data of Pure Substances* (Ihsan 1995). The data are presented as tabulated Gibbs free energy, enthalpy, entropy, and heat capacity values in increments of temperature. Care must be taken to ensure consistency of standard states. For example, it is customary to use either 1 bar or 1 atm pressure as a reference; there is therefore the potential for small but important errors that can influence calculations if data are mixed between databases that use differing reference conventions.

Perhaps the international standard for thermochemical reference data and solution descriptions is that of the Scientific Group Thermodata Europe or SGTE (SGTE Substances Database Version 2). The SGTE is a well-established and respected organization. Their work is highly cited particularly within the CALPHAD (CALculation of PHAse Diagrams) community. The SGTE continually evaluates and revises the values within the existing database as needed and updates it with new species and phases when new studies identify the need for their inclusion. One limitation is the solution phase models are generally restricted to binary systems.

The Thermodynamic of Advanced Fuels-International Database (TAF-ID) program managed by the Nuclear Energy Agency (NEA) within the Organization of Economic Cooperation and Development (OECD) is an endeavor to model highly complex nuclear fuel systems with burn-up, for example, by including fission and transmutation products. The TAF-ID effort uses the CALPHAD method (detail in Sect. 7.2), whereby fundamental unaries, binaries, and ternaries subsystems are first assessed and then combined to simulate the thermodynamic behavior of multicomponent systems. An international team of researchers and engineers are compiling and assessing the available information on the most significant compounds and solution phases in order to define the multicomponent interaction behavior. Data are either mined from the open literature and checked for consistency or generated through carefully crafted experimental measurements within existing national programs. Other information that may be cost prohibitive or difficult to assess experimentally may possibly be determined from first principles calculations. Advances in theory and computational capabilities allow for calculating thermodynamic properties in this way, for example, lattice stabilities using Density Functional Theory (DFT). Estimation techniques are also useful and can be remarkably accurate in many cases. One example is the Neumann-Kopp rule, postulated more than 140 years ago. It states "*Each element (in the solid state) has essentially the same specific or atomic heat in compounds as it has in the free state*" (*Leitner* et al. 2010). Neumann-Kopp is often modified, for example, computing the overall heat capacity of a multicomponent compound by summing the individual heat capacities of constituent simpler compounds in appropriate ratios that result in the correct

stoichiometry for the complex phase of interest. This has been done with some success for oxides (Leitner et al. 2010). The enthalpy of formation of analogue compounds often follows a linear relationship with atomic number (Konings 2012). Other trends can be identified through systematic evaluation. For example, the standard state entropy and enthalpy of formation for a series of stoichiometric oxides in the Pr-O system show a linear relationship with composition (McMurray 2016). Generalized methods for particular systems are discussed at length by Kubaschewski et al. (1993) and Spencer (1998).

7   Computational Thermodynamics

The essence of computational thermodynamics (CT) is the use of computer codes and numerical techniques to find the constrained global minimum of the total Gibbs energy of a heterogeneous system. This results in an assemblage of stable phases, their elemental composition, and their relative amounts. At the heart of this technique is the Gibbs energy models discussed earlier, many of which are quite sophisticated which, ironically, has restricted successful implementation. For example, the sublattice model has existed for quite some time but use was not practical until the advent of affordable high speed computing and the development of robust minimization routines. For this reason, CT represents a significant scientific step forward by facilitating the thermodynamic representation of multicomponent systems with a high order of possible phase assemblages using complex liquid and crystalline models. The results of CT equilibrium calculations are currently used to construct unary, binary, and ternary phase diagrams and provide thermodynamic properties and potentials that are vital for understanding material behavior.

7.1   Phase Diagrams

Phase diagrams are graphical representations of the thermodynamically distinct phases that are stable at equilibrium. A unary phase diagram is completely defined by two natural variables, usually T,P. An example for H_2O is given in Fig. 1. For a binary system, a complete diagram requires the use of three dimensional (3-D) space since it is necessary to define the system in terms of one independent compositional variable in addition to the two natural variables. For convenience, a temperature composition (T-c) plot of a two component system is sectioned along an isobar, typically 1 atm, to construct a two-dimensional (2-D) T-c binary phase diagram. An example for the Cu-Ni system is given in Fig. 2. Ternaries are even more complex since it is necessary to specify two compositional and two natural variables to define the system, thus requiring the use of four dimensional (4-D) space to completely represent the phase equilibria. An isobaric section reduces this to 3-D that can be represented by a prism as shown in Fig. 3. Here, temperature is plotted along the vertical axis and the base of the prism is formed by an equilateral compositional triangle, the so-called Gibbs triangle.

Difficulties arise when using 3-D representations on 2-D media; it is therefore useful to reduce ternary phase space plots to 2-D by sectioning. If a slice is excised along the horizontal plane of the 3-D space, i.e., the system is at a constant temperature, then an isothermal diagram results such as that shown for the Mo-Si-C system in Fig. 4. These are typically the most useful since they
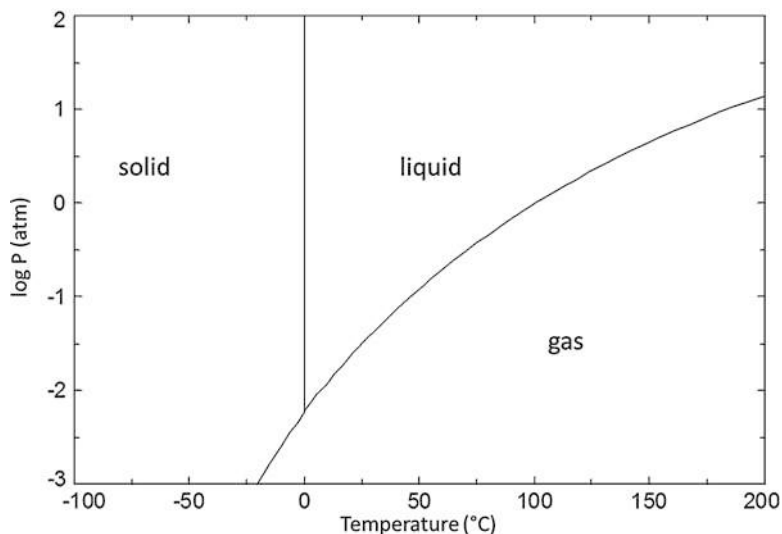
Fig. 1   Temperature-pressure phase diagram for H_2O in the vicinity of the triple point
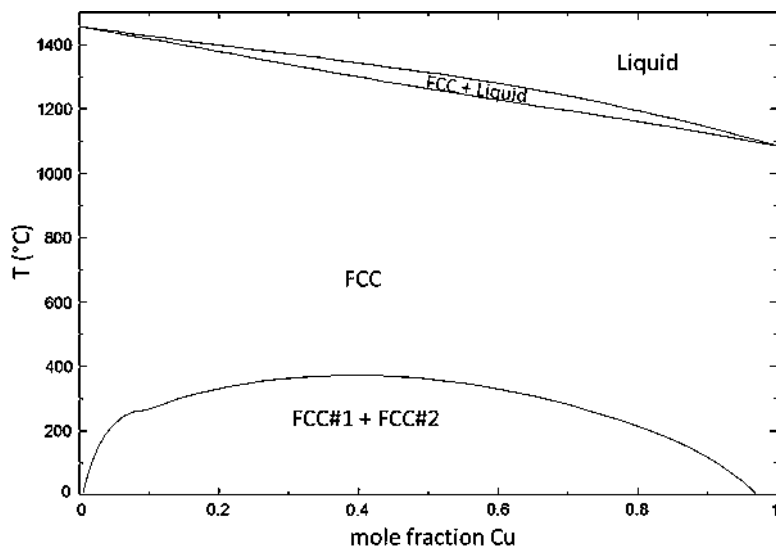
Fig. 2   The Cu-Ni binary phase diagram

Fig. 3 The constant pressure prism representing temperature along the vertical axis and composition within the so called Gibbs triangle
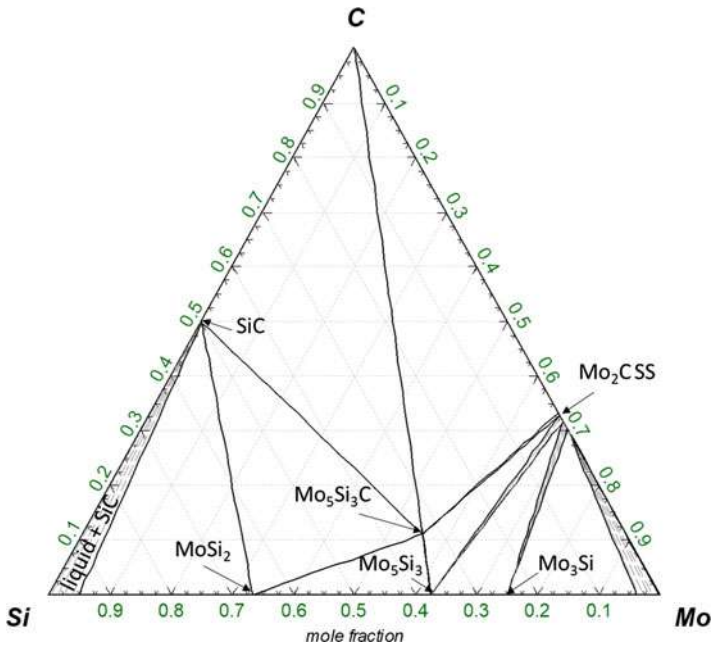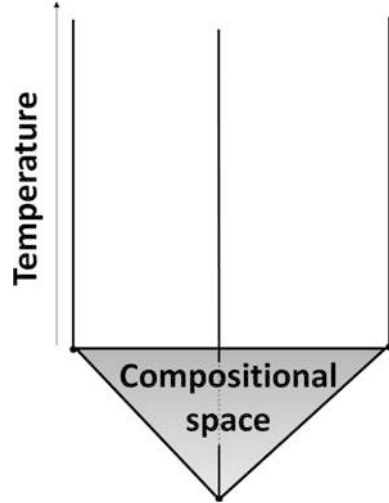
Fig. 4 Ternary isothermal section of the Mo-Si-C at 1590 °C

allow for compositional analysis while the alternatives do not (Campbell 2012). An isoplethal diagram follows a constant variable, usually composition of one (or more) component. An example of the isopleth is shown in Fig. 5.

Higher order phase diagrams are extremely difficult to visualize and plot and are therefore not very utilitarian; however, one technique is to fix the compositions of
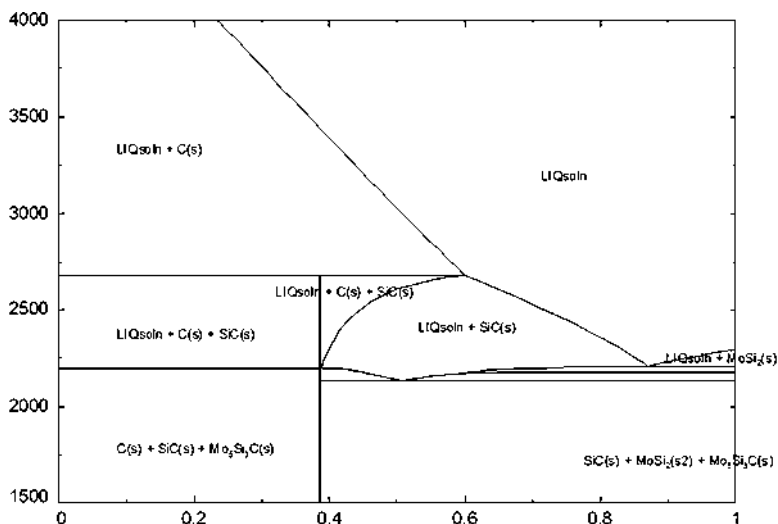
Fig. 5 Pseudobinary phase diagram of the Mo-Si-C system along the C-MoSi$_2$ isoplethal section

enough of the components to generate a pseudo-ternary isotherm or a pseudo-binary isopleth for 2-D renderings. While most systems do not attain equilibrium, phase diagrams are important since they show where nature would like to move the system given enough time to evolve, with increasing temperature generally moving systems more rapidly to equilibrium. They are a core part of developing thermodynamic models that can be used to compute chemical potentials that are fundamental for simulating many kinetic processes such as microstructural evolution based on phase-field methods.

7.2   CALPHAD Methodology

Over 40 years ago, Larry Kaufman and Himo Ansara organized the first CALPHAD (**CAL**culation of **PHA**se **D**iagrams) (Spencer 2008) meeting with the aim of developing a framework for self-consistent modeling of combined thermodynamic properties and phase equilibria – this became known as the CALPHAD method. Since then, CALPHAD has become more of a movement than a method. There is a membership organization growing in both the number of scientists employing the approach and the scope of materials to which the methodology is applied as well as a flagship journal *CALPHAD – Computer coupling of phase diagrams and thermochemistry*. Advances in science and computing performance have facilitated the implementation of theoretical first principles calculations into thermodynamic assessments and the development of sophisticated commercial computational thermodynamic software packages like Factsage and Thermocalc. These achievements have grown from and form a fundamental part of present day CALPHAD.

In the CALPHAD approach, all available data are critically assessed and utilized to build the most accurate and comprehensive representation of the phases in a system. The aim is self-consistent models that predict thermodynamic properties and equilibria to be used for phase diagram construction that can be confidently extrapolated beyond the range of experimental validation. The ultimate goal is the development of databases for specific components that form systems of technological importance that can be extended to include higher order systems ad hoc. This is facilitated by developing models based on the physical and chemical properties of the phases, for example, the crystallography, bonding, order-disorder transitions, and magnetism (CALPHAD 2014). Of course, extensions and extrapolations are just that and a core tenet of the CALPHAD methodology is the periodic re-assessment of systems, a re-examination of the existing data from which they are determined, and new experiments for validation of the models in extrapolated regions; therefore, databases and models should be constantly updated and improved.

The use of first principles methods, particularly density functional theory (DFT) (Lukas et al. 2007), has come to play an increasingly important role in CALPHAD modeling to supplement efforts to develop thermodynamic data or to determine thermodynamic values that are difficult or otherwise impossible to obtain experimentally, such as properties of meta-stable or unstable phases. The structure of crystalline materials should inform the selection or development of models that can represent the defects and site occupancy of the species comprising the solid; this information can generally be determined from neutron and X-ray diffraction techniques. Once a model is chosen to represent the phase, thermodynamic data determined from experimental measurements or ab initio simulations are used to optimize the adjustable parameters of models using a computational software package.

7.3   Lattice Stabilities

A fundamental CALPHAD concept is *lattice stability* (Saunders and Miodownik 1998). It comes from the idea that the Gibbs energy descriptions can be represented for all elements in every phase of a system, that is, it can be extrapolated into regions of instability, expressed graphically in Fig. 6. Lattice stabilities can be obtained by traditional CALPHAD techniques or from first principles calculations such as DFT.

7.3.1   Lattice Stabilities from Extrapolation

Most CALPHAD lattice stabilities can be deconvoluted into H and S components as follows:

$$G^{\alpha \to \beta} = H^{\alpha \to \beta} - T S^{\alpha \to \beta} \tag{59}$$

Fig. 6 Gibbs curves from the ideal solution model as a function of composition for the α and β phases in the hypothetical a-b and a-c systems. The dashed lines represent "extrapolations" into the unstable phase regions. The difference between the extrapolated end points and the pure stable phase is the lattice stability. Note that the lattice stability for pure *a* in both the *a-b* and *a-c* binaries are the same, i.e., the $G^{\alpha \to \beta}$ value is independent of the system
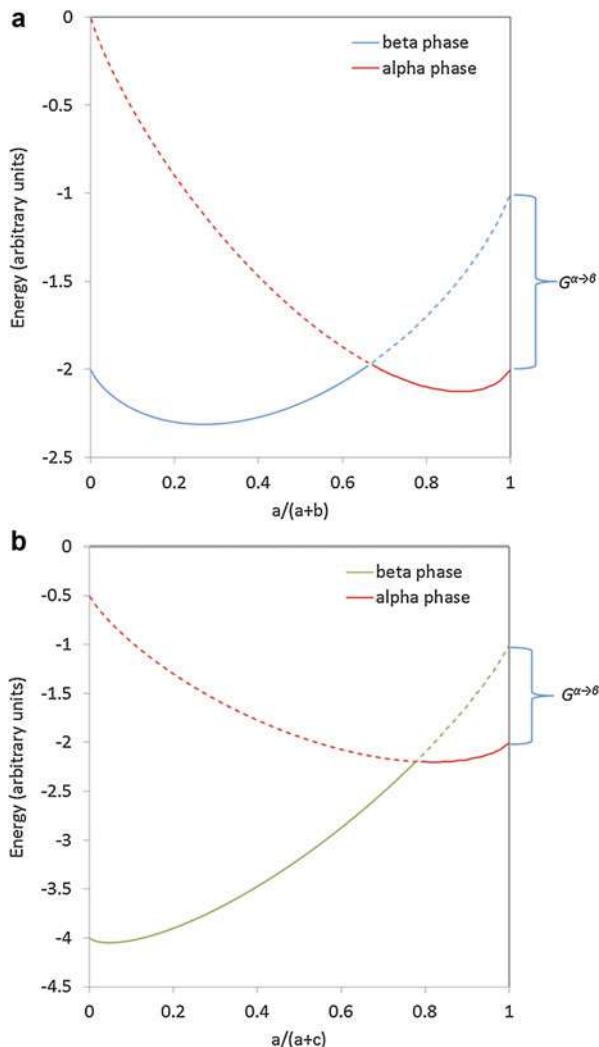
where $G^{\alpha \to \beta}$ is the lattice stability and $H^{\alpha \to \beta}$ and $S^{\alpha \to \beta}$ are the transition enthalpy and entropy for unstable β relative to the stable α phase, respectively. One common and convenient approach assumes the c_p to be the same for both phases; therefore, G^{β} can be obtained by adjustment to G^{α} with an $a-bT$ term. One technique is to estimate the entropy and hypothetical melting temperature of the unstable phase; an example is seen in Fig. 7 where the melting temperature is determined by extrapolation shown as the dashed line. Using the following relations:

$$H_{Cr}^{L} - H_{Cr}^{bcc} = S_{Cr}^{L \to bcc} \left(T_f^{bcc} - T \right) \tag{60}$$

Fig. 7 Melting temperature extrapolations (dashed line) of bcc Cr in the Cr-Ni and Cr-Pt systems from (Lukas et al. 2007)
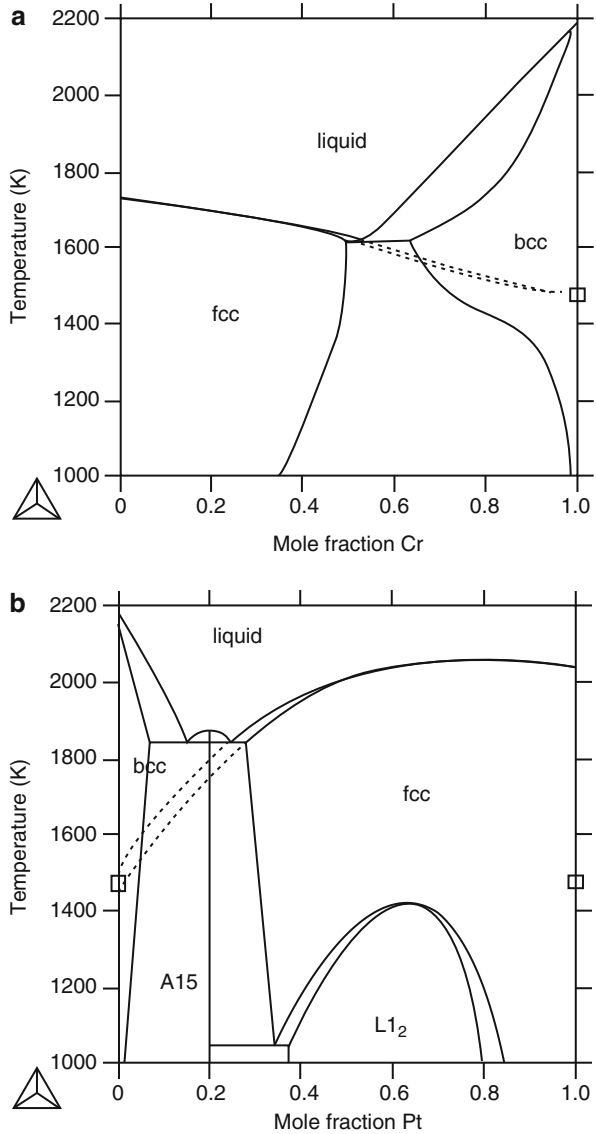
$$H_{Cr}^{L} - H_{Cr}^{fcc} = S_{Cr}^{L \to fcc} \left(T_{f}^{fcc} - T \right) \qquad (61)$$

where T_f is the melting temperature and $H_{Cr}^{fcc \to bcc}$ can be obtained by subtracting Eq. 60 from 61. The credibility of this approach is established by multiple phase relations that yield self-consistent extrapolated values (Lukas et al. 2007). Another

technique is the direct extrapolation of the Gibbs energy for a component in solution in two isostructural parent materials. For consistency, the same type of model is be used to describe the behavior for both phases and requires the end-member energies of the solute must be consistent between systems. This is illustrated graphically in Fig. 6. Finally, further confidence is established if the values for $H^{\alpha \to \beta}$ are consistent with 0 K total energy differences calculated from DFT methods.

7.3.2 Lattice Stabilities from First Principles Calculations

Ideally, DFT can provide the relative enthalpy for various crystal structures and thus the $H^{\alpha \to \beta}$ component of the lattice stability. This is accomplished by computing the total energy, E, at absolute zero for an assemblage of particles by solving the time-independent Schrödinger equation using simplifying assumptions and approximations (Lukas et al. 2007). Equations 62 and 63 show the lattice stability is, to a first approximation, the difference in E values between two distinct crystallographic arrangements.

$$\Delta E = E^{\alpha} - E^{\beta} \tag{62}$$

where

$$\Delta E \approx \Delta H \tag{63}$$

Eq. 63 is assumed to be valid since at 1 atm pressure $P\Delta V \approx 0$. These results can then be compared to estimated $H^{\alpha \to \beta}$ based on extrapolations, used directly to determine the $S^{\alpha \to \beta}$ for stable phases, or serve as an initial value for applying lattice stability in an optimization procedure.

7.4 Computational Thermodynamic Codes

In the 1950s, White et al. (2004) developed an algorithm for minimization of the Gibbs energy of multicomponent gas mixture for computer implementation using a method of steepest descent. While techniques have evolved over the years for thermodynamic equilibrium calculations for ever more complex solution phases and multicomponent systems, most are variations of the method of White et al. and all are based on one fundamental concept, which is attaining the global minimum of the total Gibbs energy satisfying the condition of conservation of mass (Piro 2011).

The two most robust and widely cited computational thermodynamic software packages in the open literature are Factsage® and Thermocalc®. Factsage uses an equilibrium solver developed by Eriksson that was the basis of his original SOLGASMIX code (Eriksson 1971, 1975; Eriksson and Rosen 1973; Piro 2011), while the Thermocalc code is based on an algorithm after Hillert and Jansson (JO Andersson et al. 2002). The fundamental operation of both programs is essentially the same allowing for versatile thermodynamic properties output and phase diagram construction. They also include an optimization module for fitting adjustable

parameters of Gibbs functions and solution models using thermodynamic values and phase equilibria data. Both Factsage and Thermocalc have an established record of use for modeling of nuclear materials systems and are capable of handling at least 48 and 20 components respectively (Piro 2011). Other commercial thermodynamic packages include Thermosuite, MTDATA, PANDAT, HSC, and MALT to name a few (Konings 2012). The software facilitate the use of the sophisticated models needed to describe many of the solution phases discussed earlier.

8 Fuel Systems

8.1 Ceramic Fuels and Ionic Crystalline Phases

Uranium monocarbide (UC_{1-x}), UC_{2-x}, and UN are of interest as advanced fuels. Urania, or $UO_{2 \pm x}$, is the most widely used commercial nuclear fuel and is interesting because of its complex oxygen chemistry. The U cation exhibits 3+,4+,5+, and 6+ oxidation states. This, along with the structure of the fluorite crystal, allows for oxygen interstitials and vacancies in $UO_{2 \pm x}$ and permit formations of a wide oxygen homogeneity as can be seen in Fig. 8. While UC_{1-x}, UC_{2-x}, and UN are all modeled with CEF, the CEF for $UO_{2 \pm x}$ developed by Guéneau et al. (2002, 2011) is used here as an example for modeling complex crystalline phases.

As can be seen in Fig. 9, the CEF provides a description of the oxygen potential, heat capacity, and point defect chemistry of $UO_{2 \pm x}$. These are critical values for understanding the behavior of the fuel in a reactor. In particular, oxygen redistributes and affects the local thermal conductivity of the phase creating a temperature-composition feedback loop. The CEF cannot, however, address the likely clustering

Fig. 8 Computed and experimental phase relations in the vicinity of the $UO_{2 \pm x}$ phase (Guéneau et al. 2011) Reprinted from The Journal of Nuclear Materials, 419, Gueneau et al., Thermodynamic modeling of advanced oxide and carbide nuclear fuels: Description of the U–Pu–O–C systems, 145–167, Copyright (2011), with permission from Elsevier https://www.sciencedirect.com/science/journal/00223115?sdc=1
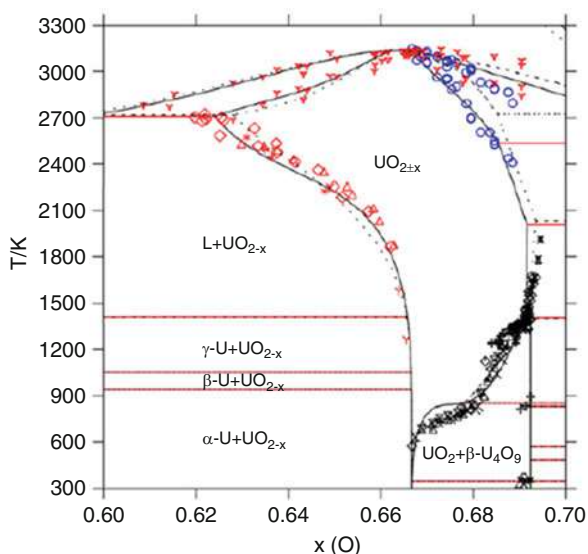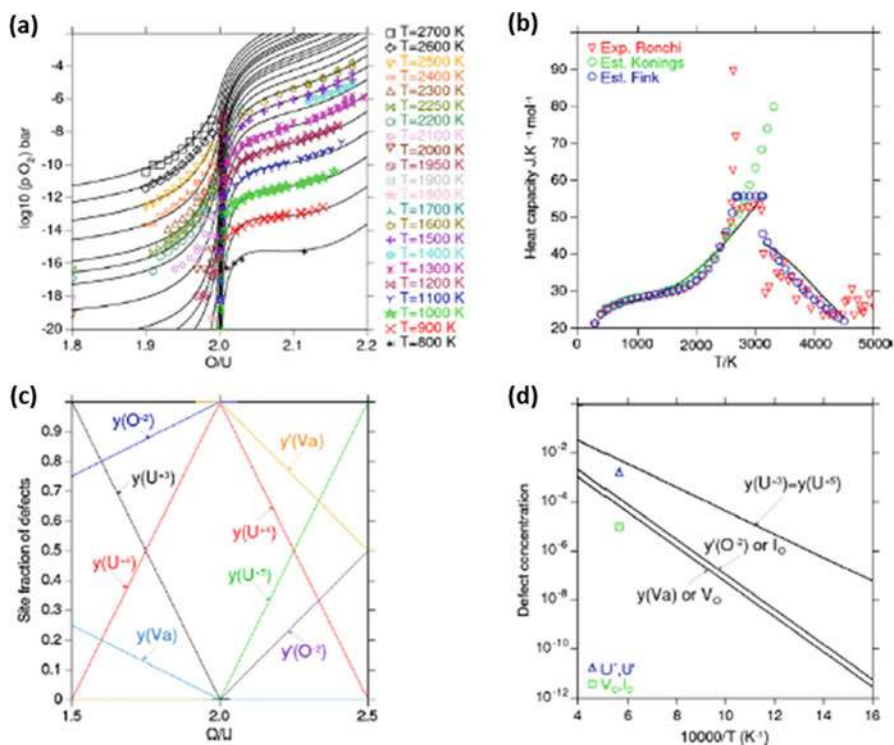
Fig. 9 Plots of **(a)** logarithm of oxygen potential versus oxygen to uranium ratio (O/U); **(b)** heat capacity versus temperature; **(c)** site fraction of defects versus O/U and; **(d)** tetrahedral oxygen vacancies (Vo), octahedral interstitials (Io), and electronic defects ($y(U^{+3}) = y(U^{+5})$) computed using the CEF for $UO_{2 \pm x}$. Reprinted from The *Journal of Nuclear Materials*, 419, Gueneau et al., Thermodynamic modeling of advanced oxide and carbide nuclear fuels: Description of the U–Pu–O–C systems, 145–167, Copyright (2011), with permission from Elsevier https://www.sciencedirect.com/science/journal/00223115?sdc=1

of oxygen interstitials in the hyperstoichiometric oxide (D. A. Andersson et al. 2009; Brincat et al. 2015) and thus the inclusion of such defects awaits further model development.

8.2   Metallic Fuels and Metallic Phases

It is standard practice to use sublattice models like the CEF to represent interstitial solutions. One example is for O dissolution in hcp Zr. The crystal structure influences the degree of solubility of interstitials and this is captured with the CEF. For example, there is one octahedral site in both the fcc (face centered cubic) and hcp (hexagonal close packed) metals. However, for the hcp structure, there are shorter distances between these sites. It is therefore generally accepted that only one-half

can be occupied (Grundy et al. 2001). The sublattice model $(M)_1(I)_{0.5}$ where M represents the metal and I the interstitial species well reproduces the solubility limit of O in both hcp Ti (Waldner and Eriksson 1999) and Zr (Wang et al. 2004).

When mixing occurs on only one sublattice and the others are either fully occupied or totally vacant, the CEF reduces to the regular solution model, sometimes called the RKMP because of the Redlich-Kister-Muggiano polynomial formalism based on Eqs. 29 and 30 that is used to define the interaction terms. The U-Th system was assessed by Li et al. (2009) and the binary phase diagram is reproduced in Fig. 10. Here, the CEF is used for the crystalline metallic solutions but with vacancies exclusively on the anion sublattice. While this is tantamount to simply using the RKMP, it is good practice to apply the CEF as it allows extension of the models into higher order component space. For example, in the ternary U-Th-O or U-Th-C systems, the O and/or the C would likely dissolve and be accommodated on the interstitial sites.

It is worth pointing out that the database focused on metallic fuels are few with limited thermochemical information, for example that of Kurata (Kurata 1999). Another, dedicated to Zr alloys of interest for nuclear applications, is ZIRCOBASE (Dupin et al. 1999) concentrating on Zr containing binary systems and some ternaries
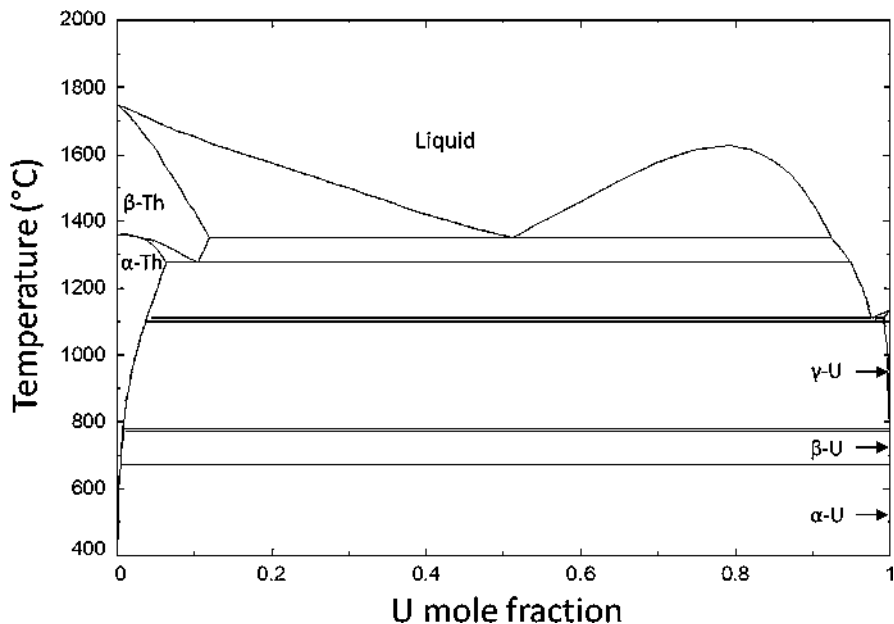
Fig. 10 Computed Th-U binary based on the models from (Li et al. 2009)

8.3   Ionic and Metallic Liquids

The MQM and the two-sublattice ionic liquid model (TSLM) are extensively used within the CALPHAD community to represent melts like ionic liquids such as molten salts. Of the two, the MQM incorporates more physics and better describes the configurational entropy by accounting for nearest neighbor short range ordering, as described in detail in Sect. 5.5. However, the TSLM is convenient to use and can do as well as the MQM in representing the Gibbs energy of both ionic and metallic melts. In the latter case, the TSLM reduces to the regular solution model. Molten salt phases are most often represented with the MQM. Modeling efforts based on the MQM for nuclear-related fluoride systems have been significant (Capelli et al. 2015; Capelli et al. 2014) and efforts are expanding to use the MQM for chloride based molten salt fuels as well.

While ionic liquids can certainly be represented with the MQM, the TSLM is also commonly used as is the case for the U-O system (Guéneau et al. 2002; Guéneau et al. 2011). For metallic melts, the RKMP is sufficient. However, just as for purely substitutional crystalline solutions, the TSLM is preferred to allow for extension into multicomponent space. A caveat is, for the TSLM both FactSage and Thermocalc currently have issues representing a liquid metal without a constituent on the anion sublattice; therefore, both TSLM and RKMP models are sometimes included in multicomponent assessments as a work around.

9   Outlook

Thermodynamics of nuclear fuels is an important and continuing area of intense research. It has broad based utility in developing new materials, designing of processes, and in predicting and optimizing reactor performance. The current state of the science can be summarized as: Much has been accomplished, but more is needed. Understanding the fundamental thermochemistry of fuel systems with burnup continues to be a challenge for two reasons. First, including actinides makes measurements required for model development costly and difficult. Second, fuel systems become highly multicomponent and inhomogeneous with burnup and are constantly evolving from beginning to end of life. To attack these challenges, over the past decade efforts have been devoted to developing multiscale-, multiphysics-based computational capabilities that couple thermodynamic inputs to transport, phase field, and continuum mechanics simulations to predict the chemical, thermal, and structural behavior of reactor core components (Devanathan et al. 2010). This is the future of fuel performance modeling. At the beating heart is thermodynamics.

References

Andersson J, Helander T, Höglund L, Shi P, Sundman B (2002) Thermo-Calc & DICTRA, computational tools for materials science. Calphad 26(2):273–312

Andersson DA, Watanabe T, Deo C, Uberuaga BP (2009) Role of di-interstitial clusters in oxygen transport in UO_2+x from first principles. Phys Rev B 80:060101. https://doi.org/10.1103/PhysRevB.80.060101

Barin I (1995) Thermochemical data of pure substances. VCH, New York

Brincat NA, Molinari M, Parker SC, Allen GC, Storr MT (2015) Computer simulation of defect clusters in UO_2 and their dependence on composition. J Nucl Mater 456:329–333. https://doi.org/10.1016/j.jnucmat.2014.10.001

CALPHAD. (2014). http://www.thermocalc.com/academia/researchers/calphad/. Accessed

Campbell FC (2012) Phase diagrams: understanding the basics. ASM International, Materials Park

Capelli E, Benes O, Konings RJM (2014) Thermodynamic assessment of the LiF-NaF-BeF_2-ThF_4-UF_4 system. J Nucl Mater 449(1–3):111–121. https://doi.org/10.1016/j.jnucmat.2014.03.009

Capelli E, Benes O, Konings RJM (2015) Thermodynamic assessment of the LiF-ThF_4-PuF_3-UF_4 system. J Nucl Mater 462:43–53. https://doi.org/10.1016/j.jnucmat.2015.03.042

Chang R (1998) Chemistry, 6th edn. McGraw-Hill, NY

De Groot SR, Mazur P (2013) Non-equilibrium thermodynamics. Courier Corporation. Dover Publications, Inc., New York

DeHoff R (2006) Thermodynamics in materials science. CRC Press, Boca Raton

Devanathan R, Van Brutzel L, Chartier A, Guéneau C, Mattsson AE, Tikare V et al (2010) Modeling and simulation of nuclear fuel materials. Energy Environ Sci 3(10):1406–1426

Dupin N, Ansara I, Servant C, Toffolon C, Lemaignan C, Brachet J (1999) A thermodynamic database for zirconium alloys. J Nucl Mater 275(3):287–295

Eriksson G (1971) Thermodynamics studies of high temperature equilibria .3. SOLGAS, a computer program for calculating composition and heat condition of an equilibrium mixture. Acta Chem Scand 25(7):2651. https://doi.org/10.3891/acta.chem.scand.25-2651

Eriksson G (1975) Thermodynamic studies of high-temperature equilibria. XII. SOLGASMIX, a computer-program for calculation of equilibrium compositions in multiphase systems. Chem Scr 8(3):100–103

Eriksson G, Rosen E (1973) Thermodynamic studies of high-temperature equilibria. VIII. General equations for calculation of equilibria in multiphase systems. Chem Scr 4(5):193–194

Fowler RH, Guggenheim EA (1949) Statistical thermodynamics (second impression with corrections.). Cambridge University Press, London

Grundy AN, Hallstedt B, Gauckler LJ (2001) Thermodynamic assessment of the lanthanum-oxygen system. J Phase Equilib 22(2):105–113

Guéneau C, Baichi D, Labroche C, Chatillon B, Sundman B (2002) Thermodynamic assessment of the uranium–oxygen system. J Nucl Mater 304:161

Guéneau C, Dupin N, Sundman B, Martial C, Dumas JC, Gosse S et al (2011) Thermodynamic modelling of advanced oxide and carbide nuclear fuels: description of the U–Pu–O–C systems. J Nucl Mater 419:145–167

Hillert M, Jansson B, Sundman B (1985) A two-sublattice model for molten solutions with different tendency for ionization. Metall Trans A 16(1):261–266

Konings RJM (2012) Comprehensive nuclear materials. Elsevier, Amsterdam

Kubaschewski O, Alcock CB, Spencer P (1993) Materials thermochemistry. Pergamon Press, Oxford New York

Kurata M (1999) Thermodynamic assessment of the Pu-U, Pu-Zr, and Pu-U-Zr systems. Calphad 23(3–4):305–337

Leitner J, Voňka P, Sedmidubský D, Svoboda P (2010) Application of Neumann–Kopp rule for the estimation of heat capacity of mixed oxides. Thermochim Acta 497(1):7–13

Li Z, Liu X, Wang C (2009) Thermodynamic modeling of the Th–U, Th–Zr and Th–U–Zr systems. J Alloys Compd 476(1):193–198

Lindemer TB, Sutton AL Jr (1988) Study of nonstoichiometry of <$U_{1-z}Gd_zO_{2\pm x}$>. J Am Ceram Soc 71(7):553–561

Lukas HL, Fries SG, Sundman B (2007) Computational thermodynamics: the Calphad method. Cambridge University Press, Cambridge

McMurray JW (2016) Thermodynamic assessment of the Pr-O system. J Am Ceram Soc 99(3):1092–1099

NIST-JANAF (1998) Thermochemical Tables, NIST Standard Reference Database 13, 1998, https://doi.org/10.18434/T42S31, https://janaf.nist.gov/

Pelton AD, Chartrand P (2001) The modified quasi-chemical model: part II. Multicomponent solutions. Metall Mater Trans A 32(6):1355–1360

Pelton AD, Degterov S, Eriksson G, Student CRPD, Dessureault Y (2000) The modified quasi-chemical model I – binary solutions. Metall Mater Trans B 31(4):651–659

Piro M (2011) Computation of thermodynamic equilibria pertinent to nuclear materials in multi-physics codes. Royal Military College of Canada, Kingston

Prausnitz JM, Lichtenthaler RN, de Azevedo EG (1998) Molecular thermodynamics of fluid-phase equilibria. Pearson Education, Upper Saddle River

Redlich O, Kister AT (1948) Algebraic representation of thermodynamic properties and the classification of solutions. Ind Eng Chem 40(2):345–348. https://doi.org/10.1021/ie50458a036

Saunders N, Miodownik AP (1998) CALPHAD (Calculation of phase diagrams): a comprehensive guide: a comprehensive guide. Elsevier/Thermotech, Guildford

SGTE Substances database https://www.sgte.net/en/thermochemical-databases

Smith EB (2013) Basic chemical thermodynamics. World Scientific, Singapore

Smith JM, Van Ness HC, Abbott MM (1996) An introduction to chemical engineering thermodynamics. McGraw-Hill, New York

Spencer P (1998) Estimation of thermodynamic data for metallurgical applications. Thermochim Acta 314(1):1–21

Spencer P (2008) A brief history of CALPHAD. Calphad 32(1):1–8

Temkin M (1945) Mixtures of fused salts as ionic solutions. Acta Phys Chim USSR 20:411

Van Wylen GJ, Sonntag RE, Borgnakke C, Wiley J (1994) Fundamentals of classical thermodynamics. Wiley, New York

Waldner P, Eriksson G (1999) Thermodynamic modelling of the system titanium-oxygen. Calphad 23(2):189–218

Wang C, Zinkevich M, Aldinger F (2004) On the thermodynamic modeling of the Zr–O system. Calphad 28(3):281–292

White WB, Johnson SM, Dantzig GB (2004) Chemical equilibrium in complex mixtures. J Chem Phys 28(5):751–755

Willard GJ (1961) The scientific papers of J. Willard Gibbs. Dover, New York

Part XIV
Radiation Damage

Modeling of Radiation Damage in Materials: Best Practices and Future Directions

96

Kai Nordlund and Michael P. Short

Contents

Abstract

In this section of the *Handbook of Materials Modeling*, modern methods to model radiation damage in materials are considered. This introductory chapter puts the topic into context, briefly overviews the multiscale nature of radiation effects, and surveys most important methods used to advance this field of physics. Emphasis is given to advances from the past 10 years, highlighting specific bodies of work, some of which are reviewed in this section, which have advanced both our scientific understanding of and best practices for simulating radiation damage. This section differs somewhat from that of Devanathan and Marian entitled "Nuclear Materials," in that this section focuses more on modeling techniques for radiation damage as applied to pure materials, not necessarily nuclear materials. However, much can be learned from reading the shared perspectives in both

K. Nordlund
Department of Physics and Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland
e-mail: kai.nordlund@helsinki.fi

M. P. Short (✉)
Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: hereiam@mit.edu

sections, as many similar topics are addressed from different points of view. The more specific chapters in this part are introduced in this overview, with emphasis on predicted directions in the field for the next 10 years.

1 Introduction: A Historical Perspective

The history of the human race is closely linked with developments in materials processing. Societal development is routinely divided into the Stone Age, Bronze Age, and Iron Age, and sometimes the last half century is characterized as the silicon age. The varieties of materials processing that enabled the first three ages are of course traditional and can be characterized in terms easily understood to a layman: chiseling, cracking, melting, cooling, and so on. In a physicist's terminology, these can be characterized as being equilibrium or near-to-equilibrium processes.

In this range of ages, the silicon age brought with it a fundamentally different way of processing materials: ion irradiation enabled by particle accelerators. Taken into use in the industry in the 1960s, it is an inherently far-from-equilibrium process. In fact, during the very first stages, when atoms ballistically collide two at a time (a process known as a collision cascade), classical thermodynamics does not in fact apply at all. Even the very idea of temperature requires an ensemble of atoms, which does not really exist at this stage. Later on, after about 0.2 ps or so, the initially energetic atoms have slowed down to kinetic energies comparable to the chemical bond energy (still tens of thousands of Kelvin, thermodynamically speaking), and atomic motion becomes many-body in character. At this stage, thermodynamic concepts like temperature and pressure become meaningful (Diaz de la Rubia et al. 1987), and the system can be characterized as in a far-from-equilibrium thermodynamic state. After another 10 ps or so, all atoms have thermalized to temperatures close to that of the surrounding material, and regular thermodynamics and kinetics, such as the heat diffusion equation, can be used to characterize the system.

Although the silicon semiconductor industry is by far the most economically important user of ion irradiation for materials modification, it did not invent the technology. Radiation does of course occur everywhere in nature due to radioactive decay and cosmic radiation. The first reported human-controlled use of irradiation effects for materials modification was in 1852, when Sir Grove built a metal needle in a vacuum chamber and charged it up from a coil (Grove 1852). This caused visible surface modifications, which in retrospect were deduced to be due to rest gas atoms being ionized and hitting the surface, causing sputtering or implantation and a color change associated with that – essentially a very early precursor to the field ion microscope (Mueller and Bahadur 1956).

The experimental and theoretical developments of the study of radiation effects in materials are proceeded by famous scientists like Rutherford, Curie, Bohr, Fermi, and Oppenheimer (Rutherford 1911; Bohr 1913; Fermi and Teller 1947). Initially the efforts were largely focused on understanding the slowing-down mechanisms of the energetic ions and neutrons that were obtained first from radioactive sources,

and later from particle accelerators. Without much exaggeration, one can say that the basic physics of the ion slowing down was understood by the 1960s, when Jens Lindhard (a student of Niels Bohr) and coworkers formulated a coherent theory of the nuclear and electronic stopping powers and their energy deposition in materials (Lindhard et al. 1963). This leads soon to a predictive simulation capacity regarding some aspects of radiation effects, such as calculating within 10% or so the penetration depth profile of keV or MeV ions in amorphous materials (Robinson and Torrens 1974).

The efforts to understand how materials are actually modified by energy deposition from ionizing radiation started to become more systematic around the same time, as nuclear fission reactors became widely used, and the idea to use ion implantation for semiconductor processing arose. However, it turned out that reaching a predictive capability for damage production was a much harder problem than that of ion passage in materials, and the efforts to achieve this understanding are still underway. This collection of chapters in the *Handbook of Materials Modeling* is in essence a description of the state of the art of attacking this complex problem, which even now is far from solved. Similarities will be noted in the part by Devanathan and Marian entitled "Nuclear Materials," though theirs focuses more on actual nuclear materials while this part focuses squarely on model materials to illustrate modeling techniques. Nevertheless, the reader is referred to the part by Devanathan and Marian for different perspectives on many similar modeling issues, as seeing similar high-level concepts from different points of view, or as applied to different materials systems, is often highly instructive.

The chapters in this part are largely written with examples focused on the studies of ion and neutron effects in metals. However, we emphasize that this is not the only regime where the methods are applicable. All the same physics and approaches can also be used to study semiconductors (Chason et al. 1997), ceramics (Meldrum et al. 1998) and carbon-based materials (Krasheninnikov and Nordlund 2010), and to some extent also damage in organic materials (Polvi et al. 2012). The radiation damage in biological materials is, however, quite different in that the damage on the biologically active molecules, such as DNA and proteins, seems to be to a large extent a secondary effect from radicals formed in the water surroundings. Discussion of the last case is beyond the scope of this part.

2   Radiation Effects as a Multiphysics and Multiscale Problem

The evolution of radiation damage in the bulk is an inherent multiphysics and multiscale problem. The physical nature of this evolution is illustrated schematically in Figs. 1, and 2 provides an overview of the length and time scales involved.

Before proceeding with the discussion of this, it is worth noting that not all aspects of radiation effects, however, are multiscale in character. Clear examples of this are the ion penetration depth problem mentioned in the historical overview above and sputtering (energetic particle-induced atom erosion from surfaces) and ion beam mixing (relocation of atoms by irradiation) at low temperatures. Both
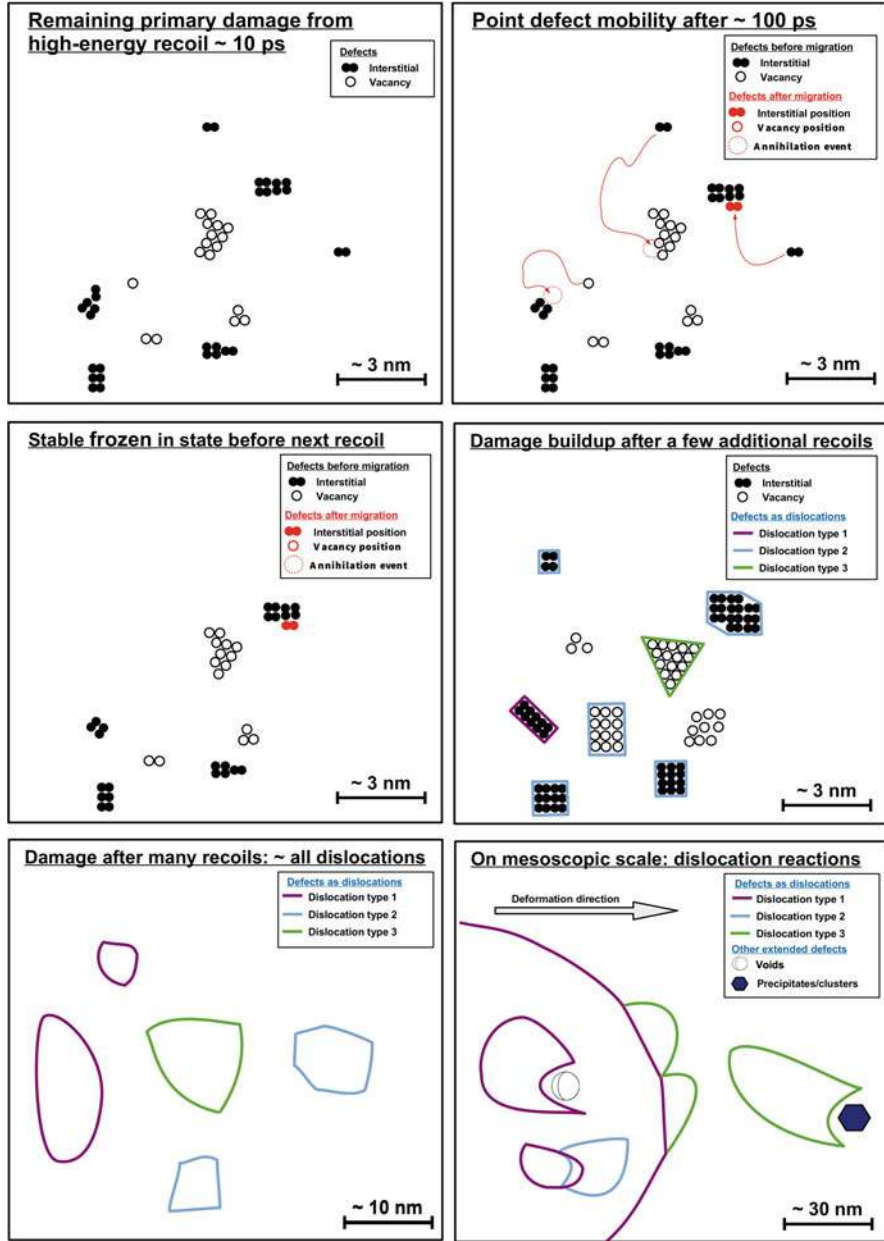
Fig. 1 Graphical illustration of radiation damage production physics
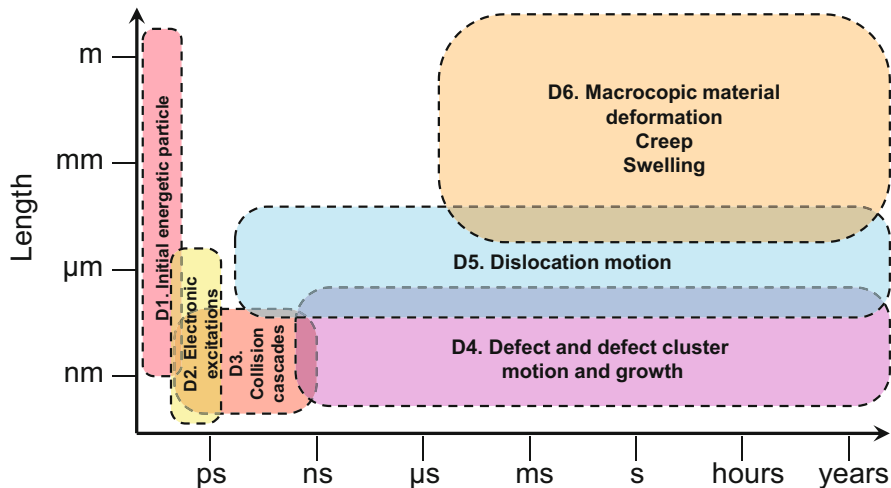
Fig. 2 Multiple levels of physics involved in radiation effects. The codes D1–D6 are used as cross-references in the main text. The box limits in space and time are very rough and should not be taken as any kind of sharp limits

effects are under typical ion or neutron fluxes caused by atom movement in the collisional and heat spike stages of single collision cascades, which are over in less than 10 ps. Hence they in essence do not involve multiscale physics, nor is multiscale modeling needed to describe these effects. Indeed, the phenomena have – at least in a few simple materials – been very well described by a single simulation technique, molecular dynamics (Gades and Urbassek 1995; Nordlund et al. 1998).

In the multiphysics regime, achieving a predictive description level of radiation effects is very challenging and has not been achieved yet. In this chapter, we describe very briefly the different levels of physics known to be involved and, in the next ones, the modeling tools used to study them, pointing out which chapter belongs to which method.

The initiator of a "radiation event" is an energetic (ionizing) particle, **box D1** in Fig. 2. These can, for instance, be fission fragments from radioactive decay of natural or man-made isotopes, neutrons from a nuclear reactor, ions from a particle accelerator, electrons from an electron gun, or photons from a synchrotron. These slow down by different mechanisms involving interaction with the electrons and nuclei in the materials. While the slowing-down mechanisms are beyond the scope of this chapter, the important physics point is that for an individual energetic particle, the slowing down to thermal energies is a very rapid process, occurring on femtosecond to picosecond time scales. Due to the energy transfer, the struck electrons and nuclei can in turn gain high kinetic energies, which in turn leads to material damage.

In this section, we focus on the damage caused by the nuclei that have gained a recoil energy from any kind of initial energetic particle. The damage by electronic excitations (**box D2**) is discussed in a separate chapter by Djurabekova and Nordlund (▶ Chap. 88, "Molecular Dynamics Simulations of Non-equilibrium Systems"). Regardless of the type of the initial particle, after a nucleus in the material has received a recoil kinetic energy, the damage production mechanism is the same. This motivates that the remainder of this chapter is a thorough discussion of damage production by atomic recoils.

After an atom has obtained a recoil energy, higher than the threshold displacement energy of the material (Nordlund et al. 2005), it will collide with atoms in it, giving over some of its kinetic energy to these. Thus these atoms in turn become recoils themselves and can lead to further collisions. The sequence of collisions is called a collision cascade (**box D3**). What is very important to realize, however, is that in dense materials, eventually the different branches of a collision cascade almost always overlap, and the process becomes a set of simultaneous many-body collisions known as a heat spike (Seitz and Koehler 1956). In the heat spikes, the damage produced is not only single Frenkel pairs (pairs of interstitials and vacancies) but can be clustered (Kojima et al. 1989), regions that have undergone a phase transition (Meldrum et al. 1998), or, in the case of amorphizing materials, an amorphous zone (Ruault et al. 1984).

The collision cascades cool down rapidly (the cooling rate of heat spikes in metals is of the order of 10^{15} K/s, which is an amusing number from a macroscopic point of view). After they have cooled down to the ambient temperature, which occurs in 10–100 ps, the system can be considered an equilibrium thermodynamic system, albeit with a high local defect concentration. In most materials, at least one of the single point defects (vacancies or interstitials) is mobile already at room temperature. Hence, after the athermal collision cascade, thermal motion of defects will occur (**box D4**). This occurs until either a next cascade hits the same region or all defects have recombined.

In many materials, and especially in metals, the damage typically takes the form of dislocations after a few damage overlaps (illustrated in lower part of Fig. 1), **box D5**. In fact, in some metals where the heat spikes are dense, such as Cu, Ni, and W, already the primary damage state may include dislocations (Kojima et al. 1989; Nordlund and Gao 1999; Yi et al. 2015). Since the mechanical properties of metals under deformation are completely dominated by dislocations, understanding the high-dose damage in metals requires good understanding of how dislocations form and develop under irradiation conditions and how they behave after the irradiation is finished.

Macroscopic materials modification after high-dose irradiation may take many forms (**box D6**). The buildup of high dislocation concentrations by irradiation generally speaking tends to harden metals, but also makes them more brittle. In the case that vacancy mobility leads to their coalescence into voids, the material will swell – and this effect can be very strong, tens of percentages of volume change. Moreover, especially at elevated temperatures, both point defect and dislocation mobility can lead to permanent shape changes of the material, an effect known as creep.

3   Modeling Methods for Studying the Multiple Scales

In this chapter, we very briefly introduce the basic ideas of the key multiscale simulation tools used in radiation effects modeling and point to which physics levels and separate chapters in this section they relate to. The reader is also referred to the part entitled "Atomistic Simulations and Extensions Towards Multiscale" in Vol. 1 of this handbook for a more general overview of modern methods for atomistic-based multiscale modeling.

The initial penetration of energetic (energies of keV or higher) particles in matter can be calculated with various codes that treat the particle as moving in straight paths between collisions. For neutrons, the collisions occur only with nuclei, and the collision probability can be obtained from nuclear reaction cross sections. Such **neutronics** calculations were actually the first Monte Carlo computer simulations ever performed, in the Manhattan Project (Fermi and Richtmyer 1948). For ions, the collision cross section is much larger, but can be well modeled with the **binary collision approximation (BCA)** method (Robinson and Torrens 1974). The movement of electrons and gamma particles is interlinked at MeV energies and can be described well with so-called **electron-gamma cascade** codes (Nelson 1978). Since all of these methods are well established in their basic formulations on one hand, and do not directly predict damage production on the other, there are no detailed articles related to them. The most popular BCA method by far to simulate radiation damage is SRIM, or the Stopping and Range of Ions in Matter , by Ziegler et al. (1985). This approach, while simple and very easy to use, fails to describe the interaction of radiation damage cascades with material structures of a similar length scale. ▶ Chapter 97, "More Efficient and Accurate Simulations of Primary Radiation Damage in Materials with Nanosized Microstructural Features or Ion Beams," by Li, Yang and Short shows a new approach to directly simulating the interaction of swift ions and neutrons with nanostructures, both isolated and within materials, showing that they are more resilient to damage by many orders of magnitude. Because radiation interactions with nanostructures form more of the underpinnings of material behavior, from structural composites to nanotube-based detectors, this ability is projected to be increasingly important in the years to come.

Reality is of course fundamentally quantum mechanical, and thus all dynamic materials properties are in principle determined by the Schrödinger equation (or for heavy elements its relativistic equivalent to the Dirac equation) for the system of all electrons and all nuclei. Unfortunately solving the true Schrödinger equation for a many-atom system is, even with current computer capacity, intractable. However, the approximate quantum mechanical method, **density functional theory (DFT)**, has proven to be quite accurate and efficient enough to handle system sizes of a few hundred atoms (Jones and Gunnarsson 1989). Hence it is very well suited to examine the properties of point defects and small defect clusters and also defect production dynamics in the smallest collision cascades (Holmström et al. 2008; Olsson et al. 2016). The method is very widely used and reviewed elsewhere for defect calculations (Freysoldt et al. 2014) and hence there is not a separate chapter for this topic.

The few hundred atoms, to which DFT is limited, are certainly not enough to describe full collision cascades. For simulating these, **molecular dynamics (MD)**

simulation utilizing classical potentials is a very appropriate method, as it can easily handle millions of atoms to nanosecond time scales, In MD, Newton's equation of motion is solved for forces acting between atoms, obtained from analytically or numerically defined interatomic potentials. If the interatomic potential is accurate and the electronic stopping power known, MD can describe perfectly the atom motion in collision cascades all the way from the initial high-energy collisions down to the final thermalization to the ambient temperature. Use of MD for collision cascade studies is discussed further in the chapter by A. E. Sand et al., ► Chap. 98, "Incorporating Electronic Effects in Molecular Dynamics Simulations of Neutron and Ion-Induced Collision Cascades"

The thermally activated defect motion between collision cascades can occur on time scales from 10 ps to years. While both DFT and MD can be useful methods to obtain the migration activation energy and Arrhenius equation prefactor for individual defects, they are too slow to simulate the evolution of a complex defect ensemble over long time scales. The **kinetic Monte Carlo (KMC)** method, on the other hand, is very useful for simulating the long-term defect dynamics and is first addressed in the part entitled "Long Timescale Atomistic Simulations: Accelerated Molecular Dynamics and Adaptive Kinetic Monte Carlo" of the MTM set of this handbook. This method needs as input the defect migration rates and a full description of defect reactions (e.g., recombination radius between a vacancy and an interstitial, emission of point defects from clusters, even extended defect reconfigurations and rotations). If all of these are known, the KMC algorithm can describe defect evolution perfectly with the correct time scale and correct relative jump rates of all defects. The object simulated can be atoms (atomic KMC, AKMC) or "defect objects" (object KMC, OKMC). ► Chapter 99, "Atomistic Kinetic Monte Carlo and Solute Effects," by Becquart et al., describes the use of AKMC for simulating solute evolution in metals, a notoriously difficult problem. Multibody potentials must be utilized to enumerate solute-defect reactions, and many solute-defect effects take place at fantastically low concentrations of tens of parts per million, such as the slowdown of vacancy migration by tiny amounts of carbon in iron (Takaki et al. 1983).

Solid links between such disparate length/time scales such as DFT and KMC have characterized progress in the past 10 years, as exemplified by Martinez et al., ► Chap. 100, "DFT-Parameterized Object Kinetic Monte Carlo Simulations of Radiation Damage" This chapter details the use of OKMC as informed by lower length scale simulations from DFT describing precise defect reactions and energetics. This direct link between the atomic/femtosecond scales and the micron/millisecond scales is a prime example of what has only been possible in the past 10 years, as computational power increases enough to parallelize both DFT wave function calculations and larger/longer OKMC simulations, allowing the inclusion of more and rarer transitions between defect states.

The dynamic behavior of dislocations, in turn, can be modeled at a higher length scale with a method known as **discrete dislocation dynamics (DDD)**, addressed in "Advances in Discrete Dislocation Dynamics Simulations" by LeSar

and Capolungo. In this approach, dislocations are described as a chain of connected dislocation segments, each of which giving rise to its own strain field. In the simplification of elastically isotropic materials, these strain fields are known exactly from classical dislocation theory (Hirth and Lothe 1992). Each segment interacts with each other in the system, and their motion is described for forces derived from the strain fields, similar to the basic molecular dynamics equation. The big challenge in DDD is being able to describe all possible reactions of dislocations with each other, voids, and precipitates, for all stresses and temperatures. Such information can be derived from MD, and while serious efforts are underway to do this (Queyreau et al. 2010; Terentyev et al. 2009; Byggmästar et al. 2017), much clearly remains to be done.

Both KMC and DDD can handle micrometer length scales. However, for a large, complicated defect population where the balance of defects is important, neither method can handle a sufficiently large amount of defects. The **cluster dynamics (CD)** method as presented in ▶ Chap. 101, "Rate Theory: Cluster Dynamics, Grouping Methods, and Best Practices"by Jourdan et al., is an efficient rate theory approach, which can describe the time evolution of even very complicated defect populations. It does not have a length scale explicitly; however, from the nature and concentration of defects it describes, one can nevertheless give a rough description of its range of validity in length scales, as shown in Fig. 3. The inputs to CD can be obtained from DFT or MD and are often the same as the inputs needed for KMC. Jourdan's chapter details critical lessons learned since the inception
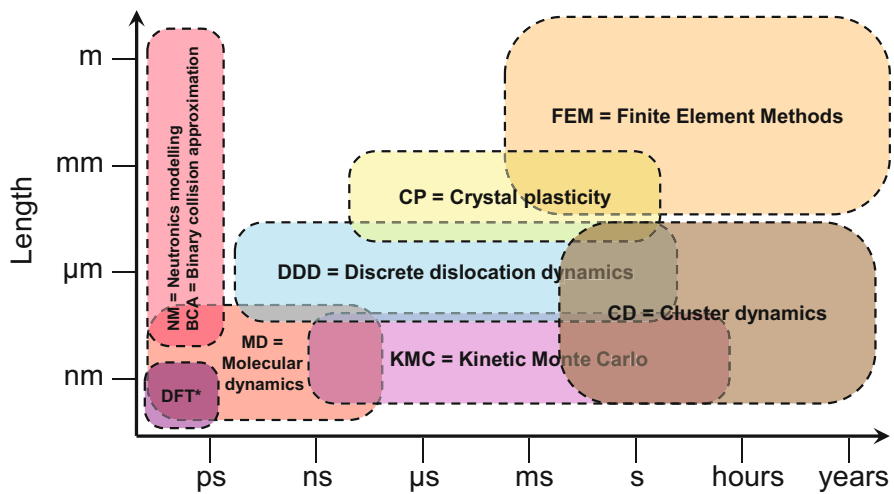
Fig. 3 Multiscale modeling methods for radiation effects in materials. *DFT* density functional theory. The limits in the figure are very rough and indicative only of the typical most relevant time and length scales of the method. In some cases the methods can be applied on a much smaller scale; for instance finite element methods can for linear elastic properties often used down a nm length scale

of cluster dynamics, in terms of how best to run stable and efficient large-scale simulations. This is especially enabled by grouping sets of similarly behaved defects into single ensembles, ensuring that total fluxes between sizes and types of defects are conserved while reducing the number of differential equations required by factors of thousands or more.

Structural engineering materials, especially metals, are almost always polycrystalline, with grain sizes typically in the 1–100 μm. It is well known that grain boundaries can be sinks for point defects; dislocation movement follows crystal direction and is hindered by grain boundaries. Hence a complete description of these materials should also include the effect of different crystal orientation in different grains and grain boundary effects. While grain boundary effects can be described by DFT, MD, KMC, and DDD, these methods are too slow or small-scale to describe the deformation of a large number of differently oriented grains adjacent to each other. The **crystal plasticity (CP)** approach deals with this well (Lebensohn and Tomé 1993) and is discussed thoroughly in the part entitled "Crystal Plasticity: From the Atomic Scale to the Macroscale" in Vol. 1 of this handbook.

In the scale of simulation methods mentioned in this overview, we must address **finite element modeling (FEM)**. In this approach, space is divided into a set of elements, typically triangular (2D) or prism-shaped (3D), all connected to each other in the corners, and a given equation system is solved in this grid. The solution can be either static or dynamic. For the purpose of this chapter, the most relevant use of FEM is the solution of elastoplastic equations describing the response of a material to external forces, as it is widely applied for structural design in mechanical engineering. FEM is, because of this wide use, very reliable for describing stresses and deformation in standard engineering materials (steels, aluminum, etc.) under usual processing conditions. However, the radiation condition is not a usual one, and due to all the physics effects described above, the elastoplastic response (constitutive equations) of a material will be significantly different from the initial one after a prolonged exposure to radiation. Thus, from a nuclear (fusion or fission) development point of view, the ultimate aim of the entire multiscale modeling can be considered to be obtaining both understanding and predictive modeling capability in all the lower levels, which enable formulating physics-based input laws including all radiation effects for FEM.

Finally, we mention a very interesting method that can be used on many different levels of modeling (and hence is not explicitly shown in the multiscale modeling schematic): **artificial intelligence (AI)**. This topic is more generally addressed in the part entitled "Materials Informatics" in Vol. 1 of this handbook. Although the exact definition of the term varies, in modern usage it generally means algorithms that have self-reinforcing or iterative features that enable a program to mimic human learning. With this definition, the term is almost identical to **machine learning (ML),** such as is discussed in Chap. 88, "Machine Learning and Big-Data in Computational Chemistry" in Vol. 1 of this handbook. AI or ML approaches can be useful in many ways in materials modeling, starting from optimizing DFT wave functions, to crafting more accurate MD interatomic potentials, even to mapping out potential defect reactions in KMC. In the future, AI approaches may become

crucial for transferring large data sets on one multiscale level to another one effectively. ▶ Chapter 103, "Modeling Radiation-Induced Segregation and Precipitation: Contributions and Future Perspectives from Artificial Neural Networks" by Castin and Malerba, describes one fairly well-established application of AI in materials modeling, predicting diffusion barriers efficiently using trained neural networks. Such examples are rare so far, as most parameters regarding activation energies, defect reactions, and even structures of defects themselves have been painstakingly acquired through a combination of single simulations and well-crafted validation experiments. Thus ML/AI represent a fundamental paradigm shift, where unleashing automated pattern recognition can point the human investigators toward potentially useful parameters and information far quicker than they could achieve by themselves.

Although this book series deals with materials modeling, it is of course very important to always keep **experimental inputs** in mind. The only theory of all those mentioned above that is known to be exactly correct at all relevant levels of accuracy needed for materials modeling are the Schrödinger/Dirac equations (which as noted are not practically usable for materials). All higher levels involve approximations, and sometimes also assumptions. Due to the rapid emergence of complexity in materials modeling when going up from a single atom, the accuracy of the higher-level methods is difficult to assess from theory alone. Hence it is crucial to utilize the available experimental knowledge and data for formulation and testing of the higher-level models. When the models develop, it is also natural to formulate and perform well-defined experiments that can be used to test the models. ▶ Chapter 102, "Experimental Validation of Models: *In Situ* TEM for Radiation Damage," by Arakawa and Short, deals with the experimental validation of models of radiation damage by *in situ* transmission electron microscopy experiments as an exemplar of how to elegantly confirm key simulations of radiation effects in materials.

4   Conclusion

In conclusion, this introductory article has presented an overview of the physics of radiation effects in solids, as well as the perhaps most widely used models for studying them. The aim was solely to put the following articles into a perspective and indicating how the different levels of modeling contribute to the whole.

References

Bohr N (1913) On the theory of the decrease of velocity of moving electrified particles on passing through matter. Philos Mag 25:10

Byggmästar J, Granberg F, Nordlund K (2017) Molecular dynamics simulations of thermally activated edge dislocation unpinning from voids in alpha-Fe. Phys Rev Mater 1:053603

Chason E, Picraux ST, Poate M, Borland JO, Current MI, Diaz de la Rubia T, Eaglesham DJ, Holland OW, Law ME, Magee CW, Mayer JW, Melngailis J, Tasch AF (1997) Ion beams in silicon processing and characterization. J Appl Phys 81:6513

Diaz de la Rubia T, Averback RS, Benedek R, King WE (1987) Role of thermal spikes in energetic collision cascades. Phys Rev Lett 59:1930 See also erratum: Phys Rev Lett 60:76 (1988)

Fermi E, Richtmyer RD (1948) Note on census-taking in Monte Carlo calculations. A declassified report by Enrico Fermi. From the Los Alamos Archive. Technical Report Number LAMS-805, Series A (July 11)

Fermi E, Teller E (1947) The capture of negative mesotrons in matter. Phys Rev 72:399

Freysoldt C, Grabowski B, Tilmann H et al (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86:253

Gades H, Urbassek HM (1995) Simulation of ion-induced mixing of metals. Phys Rev B 51:14559

Grove WR (1852) VII. On the electro-chemical polarity of gases. Philos Trans R Soc 142(I):87

Hirth JP, Lothe J (1992) Theory of dislocations, 2nd edn. Krieger, Malabar

Holmström E, Kuronen A, Nordlund K (2008) Threshold defect production in silicon determined by density functional theory molecular dynamics simulations. Phys Rev B 78:045202

Jones RO, Gunnarsson O (1989) The density functional formalism, its applications and prospects. Rev Mod Phys 61:689

Kojima S, Satoh Y, Taoka H, Ishida I, Yoshie T, Kiritani M (1989) Confirmation of vacancy-type stacking fault tetrahedra in quenched, deformed and irradiated face-centred cubic metals. Philos Mag A 59:519

Krasheninnikov AV, Nordlund K (2010) Ion and electron irradiation-induced effects in nanostructured materials. J Appl Phys 107:071301

Lebensohn RA, Tomé CN (1993) A self-consistent anisotropic approach for the simulation of plastic deformation and texture development of polycrystals: application to zirconium alloys. Acta Metall Mater 41(9):2611–2624

Lindhard J, Scharff M, Schiott HE (1963) Range concepts and heavy ion ranges. Kgl Danske Vid Selsk Mat Fys Medd 33(14):1

Meldrum A, Zinkle SJ, Boatner LA, Ewing RC (1998) A transient liquid-like phase in the displacement cascades in zircon, hafnon and thorite. Nature 395:56

Mueller E, Bahadur K (1956) Field ionization of gases at a metal surface and the resolution of the field ion microscope. Phys Rev 102:624

Nelson WR (1978) Solution of the electromagnetic cascade shower problem by analog Monte Carlo methods – EGS. In: Nelson WR (ed) Computer techniques in radiation transport and dosimetry. New York/London, Plenum Press, pp 173–196

Nordlund K, Gao F (1999) Formation of stacking fault tetrahedra in collision cascades. Appl Phys Lett 74:2720

Nordlund K, Ghaly M, Averback RS, Caturla M, Diaz de la Rubia T, Tarus J (1998) Defect production in collision cascades in elemental semiconductors and fcc metals. Phys Rev B 57:7556

Nordlund K, Wallenius J, Malerba L (2005) Molecular dynamics simulations of threshold energies in Fe. Nucl Instr Meth Phys Res B 246:322

Olsson P, Becquart CS, Domain C (2016) Ab initio threshold displacement energies in iron. Mater Res Lett 4:216

Polvi J, Luukkonen P, Nordlund K, Järvi TT, Kemper TW, Sinnott SB (2012) Primary radiation defect production in polyethylene and cellulose. J Phys Chem B 116:13932

Queyreau S, Monnet G, Devincre B (2010) Orowan strengthening and forest hardening superposition examined by dislocation dynamics simulations. Acta Mater 58:5586

Robinson MT, Torrens IM (1974) Computer simulation of atomic-displacement cascades in solids in the binary-collision approximation. Phys Rev B 9:5008

Ruault MO, Chaumont J, Penisson JM, Bourret A (1984) High resolution and *in situ* investigation of defects in Bi-irradiated Si. Philos Mag A 50:667

Rutherford E (1911) The scattering of α and β rays by matter and the structure of the atom. Philos Mag 6:31

Seitz F, Koehler JS (1956) Displacement of atoms during irradiation. In: Seitz F, Turnbull D (eds) Solid state physics, vol 2. Academic Press, New York, p 307

Takaki S et al (1983) The resistivity recovery of high purity and carbon doped iron following low temperature electron irradiation. Rad Eff 79(1–4):87–122

Terentyev D, Klimenkov M, Malerba L (2009) Confinement of motion of interstitial clusters and dislocation loops in bcc Fe–Cr alloys. J Nucl Mater 393:30

Yi X, Sand AE, Mason DR, Kirk MA, Roberts SG, Nordlund K, Dudarev SL (2015) Direct observation of size scaling and elastic interaction between nano-scale defects in collision cascades. Europhys Lett 110:36001

Ziegler JF, Biersack JP, Littmark U (1985) The stopping and range of ions in matter. Pergamon, New York

More Efficient and Accurate Simulations of Primary Radiation Damage in Materials with Nanosized Microstructural Features or Ion Beams

97

Ju Li, Yang Yang, and Michael P. Short

Contents

Abstract

Primary radiation damage, resulting from the direct ballistic collisions of energetic particles with matter, is the starting point for all of radiation damage. It has traditionally been simulated using binary collision approximations (BCA) Monte Carlo (MC) method in bulk or layered materials, which has served the community well until now. The introduction of nanosized features into materials, whether as 0D dispersoids to pin grain boundaries, 1D dispersoids for removal of helium and sinking defects, or other nanosized features, introduces major errors into conventional bulk/multilayer BCA-MC simulations. This is due to an inability to simulate the exchange of ions at internal/external phase boundaries and the creation of struck atoms with ranges larger than or comparable to the

J. Li (✉) · Y. Yang · M. P. Short
Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: liju@mit.edu; yyn@mit.edu; hereiam@mit.edu

2381

smallest microstructural feature. We first review the fundamentals of BCA-MC simulations with continuous electronic slowing down for energetic ions, followed by identifying where the traditional approaches fail, and ending with a new full-3D simulation capability to correctly model such features. Such simulations drive the planning and interpretation of radiation exposure campaigns, ion implantation, ion modifications, and even the basic definition of radiation damage. A comparison to more accurate but computationally more expensive, molecular dynamics (MD) simulations of radiation damage will be discussed.

1   Introduction to Primary Radiation Damage

Radiation damage, by nature, refers to permanent atomic displacements in solid matter caused by energetic particles, including neutrons, charged particles (e.g., electrons, protons, and α particles), and photons. During the interaction between radiation and solids, a series of atomic displacements may arise, creating damage in materials that impacts their material properties. For example, energetic particles create abundant vacancies and interstitials in materials that may lead to swelling (Short et al. 2015) and embrittlement. Understanding the interplay between radiation and matter is vital not only for designing materials and systems with improved radiation damage tolerance, but also for creating novel devices by defect engineering (Yang et al. 2018a). Atomic displacements start as primary damage, on a timescale of femtoseconds, with the phase known as "primary radiation damage" lasting roughly ∼10 ps until all atoms ballistically come to rest. Figure 1 illustrates the primary damage process where a series of knock-on atoms are generated: primary knock-on atoms (PKAs), secondary knock-on atoms (SKAs), tertiary knock-on atom (TKAs), and so on. Once the kinetic energies of the incident ions and all the knock-on atoms are low enough (usually much lower than the threshold displacement energy (Crocombette 2005)), then primary radiation damage is said to be complete.
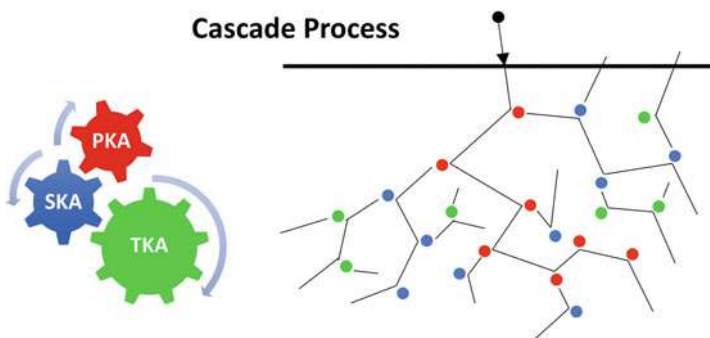
Fig. 1 Illustration of the PKA/SKA/TKA during the radiation damage cascade process. Once all knock-on atoms come to rest, then the radiation damage cascade, known as "primary radiation damage," is complete

A PKA is born when an incident energetic particle bombards a native atom of the material, transferring kinetic energy T to the native atom above its threshold displacement energy E_d. Multiple PKAs ("generation 1") can be created by the same energetic particle ("generation 0"), and these can be considered the generation-1 leaves of the generation-0 root node of a collision cascade tree. A SKA("generation 2")/TKA ("generation 3")/ . . . is generated by collisions between a PKA/SKA/ . . . and a native atom with the same energy threshold criterion. The whole process is called a "cascade" and terminates when the kinetic energy of every struck atom is below its displacement threshold energy. Rigorously speaking, PKAs, SKAs, TKAs, etc., are not mutually exclusive, since a native atom can in principle be first displaced by an external particle, settle down, and later become displaced by another SKA, although such probabilities are exceedingly small. Along with the cascade process, there is spontaneous recombination (athermal process) which happens below the timescale of picoseconds (Li et al. 2015). Above that, the defects generated will migrate and interact with each other as well as dislocations and different microstructural damage sinks, leading to microstructural evolution and modification of bulk-scale material properties. The time and length scales of these interactions are outlined in Fig. 2. Radiation damage in materials is an inherently multiscale, multiphysics problem (▶ Chap. 96, "Modeling of Radiation Damage in Materials: Best Practices and Future Directions").
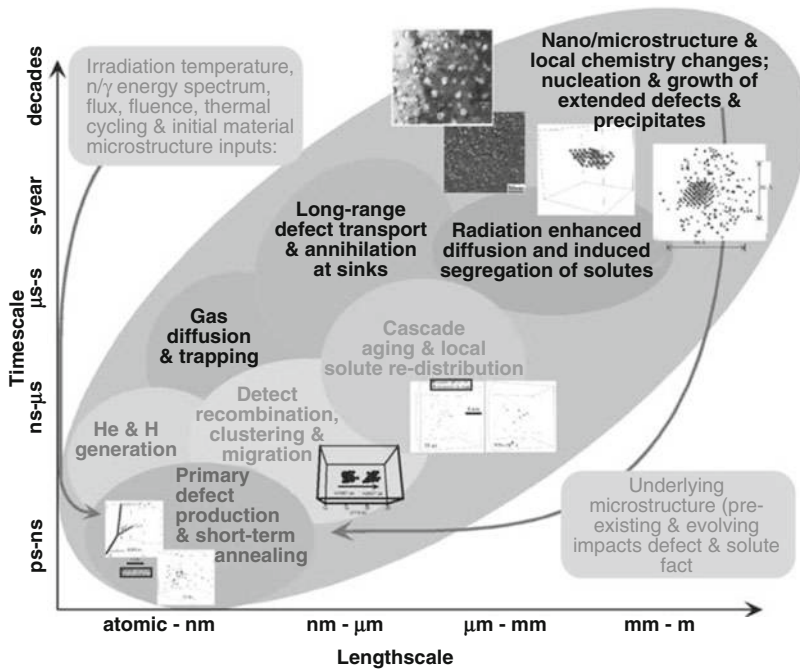
Fig. 2 Illustration of the length and timescales (and inherent feedback) involved in the multiscale processes responsible for microstructural changes in irradiated materials (Odette and Wirth 2005)

Neutrons are the main source of atomic displacement damage in nuclear reactors. Although transmutation can create charged particles such as alpha particles in reactors, this is considered as a second order effect (Woo 2005). A neutron may interact with atoms in multiple ways: elastic scattering (n, n), radiative capture (n, γ), inelastic scattering (n, n'), fission (n, f), and charged particle emission (Yip 2014). Among these mechanisms, elastic and inelastic scattering are the main ways to transfer kinetic energy to target atoms. By contrast, a charged particle may interact with materials in four different ways: (a) elastic collision with electrons, (b) inelastic collision with electrons, (c) elastic collisions with atomic nuclei, and (d) inelastic collisions with atomic nuclei (Yip 2014). (b) and (c) are the main driving forces to stop the transport of charged particles, while (d) is often called "bremsstrahlung" and only becomes prominent for charged particles at very high energy. In BCA-MC simulations, it is often assumed that (a) will only change the kinetic energies of charged particles, while it has no effect on the speed direction. In this context, (c) is the leading term for changing the direction of charged particles and creation of atomic displacements in materials.

In this chapter, we will only focus on the simulations of primary radiation damage. These atomic-scale simulations provide details of the displacement process as well as input parameters for larger scale simulations. These are used as direct inputs for longer timescale models of defect migration, clustering, elemental segregation or dealloying, and evolution of microstructure, mechanical properties, and thermal, electronic, and magnetic properties of irradiated materials. For example, primary radiation damage simulated by BCA-MC codes is used as input for point kinetics simulations of radiation-induced void swelling (Short et al. 2015), MD simulation of the re-dissolution of Xe fission gas bubbles in uranium dioxide (Schwen et al. 2009), cluster dynamics simulations of nanoscale defect agglomeration (Xu et al. 2012), and so on. Also, it is often required in the planning process of irradiating materials for performance analysis, and its high importance is underscored by being the root of all follow-on simulations. Thus, it is critical that these simulations are as accurate as possible, for as many types and forms of materials as possible. The reader is referred to all other chapters in this part of the Handbook for more information about the subsequent stages of radiation damage and the best practices with which to model them.

2 BCA-MC Simulations of Primary Radiation Damage under Ion Irradiation

In this part of the chapter, we will briefly review what BCA-Monte Carlo (BCA-MC) simulations are, followed by examples of current methods and codes to carry them out, and concluding with the common shortcomings of non-full-3D simulations. This will set the stage for the required modifications to BCA-MC models to accurately model radiation damage in certain types of irradiation scenarios – those with nanosized features (in target materials or/and incident radiation beam) smaller than or comparable to the mean free path of the ionizing radiation.

Generally, there are two types of BCA simulations for ion irradiation, classified by the assumption of target structure: (1) crystalline structure, simulated by codes such as MARLOWE (Robinson and Torrens 1974); (2) amorphous structure, simulated by codes such as SRIM/TRIM. The latter is simpler and more computationally efficient, and has been more widely used in the past decade. In the following text we will focus on BCA-MC codes based on the random phase approximation (RPA), assuming the target is in amorphous structure. MC simulations based on BCA models are robust methods to understand the detailed cascade process of primary radiation damage (Gilmer and Yip 2005). A BCA-MC simulation tracks all collision events along the trajectories of incident ions. The trajectory of each ion is called a "history." By computing random trajectories of incident ions (usually about 10^4–10^8 atoms), a MC simulation presents statistically meaningful data about the spatial distribution of the injected ions, primary radiation damage in the form of Frenkel pair defects, and energy deposition. The histories of the incoming neutrons and photons ("generation 0") are often modeled using MCNP (Briesmeister 1986), Geant4 (Agostinelli et al. 2003), and so on, while for incoming ions (generations-1 and above, considered native) the most famous and widely used tool is TRIM/SRIM (Ziegler et al. 2008). This chapter will focus solely on modeling ion-matter interactions, as the damage and defects produced by neutrons and photons are so sparse compared to ions (typically millimeter to centimeter between events, owing to far lower cross sections) that very different methods must be used. Note that in accelerator tests, the external ion accelerated by the accelerator is also considered "generation 0," whereas the displaced native ions are considered generation 1 and above.

Assumptions are vital for the simplification of the full breadth of interactions to make the physical process more easily computable. Most simulation frameworks to compute primary radiation damage make these same assumptions, so they will be introduced as a common starting point for understanding the simulation process. The cascade processes induced by different ions (generations 0, 1, 2, …) are considered to be independent, and only two-body interactions are considered for elastic scattering with target nuclei. The target atom is usually assumed to be stationary before the collision. It is further assumed that ions always travel in straight paths between two binary collisions. The line segment between two successive collisions is called the "free-flight-path," within which only inelastic scattering with electrons contributes to the change in ion energy, known as electronic energy loss. Therefore, the trajectory of each ion is composed of a series of successive free-flight-paths with elastic nuclear scattering, as depicted in Fig. 3, changing the particle's energy and direction. The moving direction of ion can only be changed during an elastic scattering event with a nucleus; thus, it is assumed that electronic interactions change the direction of the particle so little that its direction of travel is essentially unchanged. A particle's history is terminated when it exits the target with zero possibility of reentry, or its energy falls below a pre-defined cutoff energy, which is set depending on the situation being simulated. A proper cutoff energy is important for enhancing the simulation efficiency. The reader is referred to Sect. 6.2.1 in Eckstein (1991) for more details about how to set an appropriate cutoff energy. Bremsstrahlung is usually ignored.
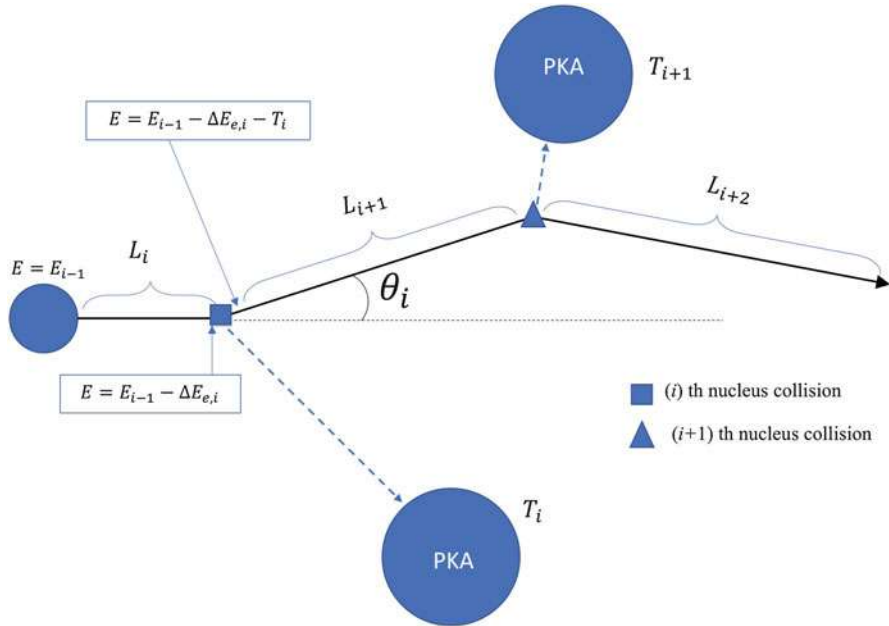
Fig. 3 Schematic of the modeling process. L is the free-flight-path length, $\Delta E_{e,i}$ is the electronic energy loss in L_i, and T_i is the kinetic energy loss during the i-th nuclear collision

A basic algorithm of the BCA-MC simulation modified from (Yang et al. 2018b) is shown in Fig. 4 in the form of pseudocode, to elucidate the general algorithm of BCA simulations while avoiding confusion and unnecessary details. A more detailed program flowchart showing the different treatments of the quick Kinchin-Pease (KP) mode and the full cascade (FC) mode can be found in the documentation of IM3D (Li et al. 2015).

There are four key ingredients which one needs in order to compute in a BCA-MC simulation, which are the free-flight-path length (L), the stopping power (determines $\Delta E_{e,i}/L_i$), the scattering angle θ_c (determines the direction of L), and the nuclear scattering energy loss (T_i). The reader is referred to existing references for the calculation of many of these parameters, such as the introduction of the algorithms in TRIM (Biersack and Haggmark 1980), a review of the computer simulation for ion-solid interactions (Eckstein 1991), textbook of SRIM (Ziegler et al. 2008), and recent corrections of stopping powers in the case of high-Z ions damaging low-Z materials (Jin et al. 2014). It is the chosen of the ion-matter interaction models and the numerical methods that differ between legacy and newer BCA-MC simulations. These differences manifest themselves not only in the sampling methods for the free-flight-path length L and the impact parameter P (Biersack and Haggmark 1980) according to the ion and target properties but also

```
for all particles do
    Initialization of position, direction and energy;
    /* Take one ion from the 'bank'                        */
    life←1;
    while life=1 do
        Calculate reduced energy (ε);
        Calculate the stopping power;
        Sample the free-flight-path length (L) and the impact parameter (P);
        Move to new location and save trajectories;
        Calculate electronic energy loss;
        Reduce energy accordingly ;

        if cross boundary then
            Correct electronic energy loss; Move to boundary; Change material;
            /* Note:  The treatment of boundary crossing
               varies in different codes.                  */
        else if exit the target then
            life←0;
        else
            /* Collision                                    */
            Sample direction;
            Calculate nuclear energy loss;
            Reduce energy accordingly ;
            If a recoil is generated, use the KP model to calculate the primary
            damage, or store the recoil in the 'bank' for full-cascade (FC);
            Add tally;
            life←0 if E is so small;
        end
    end
end
```

Fig. 4 Basic BCA/MC simulation algorithm in pseudocode, showing general code structure, modified from Ref. (Yang et al. 2018b)

in computational approaches to efficiently evaluate scattering angles and stopping powers. Below we briefly review different schemes based on these two aspects, so the reader can appreciate what has evolved in the past decade.

TRIM/SRIM (Biersack and Haggmark 1980; Ziegler et al. 2008), the most widely used tool for BCA simulations, uses semi-empirical relations to compute the electronic and nuclear stopping of ions and struck atoms. These take root in "reduced energies" (ε), by which the physical model of the ion-matter interaction is chosen. The ion energy space is further divided into three regions (low, intermediate,

and high) or four regions (very low, low, intermediate, and high), with each region assigned a unique set of semi-empirical equations for the evaluation of the impact parameter P, the free-flight-path length L, and the electronic stopping power S_e. The impact parameter P is sampled using this reduced energy, which changes its form depending on whether the ion's energy is low, intermediate, or high by assuming hard sphere collisions, a random nonlogarithmic impact parameter, or a random natural log-based impact parameter, respectively. The electronic stopping power in the early version of TRIM (Biersack and Haggmark 1980) is described by two different sets of semi-empirical relations based on different physical models in the low and the high energy regions, while in the intermediate energy region an interpolation method is used. For the ease of computation, fitting formulas for the electronic stopping power are used. A later version of SRIM provides a stopping power database using experimental results, containing over 28,000 stopping values (Ziegler et al. 2010). The scattering angle θ_c relies on the chosen interatomic potential and the impact parameter P as sampled. A key challenge here is the computationally heavy evaluation of the scattering integral (Eckstein 1991), for which various numerical recipes are applied in different codes, such as the so-called "MAGIC" fitting formula in SRIM/TRIM (Biersack and Haggmark 1980), Gauss–Legendre and Gauss-Mehler quadratures reviewed in Ref. (Eckstein 1991), the latest "fast indexing" approach in Corteo (Schiettekatte 2008), and IM3D (Li et al. 2015). In terms of the free-flight-path length L, a random impact-parameter-dependent $L(P)$, a constant L equal to the mean atomic distance in the target, or a randomly energy-dependent $L(\varepsilon)$ is applied at low, intermediate, and high energies, respectively. Two newer frameworks, Corteo (Schiettekatte 2008) and IM3D (Li et al. 2015), sample L over all energies according to a Poisson distribution, with a random inverse-logarithmic sampling of the impact parameter different from that found in SRIM. Otherwise, the basic physics of the older and newer frameworks are quite similar.

One of the major variations among the legacy and newer BCA-MC simulations is the assumption of beam shape and target geometry. SRIM/TRIM, for example, assumes a point incident beam (or "pencil beam") with bulk/multilayer target. The newer codes, however, enable the consideration of arbitrary beam profile (shape and size) as well as 3D target morphologies. The frameworks that is restricted out of the full-3D scope break down in the case of nanoscale microstructural features and radiation beams, or more generally speaking whenever ions encounter features in either the beam or the target that are close to or smaller than the mean free path of the ions. In these cases, codes like SRIM, which can only model bulk or multi-layered materials, fail to account for ions which laterally exit/enter any phase boundary and for a nanoscopically focused beam. This can result in a gross overestimation of primary radiation damage in nanosized features, whether in a radiation mask with pinholes (Bayn et al. 2015), lone carbon nanotubes, microsized mechanical test specimens, oxide dispersion strengthened (ODS) particles, or any other submicron features in materials (Yang et al. 2018a). Thus, more advanced features must be added to BCA codes to accurately simulate primary radiation damage. It should also be noted that complexity often hinders performance, so such features must be added while keeping computational efficiency in mind.

3   The Necessity of Full-3D BCA-MC Simulations

While SRIM/TRIM can perform 3D simulations, it is based on several assumptions on the beam profile and target structure. First, the beam source is assumed to be a pencil beam, meaning that ions enter the target via a single point on the surface of target all with identical velocities. Second, the target is at most a multilayered structure with interfaces perpendicular to the longitudinal (i.e., depth) direction, indicating no physical boundaries in the lateral directions. Its 1D output is the most widely used for engineering purposes. If the defect distribution in 3D has a functional form of $F(x, y, z)$, then the 1D output along depth direction z is computed by:

$$f(z) \equiv \iint_{-\infty}^{\infty} F(x, y, z)\, dx dy \tag{1}$$

$f(z)$ ignores the lateral straggling of ions, which usually ranges from a few nanometers to a few microns. However, it works as a good approximation for the traditional bulk scale experiments, where both the ion beam and the target size are larger than tens of microns (Yang et al. 2018a). Recently, nanoscale ion implantation (nanosized ion beam or target, or both) has drawn expanding interest, facilitating the fabrication of novel devices, such as quantum computers (Shinada et al. 2005; Bayn et al. 2015), magnetometers (Maletinsky et al. 2012), nanowire p-n junctions (Hoffmann et al. 2009). An excellent review about the progress (before 2011) of engineering 0D or 1D nanomaterials using ion radiation has been provided by Krasheninnikov and Nordlund (2010).

3.1   Examples Where Full 3D BCA Simulations Are Required

Because of the breakdown of several key assumptions, these traditional restricted 3D simulations as well as their 1D outputs may be misused, leading to significant errors when predicting defect distributions. Three nanosized effects of ion radiation for MC simulations have been reported, namely, the nano-beam, nano-target (Yang et al. 2018a), and nano-energetic effects (Li et al. 2015), calling for full-3D simulations. The difference between nano-beam and nano-target experiments is illustrated in Fig. 5a.

   The nano-beam effect arises where the beam size is small and thus the assumptions of 1D output, especially the pencil beam assumption, lose validity. This effect has important implications in nanofabrication, where precision and accuracy relies heavily on beam size. Advanced implantation systems (Watt et al. 2007; Tan et al. 2011; Economou et al. 2012; Yao et al. 2013; PEKIN et al. 2016) and implantation masks with nanoapertures (Pezzagna et al. 2010, 2011; Toyli et al. 2010; Bayn et al. 2015; Scarabelli et al. 2016) have been extensively developed to precisely control the width and position of ion beams. Examples include MeV proton beams with a beam-width below 14 nm (Yao et al. 2013) and silicon masks with pinholes for the creation of nitrogen-vacancy color centers (Bayn et al. 2015). The relative error of the 1D output of TRIM/SRIM in the application of nanosized ion-beam has been
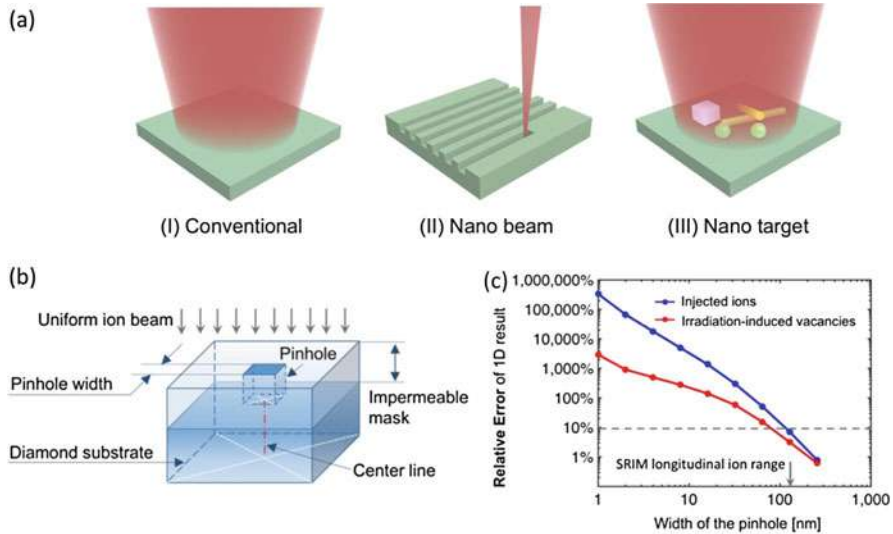
Fig. 5 (**a**) Schematic illustration of the difference between a conventional bulk-scale experiment, a nano-beam experiment, and a nano-target experiment. (**b**) Schematic diagram of the NV center creation experiment. (**c**) Relative error of the 1D output by SRIM compared to the full-3D simulation. Reproduced from Ref. (Yang et al. 2018a) with permission from The Royal Society of Chemistry

quantified using an example of implanting $^{15}N^+$ ions into diamond for the creation of NV centers (Yang et al. 2018a), as shown in Fig. 5b, c. In this case, the effective beam width is equivalent to the width of the square pinhole in the silicon mask. It is shown that the relative error of the traditional 1D output increases dramatically as the width of the pinhole decreases. A full-3D simulation is required to take into account the beam shape information (i.e., the 2D ion density distribution at entrance).

The nano-target effect, on the other hand, becomes prominent when the target size is small, leading to the breakdown of both the pencil beam and multilayer target assumptions. SRIM/TRIM ignores heterogeneity in the lateral directions, failing to consider ion exchange at physical boundaries of the target. Here we take implantation in nanowires as an example (see Fig. 10), which can be used to make nanowire-based devices such as p-n junctions (Hoffmann et al. 2009) and mechanically test ion-irradiated samples (Liontas et al. 2014). The traditional restricted 3D simulations and 1D output are incapable of considering ion leakage from surfaces, overestimating the defects accumulated in the nanowire. Meanwhile, as the target size shrinks, the shape of the defect distribution evolves accordingly, with its peak shifting to a shallower place (Yang et al. 2018a), as shown in Fig. 6b, c.

Another example of the nano-target effect is shown in Fig. 7, where the sputtering yield of spherical Au nanoparticles under the irradiation of Ar^+ and Ga^+ is shown to be enhanced when the nanoparticle radius is small (Holland-Moritz et al. 2015) by *iradina* (Borschel and Ronning 2011) simulations, in agreement with experimental
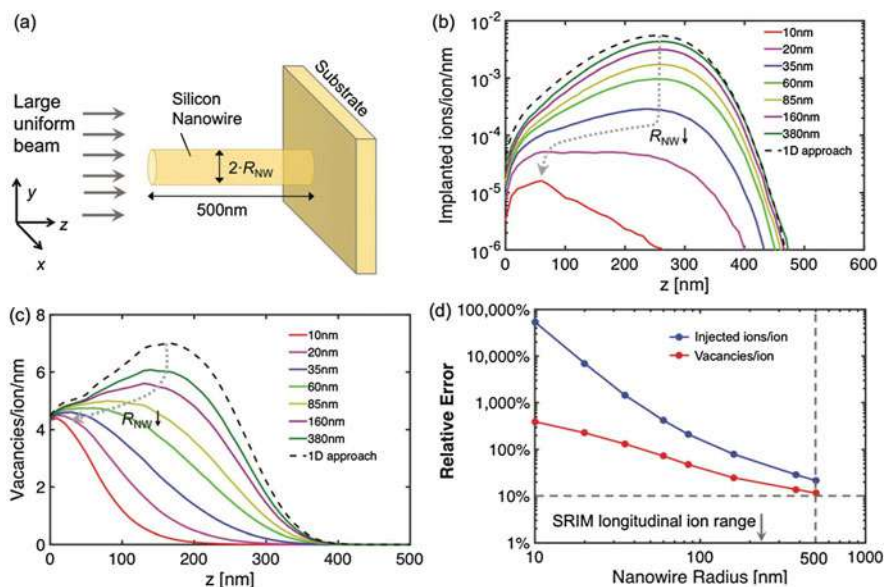
Fig. 6 (**a**) Schematic diagram of the head-on ion implantation experiment. (**b**, **c**) Distribution of point defects along *z* axis as the nanowire radius decreases. (**d**) Relative error of the 1D output by SRIM/TRIM. Reproduced from Ref. (Yang et al. 2018a) with permission from The Royal Society of Chemistry

results. The enhanced sputtering in nanostructures can be attributed to a large surface-to-volume ratio which allows more ions to exit the target, as illustrated in Fig. 7c, d. Figure 7a, b shows the energy dependence of sputtering yield in bulk and Au nanoparticles, whose diameter follows a Gaussian-shaped distribution with a mean size about 56 nm. Figure 7e, f shows the effect of nanoparticle size on the sputtering yield. The bulk-scale simulation is found to significantly underestimate the sputtering yield, while failing to predict the diameter of the nanoparticle that yields the most sputtering.

The nano-energetic effect is related to the loss of validity of bulk physical parameters at small scales due to quantum confinement, surface stress, and elastic image interactions. These fundamental parameters include electronic stopping power and the threshold energy, which could depend on the sample size. The number of athermal recombinations (Nordlund et al. 2018) could have an even stronger dependency on the size and geometry of the sample, since the sample surface and interfaces provide different venues for knock-out and recombination compared to the bulk. Figure 8 shows an example of the nano-energetic effect for the ion implantation of Ar into 3–4 nm diameter GaN nanowires. It is known that both the Kinchin–Pease (KP) and full-cascade (FC) modes in SRIM-like MC codes will over-estimate vacancy creation in solids (Stoller et al. 2013); therefore, it is particularly surprising that MC methods underestimate vacancy creation for the nanowire implantation cases, if bulk physical parameters are used (see the difference
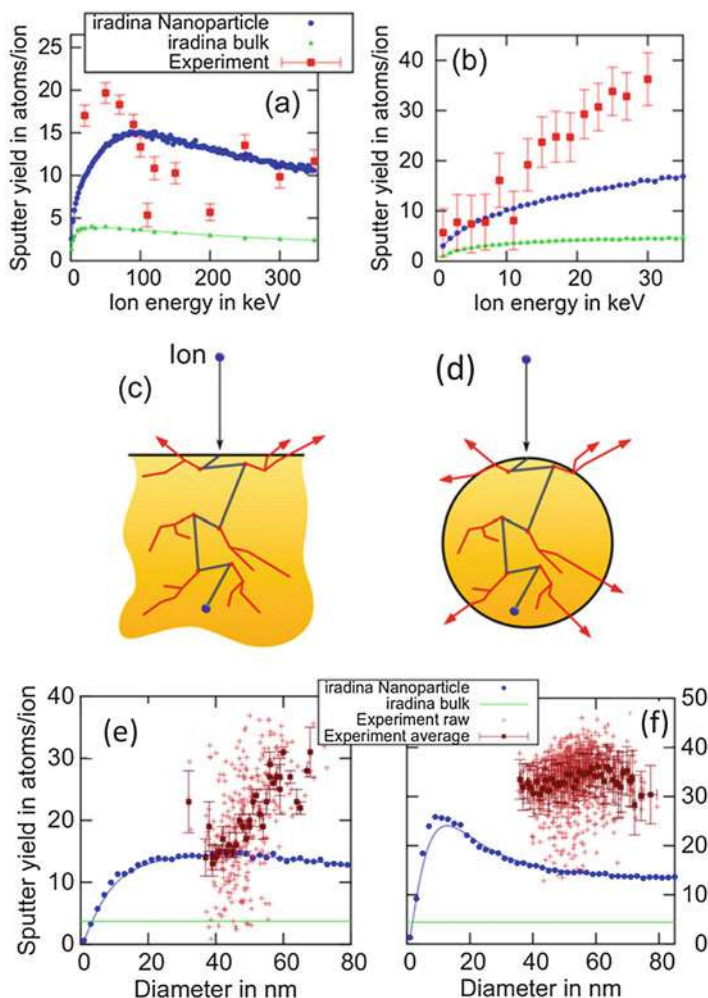
Fig. 7 The sputtering yield on bulk and Au nanoparticles as a function of ion energy: (**a**) Ar$^+$; (**b**) Ga$^+$. (**c**, **d**) Schematic illustration of the difference between sputtering on bulk sample and nanoparticles. (**e-f**) The sputtering yield as a function of the diameter of Au nanoparticles: (**e**) 95 keV Ar$^+$; (**f**) 25 keV Ga$^+$ (Holland-Moritz et al. 2015)

between MD and IM3D-1 in Fig. 8). By reducing the displacement threshold energy by half, the result of MC simulation (IM3D-2 in Fig. 8) agrees better with that of MD. Comparing Fig. 8a, b, it is also found that the nano-energetic effect is less prominent when the size of nanowire increases. The reader is referred to the supplementary document in Ref. (Li et al. 2017) for more details about this nano-energetic effect. Another example is shown in Fig. 8c, where researchers found the displacement energy of carbon nanotubes has a dependence on the diameter (Krasheninnikov and Nordlund 2010).
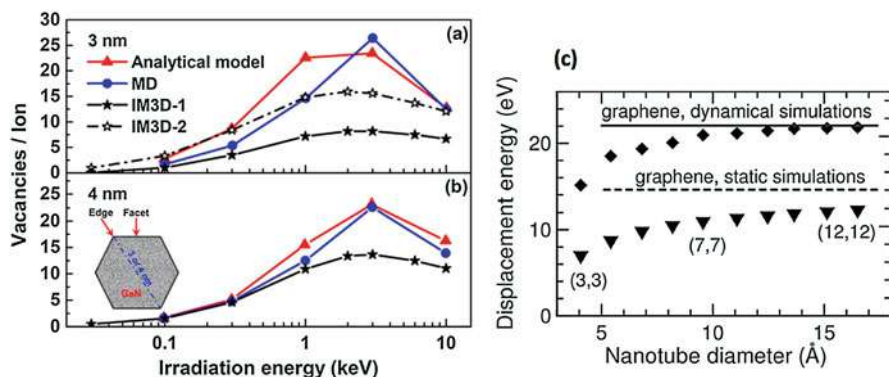
Fig. 8 Illustration of the nano-energetic effect. (**a**, **b**) The difference between IM3D-1 and IM3D-2 is that the former adopts the bulk energy thresholds, while the latter adopts half values of the bulk thresholds. Figure reprinted from (Li et al. 2015) with permission. (**c**) Dependence of the displacement energy on the carbon nanotube diameter (Krasheninnikov and Nordlund 2010)
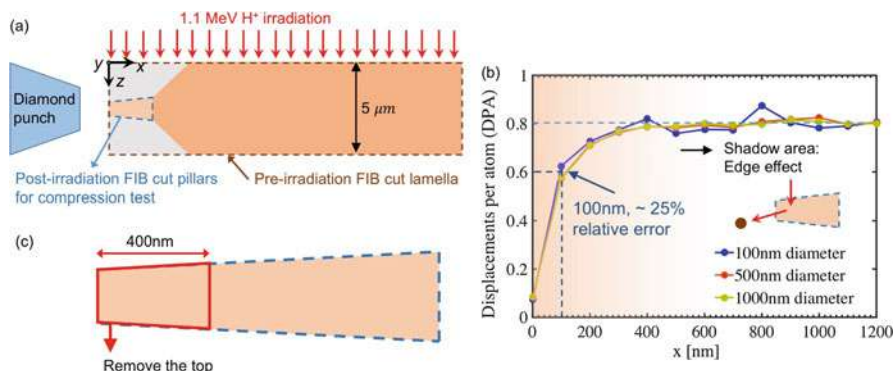
Fig. 9 (**a**) Schematic drawing of the experimental setup. (**b**) Radiation damage as a function of the distance to the edge along *x* axis. (**c**) Schematic drawing about the removal of the top of the pillar (Yang et al. 2018a)

Although the nano-beam and nano-target effects can be simulated conveniently in full-3D simulation codes such as IM3D, the nano-energetic effect requires corrected physical parameters usually obtained by MD/DFT simulations. It is recommended to consider the nano-energetic effect when the characteristic length of the target is below 20 nm, as the thermodynamic properties vary less sensitively with object sizes above 20 nm (Li et al. 2015 and 2017).

In addition to the three nanosized effects above, researchers also found an "edge-attenuation" effect and a "surface roughness" effect that require full-3D simulations, although their beam/target sizes or surface features are comparatively large. The "edge-attenuation" effect should be considered when the edge of the target is important during a traditional bulk-scale experiment. Figure 9 shows an example

(Yang et al. 2018a). Because ion radiation has very limited penetration capability, one has to perform small-scale mechanical experiments to understand the radiation hardening effect of ion irradiated samples. Researchers first irradiated a 5 µm thick Cu thin film by 1.1 MeV protons and then used a focused ion beam (FIB) to cut a small pillar at the edge for small-scale mechanical experiments, as illustrated in Fig. 9a. It is vital for the pillar to have a uniform distribution of radiation damage. However, a full-3D simulation performed in IM3D revealed that there is an "edge attenuation" of radiation damage along the x axis from the inner to the edge. Therefore, one should remove the top (i.e., from the $x = 0$ to 400 nm) to avoid uneven radiation dose distributions in the pillar. This is an absolutely critical finding for those undertaking microscale compression pillar irradiation experiments, as without these 3D BCA-MC simulations this effect would not have been discovered and would have continued to alter the consistency of microcompression experiments under irradiation compared to bulk experiments.

The "surface roughness" effect has important implications in the study of ion-surface interactions. For example, in fusion reactors, the plasma facing materials (PFMs) suffer from the impact of low energy (10–1000 eV) and high fluence (up to 10^{24} particles/m^2-s) D/T/He ions, which lead to sputtering and ion retention (Nordlund et al. 2014). The surface morphology of PFMs can be modified by the plasma. As a result, a series of surface features are formed, including fuzz (Baldwin and Doerner 2010) and surface blisters (Wang et al. 2001). Surface roughness evolution will in turn have a significant effect on sputtering and ion retention. Traditional SRIM/TRIM is only designed for bulk targets with smooth, nonevolving surfaces; thus, it will fail to predict this surface roughness effect. Previously, simulation codes based on fractal surface geometry models, such as VF-TRIM (Ruzic 1990), ACAT (Kenmotsu et al. 2005), and ITMC-F (Hu and Hassanein 2012), have been developed to understand this effect. Recently, a rough-surface geometry model based on finite element triangular mesh (FETM) method has been developed using the IM3D framework (Li et al. 2017). Since the FETM method is simpler and more realistic than the fractal model in terms of validation with experimental atomic force microscopy (AFM) images, here we will only introduce the former. In the FETM model, the surface is first divided into square pixels with a side-length of a and then subdivided into triangles by randomly selecting one diagonal of each pixel, as shown in Fig. 10a. The height Z of each point in the mesh is sampled using a truncated Gaussian distribution $f(Z) \propto \exp(-Z^2/(2\sigma^2))$, where $Z \in [-3\sigma, 3\sigma]$. Therefore, the surface roughness can be adjusted conveniently and intuitively by controlling the surface roughness amplitude 3σ and square mesh size a. The primary ion retention and sputtering yield as a function of surface 3σ and ion incident angle are shown in Fig. 10d, e. It is found that as the surface roughness increases, the primary ion retention will increase while the sputtering yield will decrease. A comparison of the experiment and IM3D simulation is provided in Fig. 10f, showing significant improvement of sputtering yield by making the surface smoother. However, it is surprising that the highest sputtering yield occurs before the surface roughness reaches zero, indicating an optimum surface roughness leading to the highest sputtering yield.
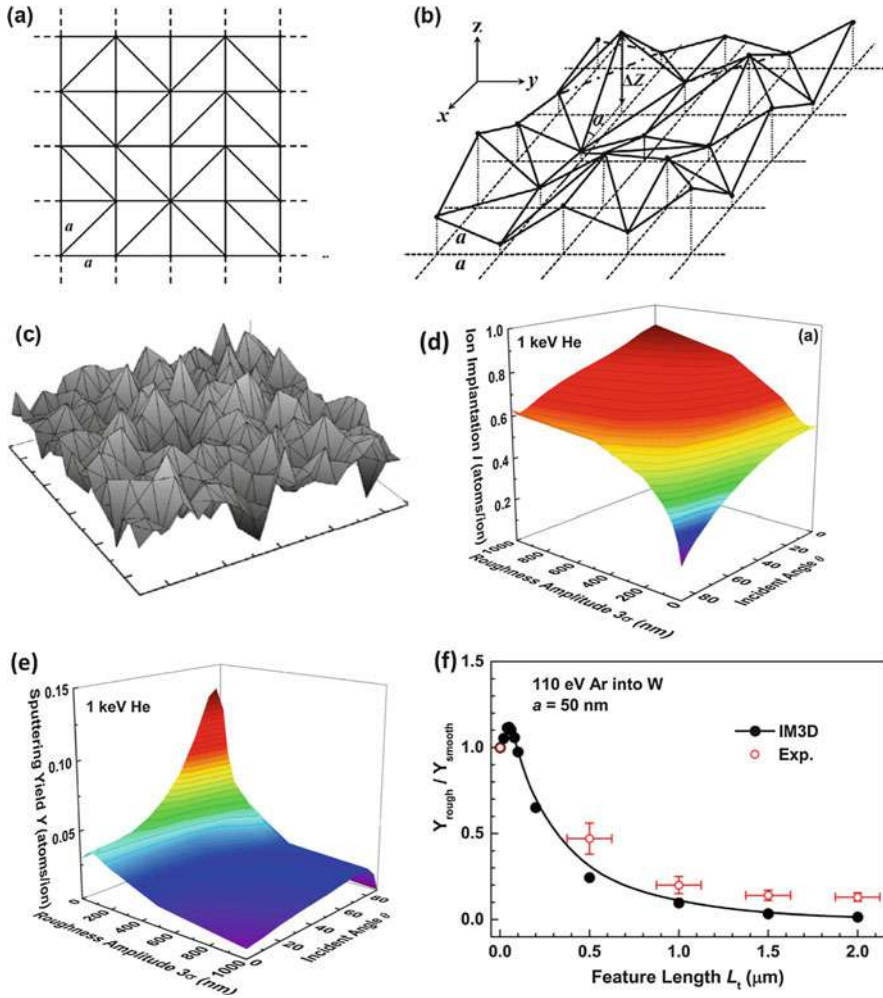
Fig. 10 (a–c) Schematic illustration of the surface roughness generated by the FETM method. Dependence of the ion retention (d) and sputtering yield (e) on the roughness amplitude and incident angle (1 keV He → W). (f) Effect of the surface roughness on the sputtering yield and validation of the IM3D simulations (1 keV He → W) (Li et al. 2017)

3.2 3D Geometry Representation and Ray Tracing in Full-3D BCA-MC Simulations

A key step in full-3D simulation is to build the 3D structure of target. It is significant for the "language" of 3D geometry description to be general and flexible. Also, during the simulation, after sampling the free-flight-path L to travel in the next step, one should determine if the projectile will enter another material or not during the

path L. If a boundary is crossed, the electronic energy loss will need correction because the free-flight-path L covers regions of different materials with different stopping capabilities. Such treatment is involved in the process called "particle tracking" or "ray tracing," and its method varies in different codes. For example, in IM3D (Li et al. 2015), the program follows a similar methodology to CD-SEM (Li et al. 2008). It calculates the sub-free-flight-path length L_i (where $L = L_1 + L_2 \sim + L_i \sim + ... + L_n$, and L_1 is the distance to the closest boundary along the direction of the ion's path at the starting point of L) in each region covered by L, according to the electronic stopping power ratios between different regions. The total electronic energy loss is corrected accordingly. Besides, the FETM module in IM3D accounts for refraction at each interface, *i.e.*, change of flight path direction due to entering different materials. In *iradina* (Borschel and Ronning 2011), the correction of electronic stopping is ignored because their free-flight-paths and ion energies are so small that nuclear stopping is the dominant mechanism of energy loss. In neutron transport simulation codes such as OpenMC (Romano and Forget 2013; MIT 2011), the free-flight-path is updated to be the distance to the closest boundary, followed by moving the ion to the intersection between the boundary and the free-flight-path, and then re-sampling L using the new material's properties. The readers are referred to Chap. 8.3 of Bielajew (2001) for details about the treatment in such neutron simulations. The selection of the 3D geometry representation model and the ray-tracing method are closely related. Here we will briefly introduce two kinds of 3D geometry representation methods with some common practices of ray-tracing.

1. The voxel method. First define the whole target space as a rectangular box, then subdivide it into equal-sized rectangular voxels (cells). Each cell is assigned to a pre-defined material, which may be a vacuum. In this case, a curved surface is approximated by a stepped surface. The progression of the ion along a path L is composed of a sequence of finding the intersection of rays with the next surrounding voxel, followed by moving the ion to the next voxel until the entire path length L is finished or a "boundary crossing" event occurs. A "boundary crossing" event means that the path L passes though voxels of different materials. The simulation accuracy of these methods increases with the number of voxels, while the computation load increases at the same time. Note that there is a lower limit of the voxel size in TRI3DRYN because of the consideration of dynamic evolution (Möller 2014). The voxel method has the advantage of easy implementation of dynamic evolution. For example, the formation of voids in the bulk can be simulated by reducing the density of certain voxels through some preset criteria. However, the voxel method may lead to unavoidable artifacts. For instance, when predicting the sputtering yield of the stepped planes, ions that have exited may re-enter and create further unphysical sputtering/cascades. Also, there is still a lack of understanding about the correction of electronic stopping power at the step edges (Li et al. 2015). Last but not least, domain size adaptivity and scaling issues should also be considered (Borschel and Ronning 2011; Möller 2014), adding to the complexity of simulations.

2. The surface method. The 3D structure is constructed by defining the surfaces of the target in real-space (x, y, z) using mathematical (vector) descriptors. For example, the vector descriptor for the surface of a spherical nanoparticle centered at the origin with radius R is $x^2 + y^2 + z^2 = R^2$. If an ion is at position (x_0, y_0), then $x_0^2 + y_0^2 + z_0^2 < R^2$, $x_0^2 + y_0^2 + z_0^2 = R^2$ and $x_0^2 + y_0^2 + z_0^2 > R^2$ means the ion is inside the nanoparticle, at the surface of the nanoparticle and outside the nanoparticle, respectively. Suppose an ion bombards an atom at (x_0, y_0), after which the new speed of the ion is $\vec{v} \equiv (u, v, w)$. We then sample the distance to travel next with a result of L. To determine if the ion will enter or exit the surface, one should first compute the minimum distance to the surface of the nanoparticle using (x_0, y_0) and $\vec{v} \equiv (u, v, w)$ as $d = -\Lambda \pm \sqrt{\Lambda^2 - \varpi}$, where $\Lambda = x_0 u + y_0 v + z_0 w$, and $\varpi = x_0^2 + y_0^2 + z_0^2 - R^2$. If there is no real solution, it means that the ray will never intersect the sphere, no matter how large L is. If $\varpi < 0$, the ion is inside the sphere at the beginning, so there will be two solutions (one positive and one negative). If $\varpi > 0$, the ion is outside the particle; thus, the solutions are both positive or negative. If $L > d > 0$, a "boundary crossing" event will occur. When the target structure is more complicated, more surfaces (mathematical descriptors) are required. If there is more than one surface, one should compute the minimum distance to each surface along the ray and sort them in order to identify the sequence of surfaces that the ion will intersect along the path L. There are several ways to construct arbitrary target structures using the surface method, here we will just briefly introduce two of them: the constructive solid geometry (CSG) method (Li and Ding 2006) and the previously introduced FETM method (see Sect. 3.1) (Zhang et al. 2012). IM3D supports both the CSG and the FETM methods. The CSG method constructs the complicated target surfaces by Boolean operations on simple geometries (such as the sphere, tetrahedron, cuboid, ellipsoid, and taper). These operations include union, difference, and intersection. Therefore, for simple nanostructures such as nanoparticles or nanowires, it provides the best balance between convenience and accuracy. The reader is referred to the documentation webpage of OpenMC (MIT 2011) for details about the CSG method. The FETM method, on the other hand, is more suitable for building arbitrarily complex structures. The triangular mesh for a complicated structure can be easily generated by programs such as Gmsh (Geuzaine and Remacle 2009). Also, triangular mesh enables easier determination of intersection points of a straight line with surfaces. The FETM method is further aided by a spatial subdividing technique (Li et al. 2008) to avoid the computation of intersections with all triangular meshing surfaces. Only the surfaces contained in the cubes along the path L are considered for such intersection searching and distance computation. For complicated geometries, the number of times for finding the intersection points with surfaces on a free-flight-path L using the FETM method is usually much smaller than that using the CSG method, leading to a higher computational efficiency. The FETM method can be extended to the finite element tetrahedron method for easier implementation of dynamic evolution in materials during irradiation.
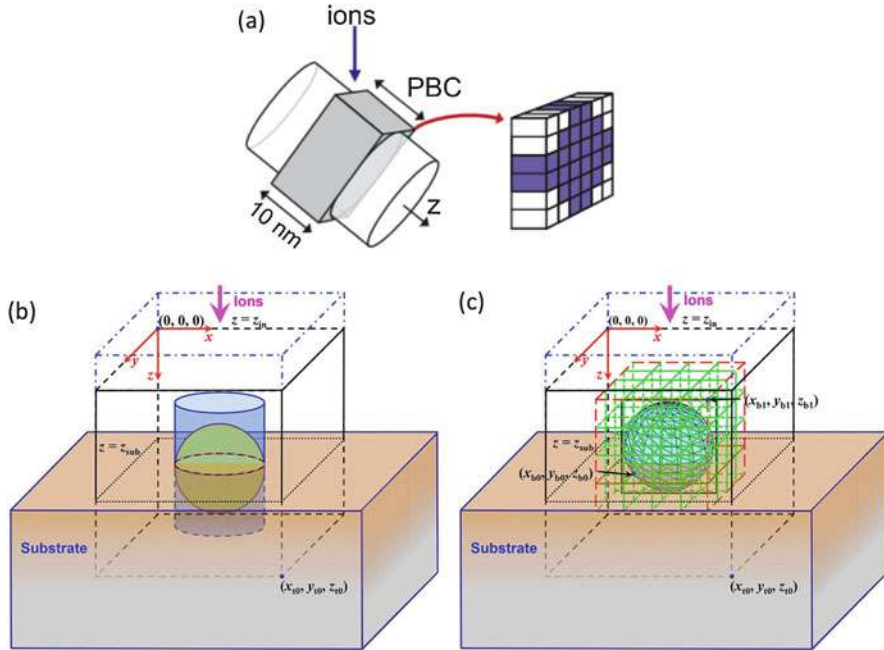
Fig. 11 Schematic illustrations of different geometry construction method: (**a**) voxels (Borschel and Ronning 2011); (**b**) CSG and (**c**) FETM (Li et al. 2015)

The differences between the voxel method and the surface method are illustrated in Fig. 11. Currently, available full-3D MC codes for the simulation of energetic particles in matter include MCNP (LANL 2018), OpenMC (Romano and Forget 2013; MIT 2011), Geant4 (Agostinelli et al. 2003), *iradina* (Borschel and Ronning 2011), TRI3DYN (Möller 2014), IM3D (Li et al. 2015), MyTRIM (Schwen et al. 2009 and 2018), and Corteo 3D (Schiettekatte and Chicoine 2016). The reader is referred to the documentation of these codes for more details about their methods for particle tracing and 3D geometry construction.

3.3 Dynamically Evolving Full-3D BCA Simulation Structures

Under high fluence of ion beam radiation, the target can be modified, leading to changes in the morphology and density in material due to sputtering and relocation of atoms by atomic collisions. A "dynamic" code is designed to take into account this evolution of the target during ion implantation, with well-documented application examples where such dynamic evolution is absolutely necessary to correctly simulate experimental results. Examples include the secondary ion mass spectrometry (SIMS) depth profiling technique, plasma–wall interactions in fusion reactors, and magnetron sputtering for thin film processing. TRIDYN (Möller and Eckstein 1984) (1D simulation) and TRI3DYN (Möller 2014) (full-3D simula-

Fig. 12 Schematic of the voxel types in TRI3DRYN. *Dark blue*, *light blue*, and *white* represent bulk, surface, and vacuum voxels, respectively (Möller 2014)
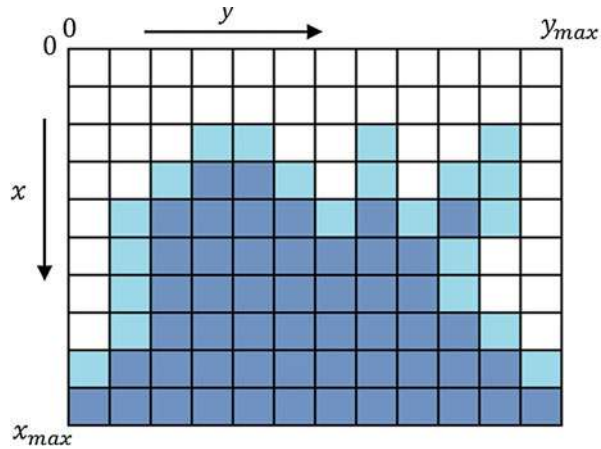
tion) have been developed to model this dynamic evolution under ion radiation. TRI3DYN uses the voxel method (see Sect. 3.2) for the construction of target 3D structures. Here we will briefly introduce TRI3DYN, for more details please refer to Möller (2014).

In TRI3DYN, the voxels are classified into three types: bulk, surface, and vacuum, as shown in Fig. 12. The neighborhoods of a voxel (say "voxel 1") are classified by the distance between voxel 1 and its neighborhood voxels. There are 6 first neighbors, 12 second neighbors, and 8 third neighbors. A lower limit for the voxel size is set ($\lambda < \min(\Delta x, \Delta y, \Delta z)$), where λ is the mean free path and $\Delta x, \Delta y, \Delta z$ are the size of voxel along the x, y, z directions, respectively, so that the path length L between two successive collisions always ends within the volume that is composed of the 26 first-to-third nearest neighbors. Suppose there are N_c types of component in the material, then the partial atomic density of a component l ($1 \leq l \leq N_c$) in the voxel with spatial index (i, j, k) is $n_l(i, j, k)$. TRI3DYN and TRIDRYN both assume that after the relaxation, a constant atomic volume is reached with the following relation:

$$\frac{1}{n} = \sum_l \frac{q_l}{n_l^0} \tag{2}$$

where n, n_l^0, and q_l denote the total atomic density, a predefined atomic density of component l and atomic fractions of component l ($0 \leq q_l \leq 1$), respectively. With this condition at a relaxed state, one can define a reduced partial density:

$$d = \sum_l d_l = \sum_l \frac{n_l}{n_l^0} \tag{3}$$

where d_l is the partial density of the component l. When $n_l = q_l \cdot n$, a "nominal" density is reached such that $d = 1$. Based on this, $d = 0$, $d > 1$, and $0 < d < 1$

represent "vacuum," "over-dense," and "under-dense," respectively. The simulation procedure is an iteration of the following pseudocode steps:

1. Perform a SRIM/TRIM-like BCA simulation for the cascade resulting from one (default) or more incident ions. Record point defect distributions (interstitial atoms and vacancies).
2. Compensation and recombination of interstitials and vacancies. First, the interstitials and vacancies within the same voxel will recombine. Next, the program will search for neighboring voxels within a predefined recombination radius and enable recombination.
3. Incorporation of the remaining point defects from step (2) into the voxels.
4. Relaxation of the voxels, so that nominal density ($d = 1$) is obtained for each bulk voxel, while the surface voxel is permitted to be under-dense ($d < 1$).
5. "Local surface smoothing" will be performed if this kernel is activated. Examples of smoothing methods are illustrated in Fig. 13.
6. Optionally, adjust the surface layers so that the top two layers are kept a vacuum.

Ion implantation is a popular tool to tailor the composition and morphology of materials (Krasheninnikov and Nordlund 2010; Borschel et al. 2011) and is promising to improve the functional properties of nanostructured materials, such as using Ar^+ to modify the electronic properties of single-wall carbon nanotube (Gómez-Navarro et al. 2005) or single layer MoS_2 (Chen et al. 2018). An example of the simulation by TRI3DRYN is shown in Fig. 14, where P^+ at 10 keV is modifying the morphology and composition gradually. The 3D surface morphology evolution of the nanoparticle (changing from a sphere to pillar) can be well captured by the TRI3DYN simulation.

Fig. 13 Schematic of the "local surface smoothing" in TRI3DRYN. (**a**) "hanging" or (**b**) "flying" occurs rarely during the simulations (Möller 2014)
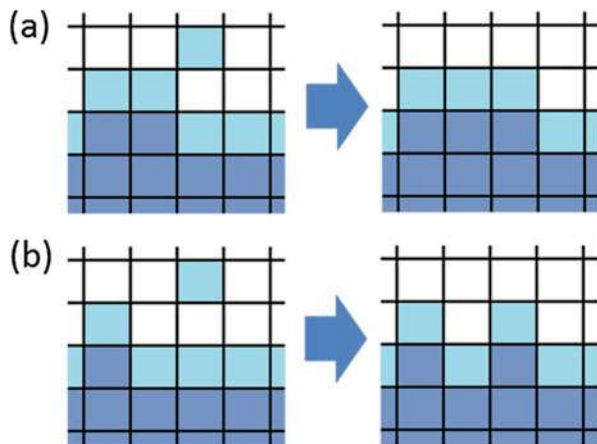
Fig. 14 Evolution of a spherical Si nanoparticle (diameter = 30 nm) on SiO_2 substrate during the implantation of 10 keV P, whose initial velocity is normal to the substrate surface. (**a–f**) Si atomic fraction. Color coding: 0.333 (*red*) → 1 (*blue*). (**g–k**) P atomic fraction. Color coding: 0 (*red*) → 0.1 (*blue*) (Möller 2014)
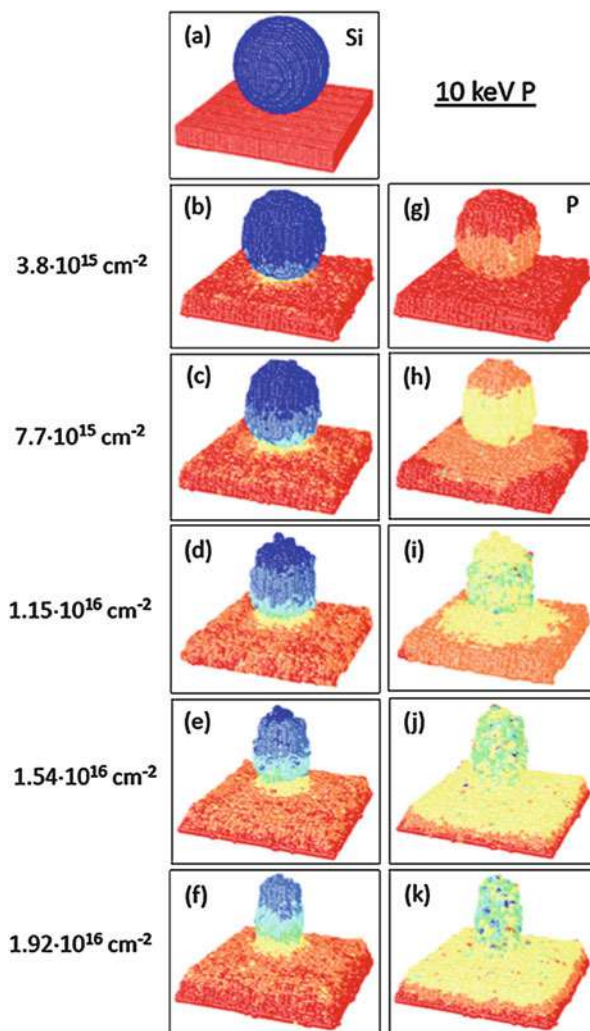
3.4   Maintaining 3D Simulation Efficiency

The usage of full-3D simulations and consideration of dynamic evolution will potentially increase the computational load. The increase of computational time can be 10–10,000× for a 3D target with a complicated surface morphology. Therefore, it is vital to enhance the efficiency of the simulation codes. Here we will briefly introduce the methods adopted in static BCA-MC simulations (such as Corteo and IM3D) for the improvement of computational efficiency.

One of the most computationally expensive parts in the simulation of radiation damage cascade is the sampling of θ_C. Traditionally, this requires one to solve the scattering integral equations, leading to a significant computational load. SRIM uses the so-called "MAGIC" formula to sample the scattering and azimuthal angles for better efficiency. Based on it, a "fast indexing" technique using database evaluation was latter proposed (Yuan et al. 1993) to further enhance the efficiency to about 18 times faster than the "MAGIC" formula. Instead of computing directly from the "MAGIC" algorithm, this method utilizes interpolated tables for the scattering angle $\sin^2(\theta_c)$ and looks up the needed values in the tables. It is worthwhile to note that the table is indexed using the binary representation of floating point numbers so that computing logarithms can be avoided (Schiettekatte 2009). This method also improves accuracy and memory usage. Similarly, one can obtain the electronic stopping power using precalculated tables (generated using "SRModule.exe" (Ziegler 2004)). This "fast indexing" method is now available in Corteo (Schiettekatte 2009), *iradina* (Borschel and Ronning 2011), and IM3D (Li et al. 2015). A comparison of the efficiency is provided in Fig. 15. The readers can refer to Schiettekatte (2009) for more details about this method. In addition, a ray-tracing technique based on the FETM 3D geometry representation and spatial subdivision is introduced in IM3D (Li et al. 2015), boosting the computational efficiency of complex 3D geometry simulations to a level similar to that in bulk simulations.

For a static BCA Monte Carlo simulation, the history of each ion is independent. Therefore, it is well suited for parallel computation. IM3D (Li et al. 2015) in particular is implemented with the Message Passing Interface (MPI) library. Its parallel scaling performance, as shown in Fig. 16, shows a greatly enhanced speed, up to 10^5 faster than SRIM.

a

2 MeV Au ions

305nm

ZrO$_2$ Si

10^5 ions

b

Method	SRIM (serial only)	*Iradina* (serial only)	IM3D
FC (sec)	473269	12089 (slow), 6345 (fast)	5760
KP (sec)	11758	—	59
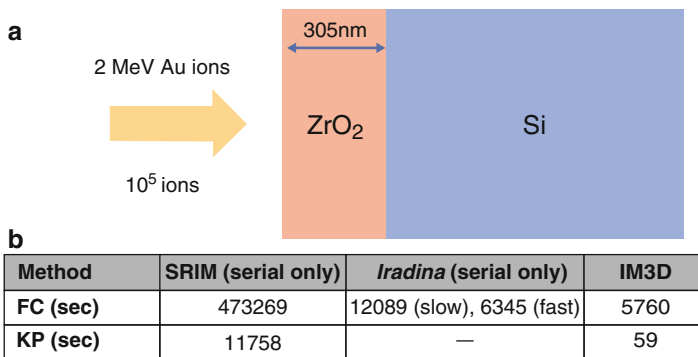
Fig. 15 A comparison of the computational efficiency (single CPU). (**a**) Schematic illustration of the simulation setup: 2 MeV Au ion irradiation on a dual-layer substrate. The top layer is a ZrO$_2$ film with a thickness of 305 nm and Si is infinitely thick. (**b**) Comparison of the time for the simulation of 10^5 ions. Reproduced from Ref. (Li et al. 2015), which is licensed under CC BY 4.0

Fig. 16 Wall clock scaling of IM3D running Message Passing Interface (MPI), for the system of a 305 nm ZrO_2 film on Si under a total of 100000 Au ion irradiations with ion energies of 2.0 MeV. Reproduced from Ref. (Li et al. 2015), which is licensed under CC BY 4.0
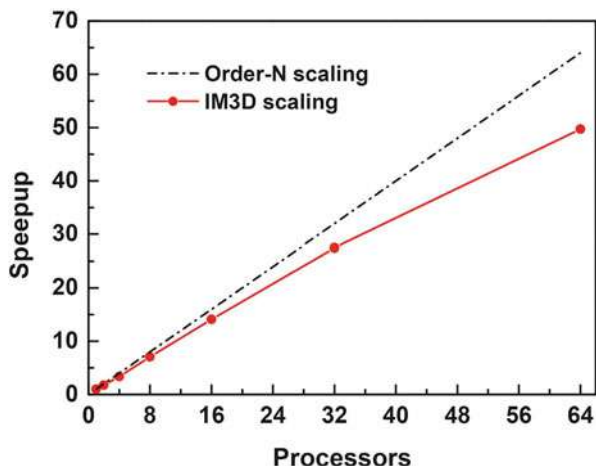
3.5   Open Sourcing and Stopping Power Databases

An open source code can greatly facilitate the development of advanced codes, as it allows an open inspection and validation of the underlying algorithms. More and more codes have joined the open-source family, including Corteo (Schiettekatte 2009), *iradina* (Borschel and Ronning 2011), MyTRIM (Schwen 2018), Mat-TRIM (Yang et al. 2018b), and IM3D (Li et al. 2015), etc. One of the main advantages of open-source is that one can easily modify or incorporate new stopping power data into the code. It is known that the accuracy of BCA Monte Carlo simulation codes for ion radiation rely greatly on the accuracy of the stopping power data. Although SRIM has provided a powerful stopping power database (Ziegler 2004), it was recently shown to be inaccurate in several circumstances. For example, for the case of very heavy ions into a light element target, a large discrepancy is shown between the ion ranges predicted by SRIM and experimental results (Behar et al. 1985; Grande et al. 1988; Friedland et al. 1998; Zhang et al. 2009, 2010; Jin et al. 2014) due to errors in the stopping power database. An example of heavy ions into SiC is shown in Fig. 17. Because of this, M-TRIM (Jin et al. 2014) has been developed based on an early version of TRIM (Ziegler et al. 1985) with updated stopping power data for Cl, Br, I, and Au ions into SiO_2 and SiC. The comparison between the ion distribution simulated by SRIM and M-TRIM and the experimental results obtained by Rutherford backscattering spectrometry (RBS) and secondary ion mass spectrometry (SIMS) are shown in Fig. 17a.

Electronic stopping powers for bulk materials may also need to be corrected for nanostructured materials. Recently, the electronic energy loss for graphene (see Fig. 18) and boron nitride (BN) sheets have been studied (Zhao et al. 2015) using time-dependent DFT. However, there is still a lack of study of the applicability of
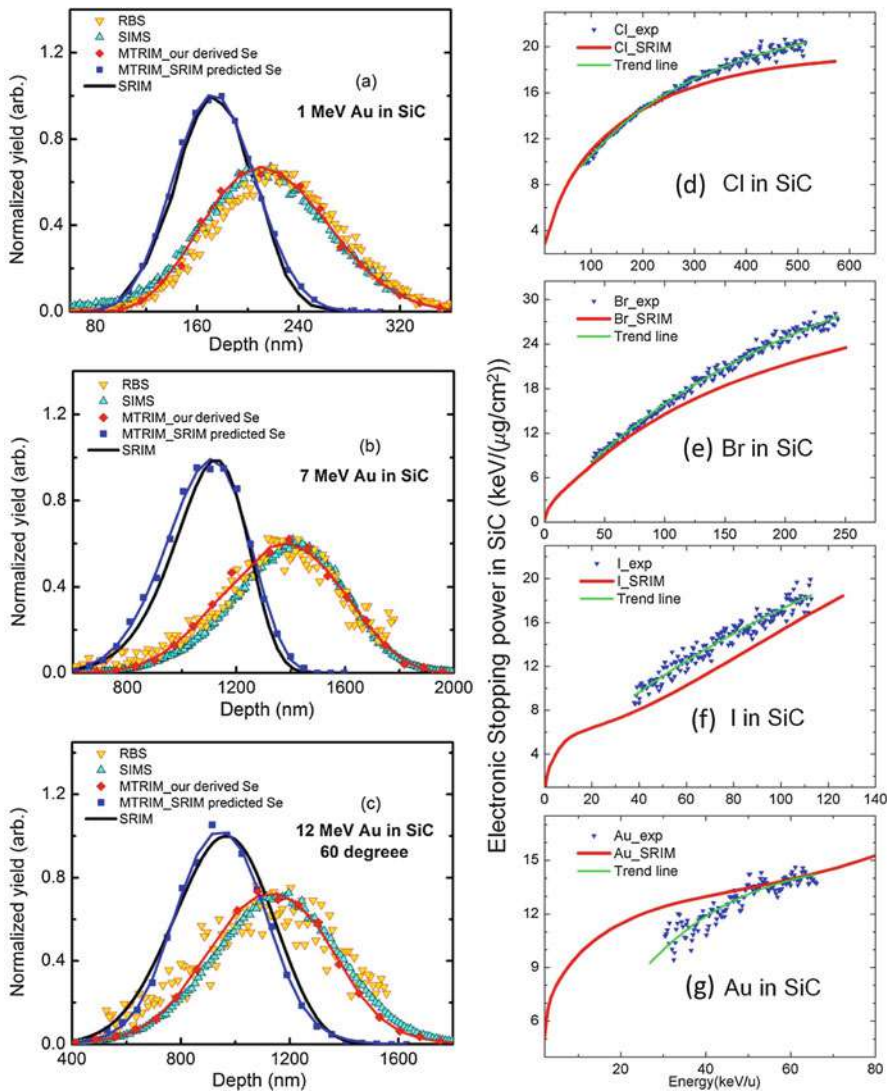
Fig. 17 (**a–c**) Injected ion distribution of 1 MeV, 7 MeV, and 12 MeV Au into SiC: a comparison between the simulations and experimental results. (**d–g**) Comparison of the electronic stopping power of Cl, Br, I, and Au into SiC between the experimental results and SRIM database (Jin et al. 2014)

bulk stopping power database in nanostructured systems. Currently, Corteo, *iradina*, and IM3D use the same stopping power tables generated by Corteo using the "SRModule.exe" (Ziegler 2004). Since these codes are open-source, in principle it is possible to generate new stopping power tables. However, there is still a need

Fig. 18 Comparison of transferred energy as a function of proton energy. The target is graphene and the collision site is point C (Zhao et al. 2015)
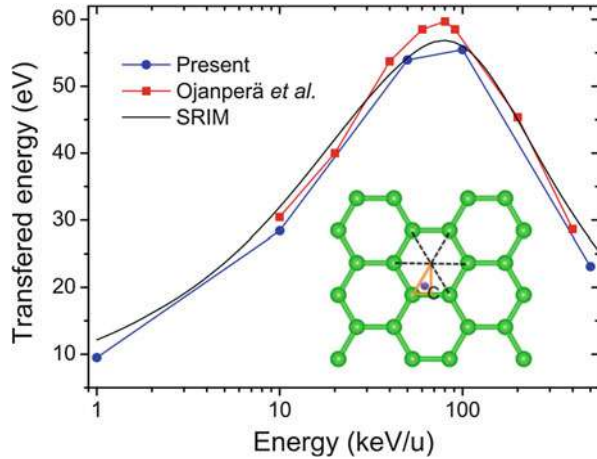
of a friendly user-interface (UI) that can help generate and conveniently manage the stopping power data for these codes.

4   Comparisons Between MC and Molecular Dynamics (MD)

Radiation damage in nanostructured materials can also be simulated by MD (Jin et al. 2018). Here we will briefly review the difference between MC and MD simulations. MC codes are generally based on more assumptions than MD. MC codes usually only consider two-body interactions and use universal potentials in many cases. Because of its high computational efficiency, MC simulations of ion irradiation serve as convenient engineering tools for the prediction of ion ranges and vacancy/energy-deposition distributions. On the other hand, MD simulations are based on solving equations of motion (Li 2005), which is in general more accurate as they consider many-body interactions and can provide more details about the cascade process; however, MD is much more computationally intense than MC, limiting the time and length scales simulated. It should be noted that the accuracy of MD relies on the accuracy of the interatomic potential used. Also, it has recently been shown that the electronic energy loss should be taken into account in MD simulations, while the majority of MD studies on radiation damage cascades ignore this effect (Rutherford and Duffy 2007; ▶ Chap. 98, "Incorporating Electronic Effects in Molecular Dynamics Simulations of Neutron and Ion-Induced Collision Cascades"). For nanostructured materials whose size is less than the mean-free-path of ions, it was recently shown that electronic stopping is the dominant energy loss mechanism for incident ions (Yang et al. 2018b), indicating the necessity of considering electronic energy loss in MD simulations. The results of MD also depend on the simulation temperature, which is difficult to consider in BCA-MC codes. Almost no BCA-MC simulations consider defect

diffusion and defect-sink interactions (kinetics at longer timescales), while MD simulations more systematically model kinetic processes (especially annealing) in materials if the time and length scales are within computable limits. For instance, previous MD simulations showed that defect production can be greatly enhanced by a large surface-area to volume ratio in nanostructured materials at low incident ion energies (Hoilijoki et al. 2011; Sun et al. 2012), which BCA-MC simulations usually fail to predict. However, the primary radiation damage estimated by MC simulations still has broad and significant engineering implications.

MC simulation is a useful and rapid method to calculate the Norgett–Robinson–Torres (NRT) displacements per atom (DPA), a standardized radiation exposure parameter. The definition of DPA is based on $T = 0$ K MD simulations, using the following procedure: First, give a selective incident particle an initial energy E_0. The particle will collide with multiple atoms in the target until its energy is small enough. Every time when it collides an atom in the lattice, if the kinetic energy transferred to the target atom exceeds the threshold energy E_d, the atom will be knocked out (at $t = t_I \sim$ fs). After t_I, there is spontaneous recombination of the Frenkel pairs if the distance between a vacancy and an interstitial atom is within a certain limit. At $t = t_F \sim$ ps, this athermal dynamic recombination/transient annealing process is finished. After that, long-term defect evolution occurs, which depends on thermal activation.

Among all MC codes, SRIM (Ziegler et al. 2008) is the most widely used for the estimation of DPA because of its powerful stopping power database and its ease of use. SRIM offers two ways for the calculation of NRT DPA: (1) The quick Kinchin–Pease (K-P) mode, where defect production by generation-1 ions are tallied by using the K-P formula instead of explicitly calculated and (2) the full-cascade (FC) mode, where generation-2, -3, -4, . . . ions are also explicitly accounted for by BCA calculations. It has recently been found that the displacements calculated by both modes are significantly larger than the results of MD at ten picoseconds after a PKA when the thermal spike has largely disappeared (Stoller et al. 2013) (Fig. 19a). Also, the displacements by the FC mode are twice of that by K-P mode (Fig. 19a–c). Thus, Stoller et al. questioned the FC mode in SRIM, exposing a fundamental problem in this *closed-source* code. Recently, this discrepancy between K-P mode and FC modes has reproduced by IM3D (Li et al. 2015), a new open-source code.

It is also worthwhile to note that MC codes usually overestimate the number of radiation defects by three times compared to MD, while only capturing 1/30th of the atomic mixing (replacement) events (Nordlund et al. 2018), as shown in Fig. 20a, b. As a result, Nordlund et al. proposed two new complementary displacement production estimators to quantify the radiation damage more realistically: (a) Athermal recombination corrected dpa, namely, "arc-dpa"; and (b) replacements per atom, namely, "rpa." Both quantities rely on parameters gained by fitting MD simulation results. Helpful discussions about the difference between MC and MD for radiation damage studies can also be found in ▶ Chap. 88, "Molecular Dynamics Simulations of Non-equilibrium Systems" within this Handbook.
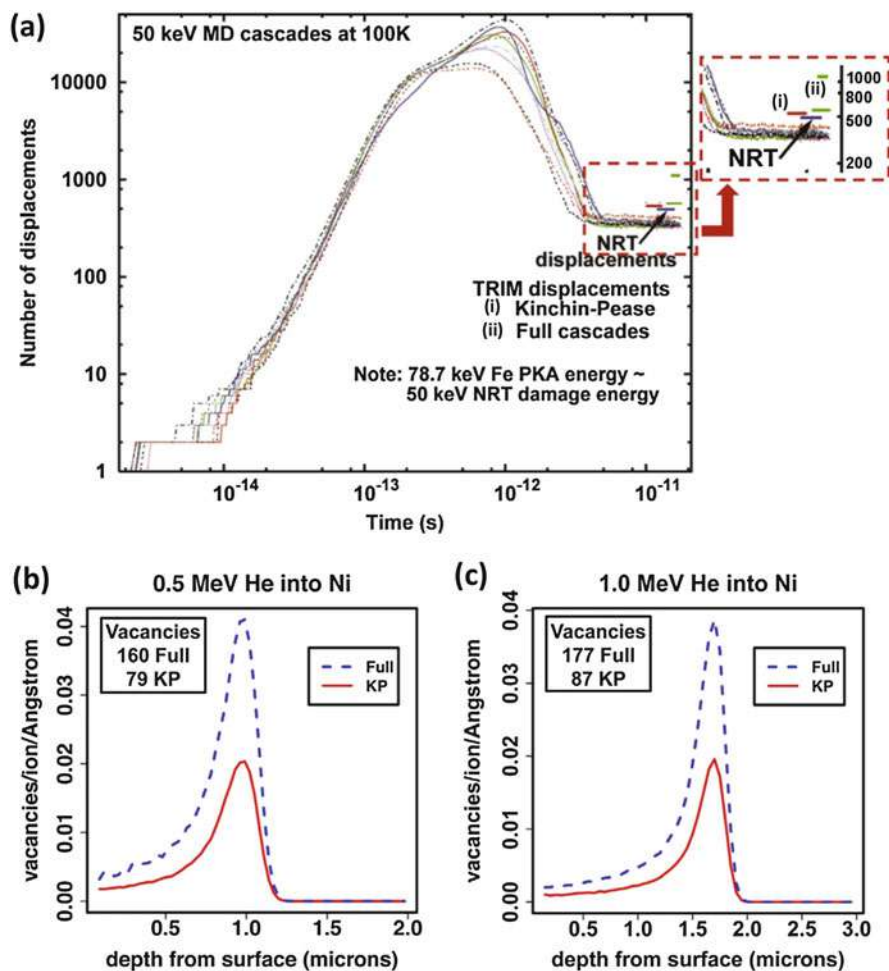
Fig. 19 (**a**) Comparison between the displacements calculated by MD, SRIM (K-P mode), and SRIM (FC mode). (**b**, **c**) Depth distribution of vacancies for He$^+$ into Ni (Stoller et al. 2013)

5   Outlook for the Next 10 Years

In summary, Monte Carlo simulation of ion radiation based on a BCA model is an efficient tool for the prediction of primary damage in materials. A series of advanced codes have been recently developed, enabling full-3D simulations, which can consider arbitrarily complex target morphologies, arbitrary beam shapes/sizes, and even dynamic evolution. The speed of these codes is also greatly enhanced by the fast indexing method of scattering angle sampling and highly parallel computation. Also, more and more advanced codes are becoming open-source, allowing verification of the algorithms and the underlying physics engine. What
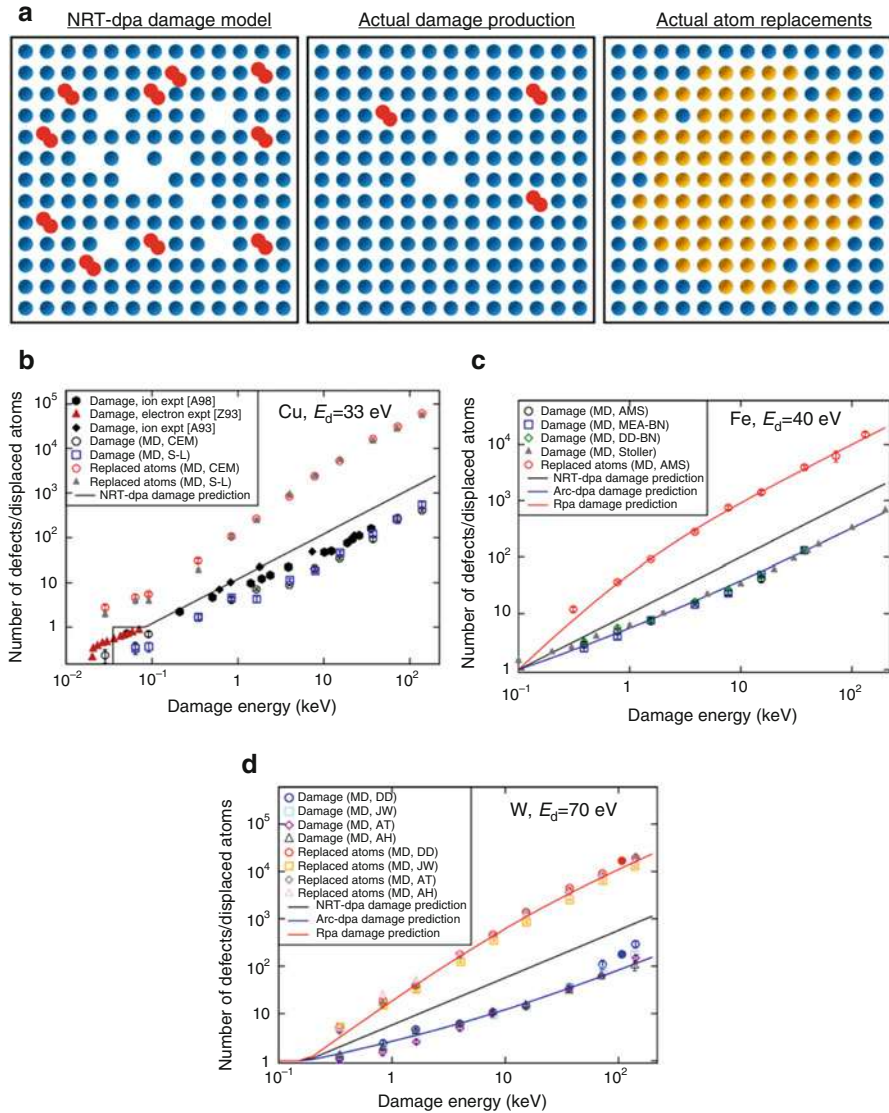
Fig. 20 (**a**) Schematic illustration of the problem of the NRT-dpa damage model. (**b**) Comparison of MD simulation and NRT-dpa model. (**c**, **d**) Two new quantification standards: Arc-DPA and RPA agree well with MD simulations. Reproduced from Ref. (Nordlund et al. 2018), which is licensed under CC BY 4.0

is more, maintaining an open-source code policy will facilitate the incorporation of BCA-MC simulations into the framework of multi-scale and multi-physics frameworks such as MOOSE (Gaston et al. 2009), dynamically providing primary radiation damage data as input for larger time- or length-scale simulations. Besides,

it has been shown that legacy codes such as SRIM will lead to significant errors in many applications, such as nano-beams, nanostructured materials, and rough-surface. However, the traditional closed-source code SRIM still enjoys a relative monopoly due to its long history, its relative accuracy for bulk-scale simulation, and especially its simple and intuitive user interface. It is important for new codes to have such a friendly user interface, allowing not only the easy input of complex target morphologies, but also convenient updates of the stopping power database. Last, but not least, increasing doubt exists on the NRT-DPA model when comparing MC to MD simulations, while the applicability of newly proposed damage units is still under exploration.

References

Agostinelli S, Allison J, Amako K et al (2003) GEANT4 – a simulation toolkit. Nucl Instrum Methods Phys Res A 506:250–303. https://doi.org/10.1016/S0168-9002(03)01368-8

Baldwin MJ, Doerner RP (2010) Formation of helium induced nanostructure "fuzz" on various tungsten grades. J Nucl Mater 404:165–173. https://doi.org/10.1016/j.jnucmat.2010.06.034

Bayn I, Chen EH, Trusheim ME et al (2015) Generation of ensembles of individually resolvable nitrogen vacancies using nanometer-scale apertures in ultrahigh-aspect ratio planar implantation masks. Nano Lett 15:1751–1758. https://doi.org/10.1021/nl504441m

Behar M, Fichtner PF, Olivieri CA et al (1985) Range profiles of implanted Bi and Au in amorphous silicon. Nucl Instrum Methods Phys Res B 6:453–458. https://doi.org/10.1016/0168-583X(85)90002-3

Bielajew AF (2001) Fundamentals of the Monte Carlo method for neutral and charged particle transport. The University of Michigan

Biersack JP, Haggmark LG (1980) A Monte Carlo computer program for the transport of energetic ions in amorphous targets. Nucl Inst Methods 174:257–269. https://doi.org/10.1016/0029-554X(80)90440-1

Borschel C, Ronning C (2011) Ion beam irradiation of nanostructures – a 3D Monte Carlo simulation code. Nucl Instrum Methods Phys Res B 269:2133–2138. https://doi.org/10.1016/j.nimb.2011.07.004

Borschel C, Spindler S, Lerose D et al (2011) Permanent bending and alignment of ZnO nanowires. Nanotechnology 22:185307. https://doi.org/10.1088/0957-4484/22/18/185307

Briesmeister JF (1986) MCNP – a general Monte Carlo code for neutron and photon transport. Technical report. Los Alamos National Laboratory, Los Alamos

Chen Y, Huang S, Ji X et al (2018) Tuning electronic structure of single layer MoS_2 through defect and interface engineering. ACS Nano 12(3):2569–2579. https://doi.org/10.1021/acsnano.7b08418

Crocombette JP (2005) Cascade Modeling. In: Yip S (eds) Handbook of materials modeling. Springer, Dordrecht, pp 987–997. https://doi.org/10.1007/978-1-4020-3286-8_49

Eckstein W (1991) Computer simulation of ion-solid interactions. Springer, Berlin/Heidelberg

Economou NP, Notte JA, Thompson WB (2012) The history and development of the helium ion microscope. Scanning 34:83–89. https://doi.org/10.1002/sca.20239

Friedland E, Kalbitzer S, Hayes M et al (1998) Range parameters of gold and lead in carbon and carbon in gold at reduced energies of $10^{-3} < \epsilon < 1$. Nucl Instrum Methods Phys Res B 136–138:147–152. https://doi.org/10.1016/S0168-583X(97)00833-1

Gaston D, Newman C, Hansen G, Lebrun-Grandié D (2009) MOOSE: A parallel computational framework for coupled systems of nonlinear equations. Nucl Eng Des 239:1768–1778. https://doi.org/10.1016/j.nucengdes.2009.05.021

Geuzaine C, Remacle J-F (2009) Gmsh: a 3-D finite element mesh generator with built-in pre- and post-processing facilities. Int J Numer Methods Eng 79:1309–1331. https://doi.org/10.1002/nme.2579

Gilmer G, Yip S (2005) Basic Monte Carlo models: equilibrium and kinetics. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht, pp 613–628

Gómez-Navarro C, de Pablo PJ, Gómez-Herrero J et al (2005) Tuning the conductance of single-walled carbon nanotubes by ion irradiation in the Anderson localization regime. Nat Mater 4:534–539. https://doi.org/10.1038/nmat1414

Grande PL, Fichtner PFP, Behar M, Zawislak FC (1988) Range profiles of medium and heavy ions implanted into SiO_2. Nucl Inst Methods Phys Res B 35:17–20. https://doi.org/10.1016/0168-583X(88)90093-6

Hoffmann S, Bauer J, Ronning C et al (2009) Axial p-n junctions realized in silicon nanowires by ion implantation. Nano Lett 9:1341–1344. https://doi.org/10.1021/nl802977m

Hoilijoki S, Holmström E, Nordlund K (2011) Enhancement of irradiation-induced defect production in Si nanowires. J Appl Phys 110:43540. https://doi.org/10.1063/1.3627234

Holland-Moritz H, Scheeler S, Stanglmair C et al (2015) Enhanced sputter yields of ion irradiated Au nano particles: energy and size dependence. Nanotechnology 26:325301. https://doi.org/10.1088/0957-4484/26/32/325301

Hu A, Hassanein A (2012) How surface roughness affects the angular dependence of the sputtering yield. Nucl Instrum Methods Phys Res B 281:15–20. https://doi.org/10.1016/j.nimb.2012.03.026

Jin K, Zhang Y, Zhu Z et al (2014) Electronic stopping powers for heavy ions in SiC and SiO_2. J Appl Phys 115:044903. https://doi.org/10.1063/1.4861642

Jin M, Cao P, Yip S, Short MP (2018) Radiation damage reduction by grain-boundary biased defect migration in nanocrystalline Cu. Acta Mater 155:410–417. https://doi.org/10.1016/j.actamat.2018.05.071

Kenmotsu T, Yamamura Y, Muramoto T, Hirotani N (2005) Simulation studies on sputtering in rough surface. Nucl Instrum Methods Phys Res B 228:369–372. https://doi.org/10.1016/j.nimb.2004.10.072

Krasheninnikov AV, Nordlund K (2010) Ion and electron irradiation-induced effects in nanostructured materials. J Appl Phys 107:71301. https://doi.org/10.1063/1.3318261

Krasheninnikov AV, Banhart F, Li JX et al (2005) Stability of carbon nanotubes under electron irradiation: role of tube diameter and chirality. Phys Rev B 72:125428. https://doi.org/10.1103/PhysRevB.72.125428

LANL (2018) A general Monte Carlo N-Particle (MCNP) transport code. https://laws.lanl.gov/vhosts/mcnp.lanl.gov/index.shtml. Accessed 11 Oct 2018

Li J (2005) Basic molecular dynamics. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht, pp 565–588

Li HM, Ding ZJ (2006) Monte Carlo simulation of secondary electron and backscattered electron images in scanning electron microscopy for specimen with complex geometric structure. Scanning 27:254–267. https://doi.org/10.1002/sca.4950270506

Li YG, Mao SF, Li HM, et al (2008) Monte Carlo simulation study of scanning electron microscopy images of rough surfaces. J Appl Phys 104:064901. https://doi.org/10.1063/1.2977745

Li YG, Yang Y, Short MP et al (2015) IM3D: a parallel Monte Carlo code for efficient simulations of primary radiation displacements and damage in 3D geometry. Sci Rep 5:1–13. https://doi.org/10.1038/srep18130

Li Y, Yang Y, Short MP et al (2017) Ion radiation albedo effect: influence of surface roughness on ion implantation and sputtering of materials. Nucl Fusion 57:16038. https://doi.org/10.1088/1741-4326/57/1/016038

Liontas R, Gu XW, Fu E et al (2014) Effects of helium implantation on the tensile properties and microstructure of Ni73P27 metallic glass nanostructures. Nano Lett 14:5176–5183. https://doi.org/10.1021/nl502074d

Maletinsky P, Hong S, Grinolds MS et al (2012) A robust scanning diamond sensor for nanoscale imaging with single nitrogen-vacancy centres. Nat Nanotechnol 7:320–324. https://doi.org/10.1038/nnano.2012.50

MIT (2011) OpenMC: theory and methodology. https://openmc.readthedocs.io/en/stable/methods/index.html. Accessed 11 Oct 2018

Möller W (2014) TRI3DYN – collisional computer simulation of the dynamic evolution of 3-dimensional nanostructures under ion irradiation. Nucl Instrum Methods Phys Res B 322:23–33. https://doi.org/10.1016/j.nimb.2013.12.027

Möller W, Eckstein W (1984) Tridyn – a TRIM simulation code including dynamic composition changes. Nucl Inst Methods Phys Res B 2:814–818. https://doi.org/10.1016/0168-583X(84)90321-5

Nordlund K, Björkas C, Ahlgren T et al (2014) Multiscale modelling of plasma-wall interactions in fusion reactor conditions. J Phys D Appl Phys 47:224018. https://doi.org/10.1088/0022-3727/47/22/224018

Nordlund K, Zinkle SJ, Sand AE et al (2018) Improving atomic displacement and replacement calculations with physically realistic damage models. Nat Commun 9:1084. https://doi.org/10.1038/s41467-018-03415-5

Odette GR, Wirth BD (2005) Radiation effects in fission and fusion reactors. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht, pp 999–1037

Pekin TC, Allen FI, Minor AM (2016) Evaluation of neon focused ion beam milling for TEM sample preparation. J Microsc 264:59–63. https://doi.org/10.1111/jmi.12416

Pezzagna S, Wildanger D, Mazarov P et al (2010) Nanoscale engineering and optical addressing of single spins in diamond. Small 6:2117–2121. https://doi.org/10.1002/smll.201000902

Pezzagna S, Rogalla D, Becker HW et al (2011) Creation of colour centres in diamond by collimated ion-implantation through nano-channels in mica. Phys Status Solidi A 208:2017–2022. https://doi.org/10.1002/pssa.201100455

Romano PK, Forget B (2013) The OpenMC monte carlo particle transport code. Ann Nucl Energy 51:274–81

Rutherford AM, Duffy DM (2007) The effect of electron-ion interactions on radiation damage simulations. J Phys Condens Matter 19:496201. https://doi.org/10.1088/0953-8984/19/49/496201

Ruzic DN (1990) The effects of surface roughness characterized by fractal geometry on sputtering. Nucl Inst Methods Phys Res B 47:118–125. https://doi.org/10.1016/0168-583X(90)90019-Q

Scarabelli D, Trusheim M, Gaathon O et al (2016) Nanoscale engineering of closely-spaced electronic spins in diamond. Nano Lett 16:4982–4990. https://doi.org/10.1021/acs.nanolett.6b01692

Schiettekatte F (2008) Fast Monte Carlo for ion beam analysis simulations. Nucl Instrum Methods Phys Res B 266:1880–1885. https://doi.org/10.1016/j.nimb.2007.11.075

Schiettekatte F (2009) Monte Carlo simulation of ion beam analysis spectra using Corteo. https://www.lps.umontreal.ca/s̃chiette/index.php?n=Recherche.Corteo

Schiettekatte F, Chicoine M (2016) Spectrum simulation of rough and nanostructured targets from their 2D and 3D image by Monte Carlo methods. Nucl Instrum Methods Phys Res B 371:106–110. https://doi.org/10.1016/j.nimb.2015.09.089

Schwen D (2018) MyTRIM. https://github.com/idaholab/mytrim. Accessed 11 Oct 2018

Schwen D, Huang M, Bellon P, Averback RS (2009) Molecular dynamics simulation of intragranular Xe bubble re-solution in UO$_2$. J Nucl Mater 392:35–39. https://doi.org/10.1016/j.jnucmat.2009.03.037

Shinada T, Okamoto S, Kobayashi T, Ohdomari I (2005) Enhancing semiconductor device performance using ordered dopant arrays. Nature 437:1128–1131. https://doi.org/10.1038/nature04086

Short MP, Gaston DR, Jin M et al (2015) Modeling injected interstitial effects on void swelling in self-ion irradiation experiments. J Nucl Mater 471:200–207. https://doi.org/10.1016/j.jnucmat.2015.10.002

Stoller RE, Toloczko MB, Was GS et al (2013) On the use of SRIM for computing radiation damage exposure. Nucl Instrum Methods Phys Res B 310:75–80. https://doi.org/10.1016/j.nimb.2013.05.008

Sun L, Lan C, Zhao S et al (2012) Self-irradiation of thin SiC nanowires with low-energy ions: a molecular dynamics study. J Phys D Appl Phys 45:135403. https://doi.org/10.1088/0022-3727/45/13/135403

Tan S, Livengood R, Hack P et al (2011) Nanomachining with a focused neon beam: a preliminary investigation for semiconductor circuit editing and failure analysis. J Vac Sci Technol B 29:06F604. https://doi.org/10.1116/1.3660797

Toyli DM, Weis CD, Fuchs GD et al (2010) Chip-scale nanofabrication of single spins and spin arrays in diamond. Nano Lett 10:3168–3172. https://doi.org/10.1021/nl102066q

Wang W, Roth J, Lindig S, Wu CH (2001) Blister formation of tungsten due to ion bombardment. J Nucl Mater 299:124–131. https://doi.org/10.1016/S0022-3115(01)00679-1

Watt F, Breese MBH, Bettiol AA, van Kan JA (2007) Proton beam writing. Mater Today 10:20–29

Woo CH (2005) Modeling irradiation damage accumulation in crystals. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht, pp 959–986

Xu D, Wirth BD, Li M, Kirk MA (2012) Combining *in situ* transmission electron microscopy irradiation experiments with cluster dynamics modeling to study nanoscale defect agglomeration in structural metals. Acta Mater 60:4286–4302. https://doi.org/10.1016/j.actamat.2012.03.055

Yang Y, Li YG, Short MP et al (2018a) Nano-beam and nano-target effects in ion radiation. Nanoscale 10(4):1598–1606. https://doi.org/10.1039/C7NR08116B

Yang Y, Short MP, Li J (2018b) Monte Carlo simulation of PKA distribution along nanowires under ion radiation. Nucl Eng Des 340:300–307. https://doi.org/10.1016/j.nucengdes.2018.09.001. Accessed 11 Oct 2018

Yao Y, van Mourik MW, Santhana Raman P, van Kan JA (2013) Improved beam spot measurements in the 2nd generation proton beam writing system. Nucl Instrum Methods Phys Res B 306:265–270. https://doi.org/10.1016/j.nimb.2012.10.033

Yip S (2014) Nuclear radiation interactions. World Scientific Publishing, Singapore

Yuan B, Yu PC, Tang SM (1993) A database method for binary atomic scattering angle calculation. Nucl Instrum Methods Phys Res B 83:413–418. https://doi.org/10.1016/0168-583X(93)95864-2

Zhang Y, Bae IT, Sun K et al (2009) Damage profile and ion distribution of slow heavy ions in compounds. J Appl Phys 105:104901. https://doi.org/10.1063/1.3118582

Zhang Y, Ishimaru M, Jagielski J et al (2010) Damage and microstructure evolution in GaN under Au ion irradiation. J Phys D Appl Phys 43:85303. https://doi.org/10.1088/0022-3727/43/8/085303

Zhang P, Wang HY, Li YG et al (2012) Monte Carlo simulation of secondary electron images for real sample structures in scanning electron microscopy. Scanning 34:145–150. https://doi.org/10.1002/sca.20288

Zhao S, Kang W, Xue J et al (2015) Comparison of electronic energy loss in graphene and BN sheet by means of time-dependent density functional theory. J Phys Condens Matter 27:025401. https://doi.org/10.1088/0953-8984/27/2/025401

Ziegler JF (2004) Srim-2003. Nucl Instrum Methods Phys Res B 219–220:1027–1036. https://doi.org/10.1016/j.nimb.2004.01.208

Ziegler JF, Biersack JP, Littmark U (1985) The stopping and range of ions in matter, vol 1. Pergamon Press, New York

Ziegler JF, Biersack JP, Ziegler MD (2008) SRIM: the stopping and range of ions in matter. Cadence Design Systems

Ziegler JF, Ziegler MD, Biersack JP (2010) SRIM - The stopping and range of ions in matter. Nucl Instruments Methods Phys Res Sect B Beam Interact with Mater Atoms 268:1818–1823. https://doi.org/10.1016/j.nimb.2010.02.091

Incorporating Electronic Effects in Molecular Dynamics Simulations of Neutron and Ion-Induced Collision Cascades

98

Andrea E. Sand

Contents

Abstract

Molecular dynamics offers an ideal method for investigating the evolution of collision cascades and the formation of the resulting primary radiation damage. However, several modifications to classical molecular dynamics are needed in order to facilitate simulating the highly non-equilibrium cascade process. One such modification, which is still a point of active research, concerns methods by which electronic effects, which are expected to be important in collision cascades, can be included in molecular dynamics simulations, where a priori atoms are treated as classical particles and electrons are only present implicitly

A. E. Sand (✉)
Department of Physics, University of Helsinki, Helsinki, Finland
e-mail: andrea.sand@helsinki.fi

through the atomic interactions represented by the interatomic potential. This chapter describes current methods by which electronic effects can be accounted for in cascade simulations, with special focus on the developments in the field that have taken place over the last 10 years.

1   Introduction

Irradiation by energetic particles, such as neutrons or ions, can in crystalline materials give rise to various forms of damage in the atomic structure, through collisions with atoms in the target material. Such irradiation events can give rise to what is known as a collision cascade: a series of elastic collisions between atoms in the material, resulting in the permanent displacement of a potentially large number of target atoms. This chapter deals with atomic-scale simulations of such cascades, and current state-of-the-art methods by which electronic effects can be explicitly included in such simulations, beyond the bonding effects which are implicitly captured in the fitting of the semiempirical interatomic potentials.

In cascade events, most of the initial energy of the incident ion, or the recoiling atom, is transferred to the atoms of the target material through the nuclear stopping process, and only a small fraction, typically around 20%, is absorbed by the electronic system. For ions with higher kinetic energy, of the order of MeV, the collision cross section decreases to the point where interaction with atoms of the target becomes negligible, and the energy of the ion is mainly absorbed by the electronic system. Heavy particles in this energy regime are known as swift heavy ions, and while electronic effects are highly relevant in the formation of damage from such particles (Toulemonde et al. 1992; Leino et al. 2015), the damage is not produced through collision cascades, due to the negligible nuclear stopping at higher velocities, and hence this phenomenon lies outside the scope of the current article.

From the first atomistic simulations of cascades in solids in 1960 (Gibson et al. 1960), molecular dynamics (MD) simulations have uncovered many important mechanisms of radiation damage formation, including the different stages of cascade development (Calder et al. 2010), defect recombination during the heat spike phase (Diaz de la Rubia et al. 1987), and athermal defect cluster formation (Diaz de la Rubia and Guinan 1991). As the interatomic potentials that describe the forces between the atoms have improved, the predictive power of MD simulations has increased (Nordlund and Dudarev 2008). As a simulation method, molecular dynamics offers the most complete picture of the processes leading to cascade damage while still being able to treat the full system of up to millions of atoms that is required to capture the whole cascade event.

In many cascade simulations to date, electronic effects have been neglected, either in part or fully. While the potential importance of such effects on the highly non-equilibrium cascade event has been recognized early on (Flynn and Averback 1988; Caro and Victoria 1989), the interactions between the atomic and electronic systems during the cascade evolution are the result of processes occurring over many orders of magnitude in energy, and capturing the behavior in a simple, physically motivated yet computationally expedient model remains a challenging

task. Nevertheless, the past 10 years have seen significant development in this area, due in part to a continued increase in computer capacity. On the one hand, modern supercomputers facilitate performing increasingly large simulations and hence investigating more energetic cascades, where the electronic effects become ever more pronounced. On the other hand, with developments in quantum mechanical techniques of simulating electron dynamics, it is now possible to explicitly follow the dynamics of the interactions between moving ions and electrons. Although full cascades are still out of reach of ab initio calculations, a number of studies of cascade relevant conditions have been performed with different methods (see, e.g., Race et al. 2010), providing insight into the nonadiabatic quantum mechanical effects. Most ab initio studies have focused on predicting electronic energy losses in various projectile-material combinations (Mao et al. 2015; Mason et al. 2007; Race et al. 2012), although the effect of electronic excitations on the interatomic forces has also been investigated (Correa et al. 2012). The initial stages of full cascades have been studied using a time-dependent tight-binding scheme (le Page et al. 2009), providing a method for assessing the accuracy of various approximations of electronic energy losses employed in MD simulations. Together, these developments in computational methods and resources have opened up new avenues for investigating the nonadiabatic effects in collision cascades.

1.1   Phases of Cascade Dynamics

A collision cascade refers to the process of displacing atoms, which is initiated by an energetic projectile undergoing an elastic collision with an atom in the target material, called the primary knock-on atom (PKA). If the kinetic energy acquired by the PKA is above a certain threshold value, termed the threshold displacement energy (TDE), the atom will leave its place in the lattice. When the kinetic energy of the PKA is significantly above the TDE, then as it collides with other atoms in the lattice, it will transfer energy to these atoms, giving rise to secondary, tertiary, etc. recoils. This constitutes the ballistic phase of the cascade, which is characterized by energetic projectiles penetrating undisturbed lattice and undergoing individual energetic collisions with atoms in the matrix as they expand the volume of material affected by the cascade. During the ballistic phase, the volume and structure of the cascade are determined, including possible splitting into subcascades. The energy at which cascades begin to split is material dependent and occurs at lower energies for materials with lower atomic mass. Splitting is a gradual transition, but as examples, in Fe and Ni, cascades begin to split at 20–30 keV, while in W and Au, splitting begins at around 150–200 keV.

In dense materials, from energetic PKAs, the disturbance of the surrounding lattice by the recoils will eventually give rise to a heat spike. This stage of the cascade evolution occurs at the end of the ballistic phase (at around 100–300 fs depending on PKA energy), when the ballistic recoils have lost most of their kinetic energy through collisions with other atoms. The heat spike phase is characterized by a highly disordered, underdense liquid-like volume (or several separate volumes if the cascade has split into subcascades), which persists for several picoseconds.
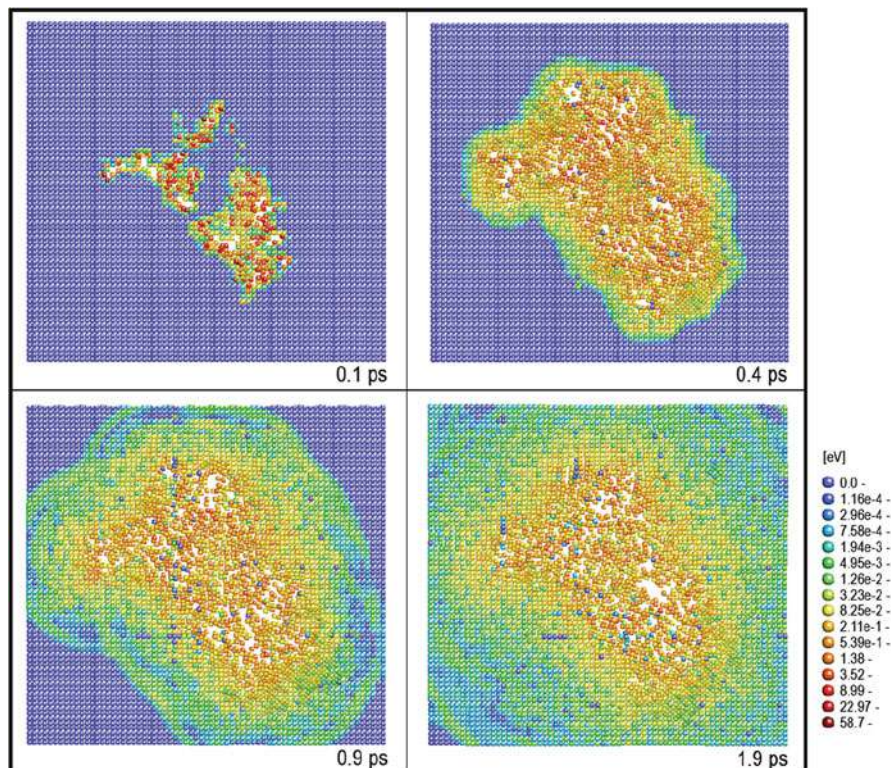
Fig. 1 The disordered core of a 150 keV cascade in W. Atoms shown are from a 2 Å thick slice through the cascade core and are colored according to their kinetic energy (scale in eV)

At this point, energetic projectiles have ceased to exist, and the cascade volume ceases to expand. During the heat spike phase, energetic atoms move randomly in the core of the cascade, resulting in total displacements which are the main causes of ion beam mixing observed in experiments (Kim et al. 1988). Figure 1 shows the disordered core of a cascade in W during the heat spike phase from a 150 keV PKA.

The disordered core eventually cools and recrystallizes, resulting in the recombination of most of the damage in the case of metals and leaving individual point defects and defect clusters as the final products. These defects, which form as a direct result of the athermal processes of the cascade but are subsequently subject to thermal evolution, are termed the primary radiation damage (Was 2007).

1.2 Electronic Effects in Collision Cascades

The ballistic phase of the cascade is dominated by the processes associated with energetic ionic projectiles in matter. It is well established, both theoretically and experimentally, that electrons play a part in the slowing down of such ions (see,

e.g., Ziegler et al. 2010). The electronic contribution to the slowing down of ions is known as electronic stopping. A detailed treatment of the various theories behind the processes involved in electronic stopping is beyond the scope of this chapter. For ion energies relevant in cascade simulations, with velocities smaller than that of the Fermi velocity of electrons, it is given, e.g., by the theory of Lindhard and Sharff (1961) as

$$S_e = \frac{\partial E}{\partial x} = \lambda E^{1/2} \tag{1}$$

where E is the energy of the ion and λ is a material-dependent proportionality factor. For our purposes, we note that the theoretical treatment is based on the assumption of an energetic particle traveling through ground-state pristine material. The SRIM software package (Ziegler 2008) provides plots of the stopping power for any projectile-target combination. The data in these plots are derived from the theoretical models, with various correction terms, and fitted to available experimental data. In many cases, the predictions of the model in SRIM are within a 10% error margin to experiments, but notable exceptions exist. The two scenarios that are most relevant to cascade simulations, where SRIM predictions fail, are in the case of channeled ions and in the low velocity limit.

In the first of the abovementioned cases, predictions fail since SRIM does not account for the crystal structure nor the dependence of the stopping power on the electron density and hence on the ion trajectory. Channeled ions experience a stopping power which may in some cases be significantly lower than that experienced by ions traveling in a random direction or through amorphous material. This has been shown experimentally for a number of cases, including for various projectiles and channels in Si (Wilson et al. 1986) and in W (Eriksson et al. 1967) and in explicit ion range simulations where electron density-dependent stopping was implemented (Cai et al. 1996; Sillanpää 2000). This spatial dependence is not of immediate concern in cascades, where secondary, tertiary, etc. recoils travel in highly random directions in the lattice. Hence the atoms visit varying regions of electron density along their path through the material, and an average stopping power is therefore representative of the average interaction with the electrons. However, as will be explored in later sections of this chapter, the question has bearing on the full treatment of the ion-electron interactions during the complete cascade evolution, including that of the later cascade phases.

In the second case, for the low velocity limit, the model assumes a linear interpolation of the average stopping power down to zero, at zero velocity, which is in good agreement with predictions at slightly higher velocities. However, such behavior is not experimentally verified for very low velocities, and in fact a clear deviation from linear dependency on velocity is seen, both experimentally and in ab initio simulations, in the form of a low-energy threshold for the stopping power of H and He in a number of materials, including insulators (Markin et al. 2009a; Pruneda et al. 2007), but also in metals such as Au and Cu (Markin et al. 2009b; Zeb et al. 2012). The existence of a threshold was predicted early on, based on considerations

of electronic densities of state (Semrad 1986). Surprisingly, recent investigations of the low-energy behavior of the stopping of Si in Si (Lim et al. 2016) have found a non-zero stopping below the band gap threshold, explained by an "electron elevator" effect due to an intermediate electron state within the band gap, induced by the projectile itself.

Another effect of electrons in cascades follows from the electron-phonon coupling, which describes the interactions between thermal ions and electrons. During the initial ballistic phase of the cascade, it is assumed that electronic transport is efficient enough in the pristine lattice to justify instantaneously removing the energy obtained by the electronic system as a result of the electronic stopping of individual recoils. However, during the heat spike phase, the precise mechanisms of energy transfer between the atomic and electronic systems, and the rate of heat dissipation from within the hot, disordered heat spike region, are less obvious. The coupling strength between electronic and ionic systems and the mean free path of electrons in the hot cascade core determine whether or not the electronic system achieves equilibrium with the ionic system during the heat spike phase (Flynn and Averback 1988). However, the energy transfer between electrons and ions in the heat spike does not lend itself well to experimental investigation, and as a result, the potential importance of the coupling on cascade evolution is not known. Early investigations in Cu and Ni (Finnis et al. 1991) indicated that the strong electron-phonon coupling in Ni will quench the cascade and result in fewer defects in total. In contrast, other cascade simulations in Ni, Pd, and Pt, where electronic stopping was included but electron-phonon coupling was disregarded, showed agreement with experiments of ion beam mixing (Nordlund et al. 1998b), a mechanism which is sensitive to the lifetime of the heat spike. These results were interpreted as an indication that electron-phonon coupling effects are negligible in cascades in these materials.

Finally, the contribution of the electrons to thermal diffusion is the dominant method of thermal conduction in metals yet is completely absent in classical molecular dynamics. Hence the heat conduction in the electronic system surrounding the cascade region will contribute to removing heat from the cascade region (Finnis et al. 1991).

2   Electronic Effects in MD Cascade Simulations

2.1   Cascades as a Function of Damage Energy

According to the theories of stopping of ions in matter, an ionic projectile is slowed down in a material through two different mechanisms: the nuclear and the electronic stopping. This allows separating the initial projectile energy into two parts. The energy lost by the ion in collisions with electrons does not directly contribute to the displacement of atoms, and hence only the remaining energy, which is available for elastic collisions with atoms, is taken as the cause of the damage and thus termed the "damage energy," T_{dam}. Especially in early cascade simulations, the energy losses due to electronic stopping were most often disregarded during the simulation but

were instead taken into account in comparisons to real irradiation events. When relating cascade simulation results to actual damage produced by ions or neutrons, the energy of the projectile in the simulation is then labeled $E_{MD} = T_{dam}(E_{PKA})$, with the corresponding (experimental) projectile energy given by E_{PKA} (Stoller et al. 1997). The difference between E_{MD} and E_{PKA}, i.e., the energy lost to electronic stopping, can be calculated, e.g., using the approximation to Lindhard's theory (Lindhard and Sharff 1961) given in (Norgett et al. (1975)). The quantity T_{dam} is also the energy used in the NRT (Norgett-Robinson-Torrens) equation (Norgett et al. 1975) for the displacement damage, $N_{dspl}^{NRT} = 0.8T_{dam}/2E_d$, i.e., the basis for the *dpa* ("displacements per atom") measure of radiation exposure (Nordlund et al. 2015). Here N_{dspl}^{NRT} is the number of defects, or displaced atoms, and E_d is the (material-specific) threshold displacement energy.

To date, many studies employ this method, in part to facilitate comparison with previously published results. While effects on the predictions of the primary damage in lower-energy cascades in many cases may be below the statistical uncertainties, recent results show that in dense, energetic cascades, such as those occurring in the heavier element W, the numbers and sizes of defects in the primary damage are clearly affected by the electronic stopping energy losses (Sand et al. 2016b). Figure 2 illustrates the increasing effect on the fraction of clustered defects that electronic stopping has for increasing PKA energy in W. The effect is strongest on the clustering of vacancies. As cascades of ever-increasing energy are being studied in various materials, given the growing power of massively parallel supercomputers able to simulate ever larger systems, including electronic stopping dynamically becomes increasingly important.

2.2 Electronic Stopping as a Nonlocal Friction Force

A straightforward method of including the electronic stopping dynamically during cascade simulations is to apply a friction force on the moving atoms. From the electronic stopping $S_e = \frac{dE}{dx} \sim v$, one arrives at an additional term in the equations of motion:

$$m = \frac{d^2\mathbf{x}}{dt^2} = \mathbf{F} - \beta\mathbf{v} \tag{2}$$

Applying such a friction term in molecular dynamics simulations is computationally efficient, and the method has been used for numerous studies of cascade damage (see, e.g., refs Zhong et al. (1998), Nordlund et al. (1998a, b), and later work by these authors).

There are a number of uncertainties pertaining to the use of a nonlocal, average stopping power (such as that given by SRIM) in classical MD cascade simulations. As mentioned earlier, the local stopping power depends on the local electronic density. For most recoils in cascades, the trajectories occur in random directions, and the average electronic stopping predicted by SRIM is largely applicable. This, however, does not hold for channeled ions, for which the nonlocal friction force may be an overestimate.
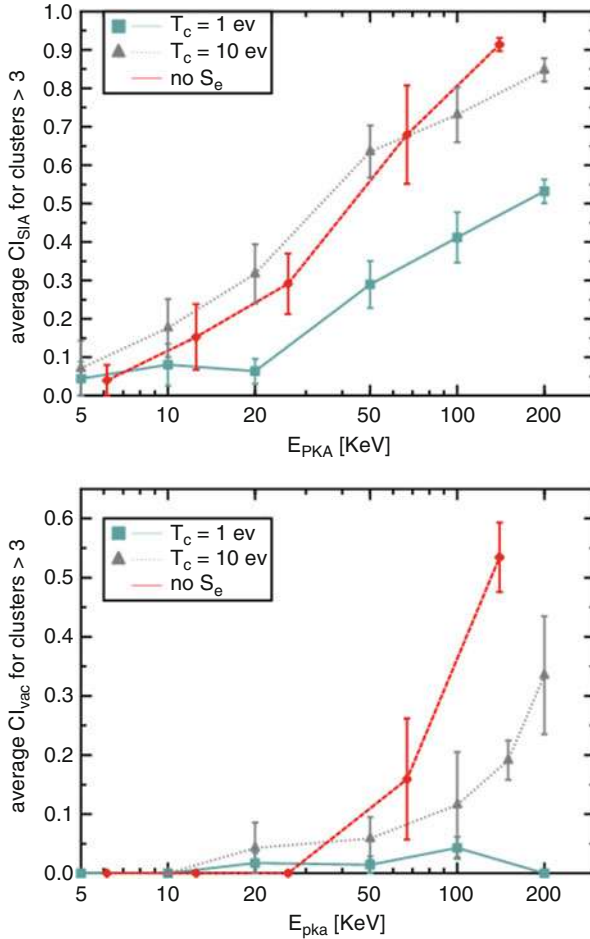
Fig. 2 The fraction of defects in clusters predicted by simulations without electronic stopping S_e and with electronic stopping included for different cutoff energies T_c (see Sect. 2.2 for details). In simulations with no S_e, the value of E_{PKA} is calculated from the initial recoil energy E_{MD} in the simulation (which in this case represents the damage energy) using an approximation to Lindhard's theory (Lindhard and Sharff 1961). In simulations with electronic stopping, E_{PKA} is the initial recoil energy in the simulation

A more serious shortcoming of the average friction term in cascade simulations is the practice of setting a cutoff T_c for the kinetic energy, below which the frictional force is not applied. This is necessary in order to avoid quenching all thermal modes, since the energy lost to electrons is simply extracted from the system. Since such a threshold in most metals is not directly motivated by any physical theory, various different ways of choosing a value of T_c can be found in the literature, including taking twice the value of the material's cohesive energy (Duffy and Rutherford 2007) or a value of 10 eV independently of the material (Zhong et al. 1998;

Nordlund et al. 1998a, b). Recent investigations of the effects of different values for the threshold in W (Sand et al. 2014) and Ni, Pd, and Pt (Sand and Nordlund 2015) have shown an effect of the choice of T_c on both the damage production and atomic mixing resulting from cascades.

Although the energy lost to collisions with electrons from an ion with low velocity becomes increasingly tiny, especially in relation to the initial PKA energy, the total number of atoms in cascades with energies of the order of the threshold value can reach into the thousands during the heat spike phase. The total energy losses therefore add up, resulting in a very large difference in the final amount of energy extracted from the atomic system, depending on the cutoff value. As an example, in tungsten it was found that a cutoff value of $T_c = 1$ eV results in roughly 40 keV of excess energy lost from the atomic system during the heat spike phase of the cascade, compared to negligible energy losses during the heat spike phase if the cutoff is set to $T_c = 10$ eV. This is illustrated in Fig. 3. The additional energy losses roughly equal the amount of energy lost to electronic stopping during the ballistic phase of the cascade, giving a total of 80 keV energy lost to electrons from 150 keV PKAs, for $T_c = 1$ eV. This does not agree with the SRIM prediction of the proportions of electronic and nuclear stopping of ions at that energy and is in contradiction also with the predictions of Lindhard's theory. The difference in the resulting primary radiation damage from such simulations is drastic, with no large clusters forming when $T_c = 1$ eV. In contrast, regularly occurring large clusters are predicted in simulations with $T_c = 10$ eV.

The impact of the electronic energy losses on predictions of the primary damage is dependent on material, with a more pronounced effect in materials with larger atomic mass. Figure 4 shows a comparison of the effects in the bcc materials W and Fe. Including electronic stopping during the simulation affects the slope of the
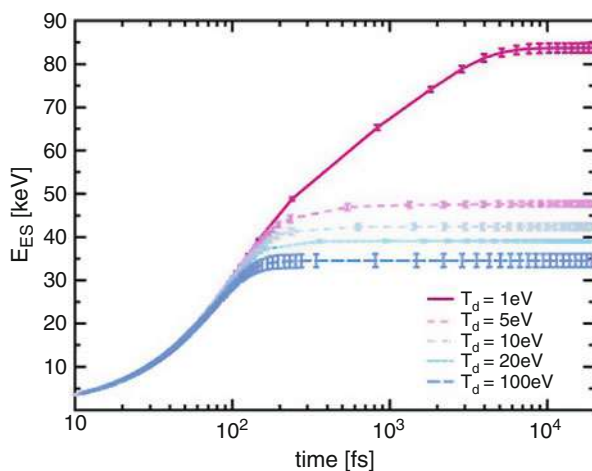
Fig. 3 Cumulative energy losses to electronic stopping during the evolution of 150 keV cascades in W, for different cutoff values. (Reproduced with permission from Sand et al. 2014)
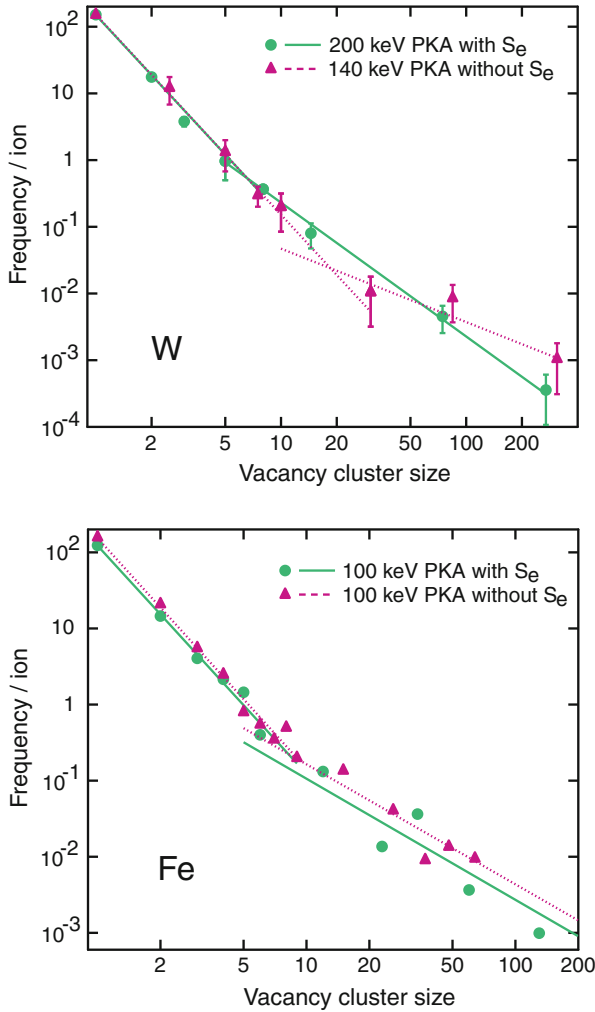
Fig. 4 Size-frequency distributions of vacancy clusters in Fe and W, with and without electronic stopping. For W, PKAs with the same damage energy are shown, while for Fe the difference in overall frequency is due to the difference in damage energy for the two cases that are plotted. (Reproduced with permission from Sand et al. 2016b)

distribution of defect cluster sizes in W quite strongly, while in Fe the slope of the distribution is largely unaffected by whether or not electronic stopping is included. This holds when the threshold T_c is set sufficiently high to leave the heat spike unaffected (for Fe the value $T_c = 5$ eV was used). This material dependence can be related to the increasing impact of electronic stopping on cascades of increasing energy, through the phenomenon of subcascade splitting. For materials with smaller atomic mass, cascades split at a relatively low energy; hence compact energetic cascade regions occur seldom in such materials, even for higher-energy PKAs.

It should be noted that because of the threshold effect, the damage energy T_{dam} associated with a given projectile energy E_{PKA} is not a well-defined quantity, when electronic stopping is included explicitly through a friction term as the sole electronic effect in cascade simulations. While the energy lost to the electronic friction during the simulation can be easily traced, the final value depends sensitively on the chosen cutoff T_c. The plots in Fig. 5 show the dependence of the "damage energy" on the cutoff value for a range of materials, with a comparison to the predictions of Lindhard's theory. While Lindhard's prediction lies within the range of these simulations, it becomes clear that relating a damage energy to a cascade simulation is rather ad hoc, and the electronic effects in the heat spike, which both Lindhard's theory and the electronic friction method fail to properly capture, potentially affect a rather large proportion of the PKA energy in the higher cascade energy regime.

Fig. 5 Energy losses to electronic stopping as a function of cascade energy in Ni, Pd, and Pt, compared to the damage energy calculated from Lindhard's theory. (Reproduced with permission from Sand and Nordlund 2015)
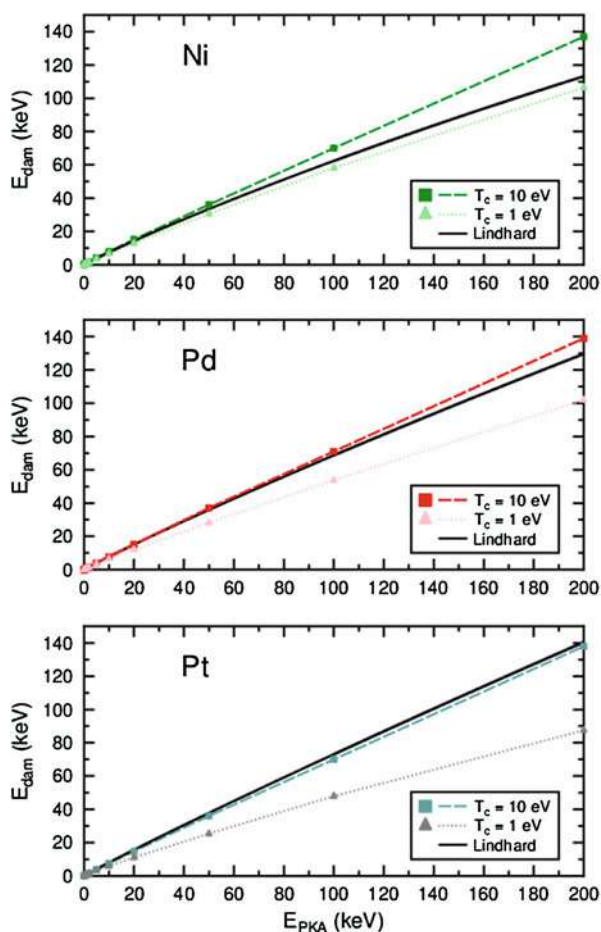
2.3 Electron-Phonon Coupling and Two-Temperature Molecular Dynamics

The issues with tuning a threshold T_c for electronic stopping are due to the fact that electronic stopping theories no longer apply once the cascade transitions into the heat spike phase. At that stage, the dynamics of the atomic system become more phonon-like, and energy transfer between ions and electrons is more correctly described by electron-phonon coupling.

A simple model which introduced electron-phonon coupling based on the method by Hou et al. (2000), as an additional damping term which tends to zero as the ionic temperature approaches a set temperature for the electronic system, was investigated for cascade simulations in Fe by Björkas and Nordlund (2009). Results showed an effect on the numbers of defects but not on the fraction of defects in clusters. However, using a time constant for the electronic stopping derived from the Sommerfeld free electron theory resulted in an underestimation of the atomic mixing with this method.

A method that explicitly includes the electronic system in molecular dynamics simulations, and accounts for both the electronic stopping and electron-phonon coupling, was pioneered for cascade simulations by Duffy and Rutherford (2007) and Rutherford and Duffy (2007). The method uses a stochastic force to represent the energy transfer from the electronic system back to the ionic system, an idea which was first suggested by Caro and Victoria (1989). The Langevin equations of motion for the ionic system are then given by

$$m\frac{\partial \mathbf{v}_i}{\partial t} = \mathbf{F}_i(t) - \gamma_i \mathbf{v}_i + \mu(t), \tag{3}$$

where $\mathbf{v}_i$ is the velocity of the ion; μ is a stochastic force, the strength of which is determined by the strength of the electron-phonon coupling; and $\gamma_i \mathbf{v}_i$ is a friction term associated with the electronic stopping. The formalism builds on the two-temperature model (Lifshits et al. 1960) (TTM), where the atomic and electronic subsystems are assumed to be separately in internal equilibrium, with different defined temperatures, which evolve toward the same value through the coupling between the subsystems. The electronic system in the MD simulation is represented by a continuum thermal bath at temperature T_e, coupled to the ionic system with temperature T_i, and governed by the heat diffusion equation:

$$c_e \frac{\partial T_e}{\partial t} = \kappa_e \Delta T_e - g_p (T_e - T_i) + g_s T_p \tag{4}$$

where c_e and κ_e are the specific heat and heat conductivity, respectively, of the electrons. Note that the corresponding lattice thermal diffusion is accounted for explicitly in the atomic MD system. The parameters g_p and g_s are related to the coefficient γ_i in the friction term in Eq. 3 and determine the strength of the coupling for the electron-phonon interaction and electronic stopping, respectively.

The second term on the right-hand side of Eq. 4 describes the heat exchange between electrons and lattice, while the third term on the right balances the energy lost from the ions through the electronic stopping friction term. The parameter T_p is determined by energy balance equations.

Although the Duffy and Rutherford model includes a description of both electronic stopping and electron-phonon coupling and allows exchange of energy both ways between the electronic and ionic subsystems, it still suffers from an unphysical cutoff parameter for electronic stopping. The model treats the two energy loss mechanisms separately, with electronic stopping applied only above an arbitrarily chosen velocity v_0. Hence the coupling term γ_i in Eq. 3 is given by

$$\gamma_i = \gamma_p + \gamma_s \quad \text{for} \ \ v_i > v_0 \tag{5}$$

$$\gamma_i = \gamma_p \quad \text{for} \ \ v_i \leq v_0 \tag{6}$$

where γ_p is the friction coefficient for the electron-phonon coupling and γ_s is the coefficient for the electronic stopping friction term. An additional uncertainty pertaining to the model is the need for a well-defined lattice temperature. In the initial stage of the cascade, a local temperature is not well-defined, leading to the choice by authors to "turn on" the electron-phonon coupling after 300 fs of the initial cascade evolution (Zarkadoula et al. 2014).

The TTM-MD method has been implemented in in-house versions of DL_POLY but is not yet available in the general DL_POLY distribution (Todorov et al. 2006). It has been used in a number of investigations of cascade damage in metals, including both Fe and W, and in some Ni- and Fe-based alloy systems (see, e.g., Zarkadoula et al. (2017) and references therein). Early investigations with the model in Fe (Rutherford and Duffy 2007) showed a sensitivity of results to the strength of the coupling parameter, with an initial increase in the number of defects with increasing coupling strength, due to a more rapid cooling rate at the later phase of the cascade evolution, resulting in less recombination at later stages (see Fig. 6). As the coupling strength was increased further, the number of residual defects decreased to below the numbers predicted in simulations without electronic effects. This was explained by the rapid early quenching of the cascade, which inhibited the initial formation of the heat spike. Later studies of energetic cascades in Fe (Zarkadoula et al. 2014) used a coupling constant derived from DFT and found little effect on the average numbers of residual defects but instead saw a discernible effect on the degree of clustering of the defects. In energetic cascades in W (Zarkadoula et al. 2015), on the other hand, the TTM-MD method predicts fewer final defects when electron-phonon coupling is included, as compared to simulations where only electronic stopping is included as a friction term.

A slightly modified version of the TTM-MD, which corrects for an energy drift in Duffy's and Rutherford's model (Phillips and Crozier 2009), has been implemented in the LAMMPS code (LAMMPS 1995; Plimpton 1995) and is freely available for use. However, this implementation enforces periodic boundary conditions in
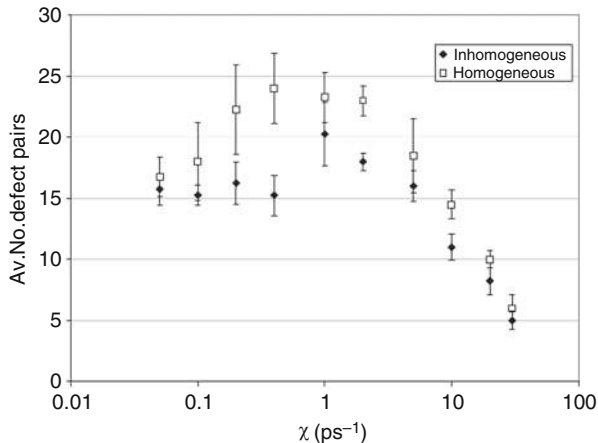
Fig. 6 Defect numbers predicted in cascades in Fe using TTM-MD with a range of coupling strengths. (Reproduced with permission from Rutherford and Duffy 2007)

all dimensions and hence removes the spatially extended borders of the electronic system, which the original TTM-MD used in conjunction with a heat sink at the electronic system borders, to facilitate electronic thermal diffusion away from the hot cascade region. A thorough investigation of the effects of boundary conditions on the cascade evolution and damage formation has not been carried out. For a perspective on the energy increase associated with cascades, one can consider a 200 keV cascade, which would require a 20 million atom system simulated in the NVE ensemble in order to maintain the average temperature to within 100 K of the initial ambient temperature. While this is well within the capacity of today's supercomputers and in pure MD simulations, such a small temperature difference would not result in significant differences in the outcome of cascades, indications are that damage production is more sensitive to temperature effects when electronic interactions are included (Zarkadoula et al. 2015). Due to this, the applicability of a periodic electronic system for energetic cascade simulations may be somewhat restricted. The method has nevertheless been used for lower-energy cascade simulations in α-quartz (Phillips et al. 2010), silicon carbide (Zhang et al. 2013), and in a pyrochlore system (Ismail et al. 2010).

These TTM-MD studies rely on a number of assumptions concerning the electronic effects, including, among others, a constant value of the coupling strength, an electronic thermal conductivity which is independent of the atomic disorder in the heat spike (and in many studies also independent of temperature), and the various thresholds such as the cutoff for electronic stopping and the time at which to include the electron-phonon coupling. The time of turning on the electron-phonon coupling was shown to affect the numbers of residual defects (Zarkadoula et al. 2017). The investigations to date have mainly been concerned with the differences in predictions of the various models and sensitivity of the TTM-MD

to the coupling constant. As no quantitative comparisons of these simulation results with experiments have been reported, the conclusions to date mainly amount to an indication that electron-phonon coupling likely has an effect on the final damage from cascades.

Mason et al. (2014) recently developed an alternative approach to including electron-ion energy transfer in both the electronic stopping and electron-phonon coupling regimes in cascade simulations, which avoids the necessity of a nonphysical cutoff. Instead, a transition between electron-phonon coupling and electronic stopping is determined based on band structure. In this model, the transition regime is found for atoms with kinetic energies on the order of hundreds of eV. Mason's model also accounts for the effects of disorder on the electronic thermal conduction. While this model has not yet been used extensively for cascade simulations, the results presented in Mason et al. (2014) demonstrate a similarity with the method of including only an electronic friction term above a cutoff set to $T_c = 10$ eV, when looking at electronic energy losses over time as the cascade develops. The observed behavior, where energy losses level off at the end of the ballistic phase (see Fig. 7, showing energy losses in 25 keV cascades in W, where the ballistic phase ends roughly after the first 100 fs), can in Mason's model be traced to the significant decrease in electronic conduction due to the disorder in the cascade core, effectively trapping the energy in the heat spike. This leads to the same qualitative behavior as that achieved artificially by a cutoff T_c of the electronic stopping, when T_c is set to a value such that energy losses cease during the transition between the ballistic and heat spike phases. The longer-term energy loss from the atomic system in the plots with a kinetic energy cutoff k_c in Fig. 7 is due to heat dissipation through the borders of the cell (by the use of a thermostat).
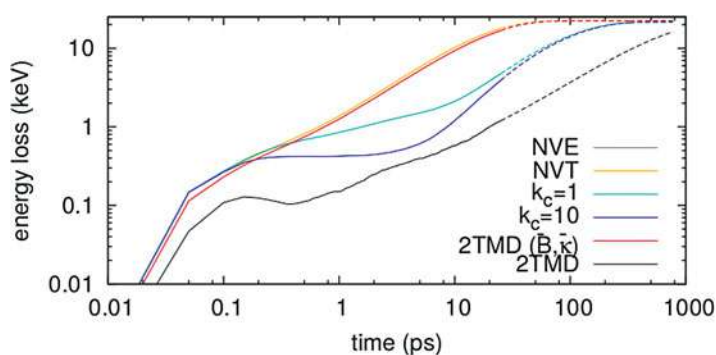
Fig. 7 Energy losses from the atomic system during 25 keV cascades in W using various methods of energy transfer. 2TMD refers to Mason's method, 2TMD $(\overline{B}, \overline{\kappa})$ refers to Duffy's model with fixed electronic thermal conductivity and electron-phonon coupling constant, and $k_c = 1$ and $k_c = 10$ refer to a model with only a friction term and cutoff at 1 and 10 eV, respectively. (Reproduced with permission from Mason et al. 2014)

3   Quantitative Comparisons with Experiment

While it is currently impossible to experimentally observe the collision cascade itself, there are a number of ways in which the effects of cascades can be measured experimentally, and quantitative comparisons with these can be used to guide simulations and give an indication of the validity of various simulation methodologies. Due to the large variability between individual energetic cascades and the wide spectrum of energies of the PKAs in, e.g., irradiation experiments with energetic noble gas atoms, a large set of cascade simulations is needed for any comparison to be possible.

Although MD simulations incorporating electronic effects can be compared to a wide variety of experiments, such as swift heavy ion tracks, ion projectile ranges, short-range and long-range disorder (applicable to alloys), and laser heating, we present here two examples that specifically address collision cascade phenomenon and are applicable also to pure materials, namely, the atomic mixing in heat spikes (captured experimentally by the use of a separate marker species) and the residual defects, or primary damage, of the cascade.

3.1   Ion Beam Mixing

As a result of cascades, atoms are displaced from their original positions, coming to rest some distance away from where they originated. In metals, displaced atoms often come to rest on a lattice site and hence do not contribute to the final damage but have nevertheless traveled a certain distance. This atomic mixing can be divided into two parts – the ballistic mixing and the heat spike mixing. The physics underlying these two mechanisms are different, and while they cannot be separated experimentally, theoretical predictions of the magnitude of ballistic mixing indicate that the total mixing in ion-irradiated metals is mainly due to heat spike mixing (Kim et al. 1988), which occurs as a diffusive process in the disordered, liquid-like core of energetic cascades, before the core recrystallizes. Simulations support this conclusion (Nordlund et al. 2015). Since the level of atomic mixing thus depends on the duration of the heat spike, a suitable choice for the cutoff energy in the electronic friction model can be found empirically from cascade simulations, by comparing with ion beam mixing (IBM) experiments.

The total atomic displacements R_{sim} can be extracted from cascade simulations as long as atoms are indexed and thus traceable. The square of the displacements $R^2{}_{sim}$ can be fitted to a function Nordlund et al. (1998b) of the form

$$R^2(E) = a \frac{E^{c+1}}{b^c + E^c},$$
(7)

where a, b, and c are fitting parameters and E is the PKA energy. The mixing efficiency derived from simulation is then obtained by integrating R^2_{sim} over the whole PKA energy range, weighted by the energy spectrum of the recoils from the

incident ion (Nordlund et al. 1998b):

$$Q_{sim} = \frac{\int_0^{E_0} R^2(E)n(E)dE}{6n_0 E_{D_n}}. \tag{8}$$

Here E_0 is the initial ion implantation energy, n_0 is the atomic density, and E_{D_n} is the damage energy of the ion beam.

As seen in Fig. 3, the choice of the cutoff T_c affects the amount of energy losses during the heat spike phase. This, in turn, affects the lifetime of the liquid core and hence the amount of mixing that takes place. The effect of the threshold is strongest in the more energetic cascades, and these are also the cascades that dominate the total mixing from the range of recoil energies produced by the impacting noble gas ions in IBM experiments. Hence the predicted mixing efficiency from cascade simulations is sensitive to the cutoff energy used for the electronic stopping.

When comparing MD cascade simulations in a pure material with experimental ion beam mixing results, the effects of impurities, varying heats of mixing and miscibility of the marker layer and substrate, and other chemical effects in the real material combination must be taken into account (Paine and Averback 1985). High-quality IBM experiments which minimize chemical effects have been performed in the metals Ni, Pd, and Pt (Nordlund et al. 1998b), and comparison of the mixing efficiencies for these materials with the predictions of cascade simulations shows that a cutoff at $T_c = 1$ eV underestimates the mixing in all cases, while a value of 10 eV gives better agreement, although in Pd and Pt, it is then slightly overestimated (Sand and Nordlund 2015).

These results, combined with the observation that raising the cutoff energy slightly above 10 eV has little or no effect on the final damage nor on the overall energy losses (Sand et al. 2013, 2014), while a lower value of 1 eV has a clear effect on both, and considering the fact that electronic stopping theory does not directly apply in the heat spike regime, conclude that erring on the side of a higher cutoff is perhaps to be preferred over introducing excessive energy losses through an electronic stopping friction term active during the heat spike (Table 1).

Table 1 Simulated and experimental values for the mixing efficiency Q.[a] Nordlund et al. (1998b),[b] Kim et al. (1988),[c] Fenn-Tye and Marwick (1986). (Reproduced with permission from Sand and Nordlund 2015)

Material beam		Q_{sim} (Å^5/eV)$T_c = 1$ eV	Q_{sim} (Å^5/eV)$T_c = 10$ eV	Q_{exp} (Å^5/eV)
Ni	600 keV Kr	2.9 ± 0.1	4.7 ± 0.1	4.8 ± 0.5[a]
Ni	650 keV Kr	3.1 ± 0.2	5.1 ± 0.1	5.0 ± 0.7[b]
Pd	600 keV Kr	6.2 ± 0.2	14 ± 1	8.4 ± 0.8[a]
Pd	400 keV Kr	6.1 ± 0.2	13.7 ± 0.5	9 ± 1[c]
Pt	1 MeV Kr	6.1 ± 0.3	20 ± 3	14 ± 2[b]

3.2   Primary Damage in TEM

It has been noted in several studies, both with the TTM-MD method and with purely electronic stopping, that the magnitude of electronic energy losses from ions affects the clustering of defects in the final damage (Zarkadoula et al. 2014; Sand and Nordlund 2015). Put in different terms, one finds that the distribution of the sizes of defect clusters in the primary damage is affected. A high fraction of clustered defects is seen in terms of the size-frequency distribution as the occasional occurrence of large clusters, with the distribution closely following a power law (Sand et al. 2013). On the other hand, a low fraction of clustered defects corresponds to a completely different distribution. Figure 8 shows the size-frequency distributions for SIA clusters in 200 keV cascades in W obtained with electronic stopping using different values of T_c. While the distributions predicted both without electronic stopping and with electronic stopping and a threshold of $T_c = 10$ eV fit well to a power law, the lower $T_c = 1$ eV inhibits all formation of larger clusters, and the size-frequency distribution tails off exponentially.

In materials where the primary damage contains defects large enough to be visible in an electron microscope, transmission electron microscopy (TEM) can be used for direct comparison between MD predictions and irradiation experiments. Such a comparison has been carried out for self-ion-irradiated tungsten (Yi et al. 2015), where irradiations were performed at cryogenic temperatures and defects observed *in situ*, thus minimizing the microstructural evolution which could take
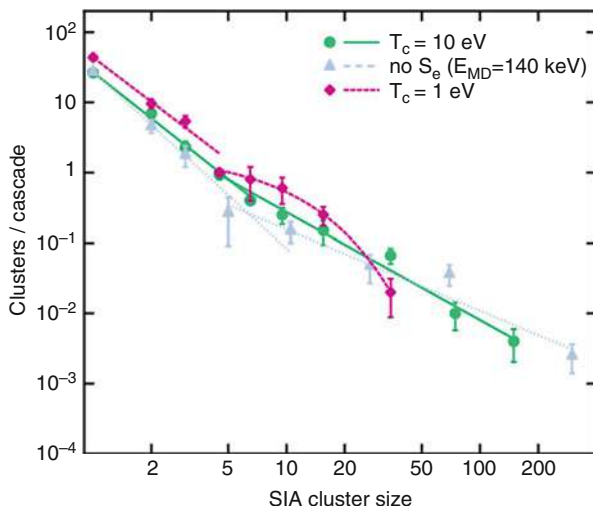
Fig. 8 Size-frequency distributions of SIA clusters, as predicted by MD simulations of 200 keV cascades in W, simulated with different cutoff values for the electronic stopping friction term but otherwise identical conditions. The defect distribution in the case of $T_c = 1$ eV is fitted to an exponential function, while the distributions predicted with the other methods fit well to power laws

place during the irradiation and between the time of irradiation and the time of observation. In particular, it was observed during the course of the measurements that loops did not visibly grow but appeared instantaneously from one frame to the next. (This is in contrast to higher temperature *in situ* irradiation experiments, where gradual growth of loops is apparent (Yi et al. 2013).) Hence the large clusters visible in the micrographs were concluded to be the products of athermal defect formation in individual cascades. The size-frequency distribution of the TEM-visible defects was extracted from the micrographs using an automated method (Mason et al. 2018) and shown to be in good quantitative agreement with MD predictions from cascades where electronic stopping was included as a friction force with $T_c = 10$ eV (Sand et al. 2017). Comparing Fig. 9 to the predictions plotted in Fig. 8, it is clear that a value of $T_c = 1$ eV leads to a direct contradiction with the experimental observation of TEM-visible loops in W.
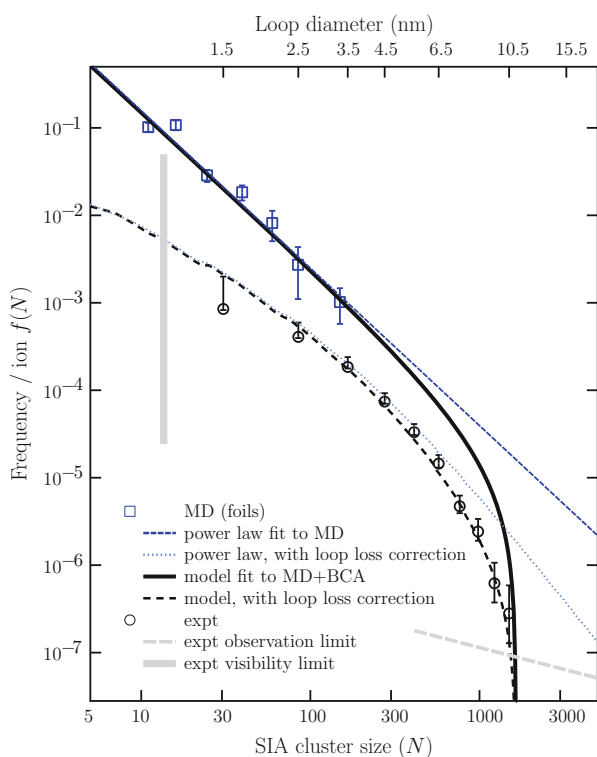
Fig. 9 Defect cluster size-frequency distribution as predicted by MD, compared to TEM observation. The experimental defect count is shown as black circles, while the blue squares were obtained with MD. The thick black line is a fit of the MD data to a full model for the size distribution of defects, where subcascade splitting is also taken into account (see Ref. Sand et al. (2017) for details). Accounting for the expected (and observed) loss of loops to the surface yielded the dotted black line, with no fitting to experiment. (Reproduced with permission from Sand et al. 2017)

4   Outlook

Historically, perhaps two factors lie most strongly behind the tendency to widely disregard electronic effects in MD cascade simulations. On the one hand, these effects are rather poorly known for the complicated processes at play in energetic collision cascades. But in addition, most early MD cascade studies were restricted to relatively low PKA energies, where the complicated processes of energy transfer at lower velocities and in the disordered cascade core are largely negligible in relation to the overall energy losses during the cascade. Today, with the continued increase in computer power, increasingly energetic cascades can be simulated with MD, with enough statistics to capture the fluctuations in the damage production. This allows investigating the sensitivity of predictions to effects of electronic energy losses in energetic heat spikes and, most importantly, of validating proposed models by comparison of the statistics of the primary damage with direct experimental observations. In addition, the relatively new time-dependent density functional theory (TDDFT) method provides novel ways of investigating the ion-electron energy transfer under various conditions and opens the way for probing unexplored regions of cascade dynamics, offering a way toward a deeper understanding of electron excitation processes in highly disordered systems and facilitating the direct parameterization of new energy transfer models.

Recently, the power of TDDFT has been put to use toward realizing the ideas originally put forward almost 30 years ago by Caro and Victoria (1989) of expressing the high-energy electronic stopping and low-energy electron-phonon coupling within one formalism in MD cascade simulations. The motivation behind this approach is the understanding that the two limiting models of energy transfer – in the high-energy and low-energy limits – are essentially due to the same physics (Koponen 1993). The proposed method has the advantage of eliminating the need for an arbitrary threshold imposed on the electronic stopping but introduces the need to parameterize the energy transfer for the whole range of electron densities visited by the recoil. Caro and coworkers (Caro et al. 2015) demonstrate the feasibility of doing this within TDDFT, with the help of modern computer capacity. They show the effect of electron density on the stopping of a moving ion and argue that low-energy thermal ions will effectively be embedded in an electron density more like that of a vacant lattice site, rather than that experienced by an interstitial atom or energetic recoil. This approach yields levels of energy transfer from ions to electrons in the two limiting regimes of an energetic recoil and a thermal ion, which correspond well to the magnitudes of the friction terms normally used for the electronic stopping and electron-phonon coupling, respectively. Figure 10 shows the dependence of the energy of the electronic system on the atom position during a TDDFT simulation of an Ni atom traversing a vacant site in the Ni lattice, representing the energy losses experienced by a thermal ion. Based on these calculations, they present an electron density-dependent stopping $dE/dx = \beta (\rho)v$ that can be conveniently introduced into MD simulations and treated with the same computational method as that used for calculating EAM-type interatomic potentials. At the time of writing, this approach has yet to

Fig. 10 The energy of the electronic system during a TDDFT simulation of a Ni atom traversing an empty site in the Ni fcc lattice. The dashed blue line shows the difference in energy between the TDDFT electronic states and the corresponding Born-Oppenheimer energy. The slope of the curve is the instantaneous rate of energy loss of the ion. (Reproduced with permission from Caro et al. 2015)
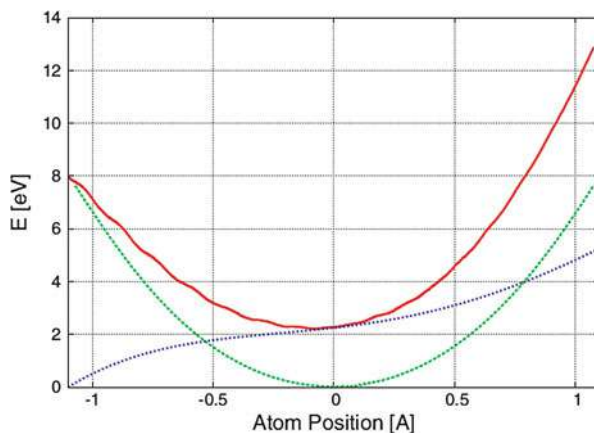
be implemented in cascade simulations but offers an intriguing possibility for overcoming many of the weaknesses of current models.

Ultimately, the question of validating ion-electron energy transfer models in MD cascade simulations will overlap with the question of the accuracy of interatomic potentials in describing the cascade dynamics (Sand et al. 2016a). For this reason, validation of models such as the TTM-MD will necessarily need to be supported by comparisons with experiments also of other phenomenon, where predictions are less sensitive to the choice of potential.

References

Björkas C, Nordlund K (2009) Assessment of the relation between ion beam mixing, electron-phonon coupling and damage production in Fe. Nucl Instr Meth B 267(10):1830–1836. https://doi.org/10.1016/j.nimb.2009.03.080

Cai D, Gronbech-Jensen N, Snell CM, Beardmore KM (1996) Phenomenological electronic stopping-power model for molecular dynamics and Monte Carlo simulation of ion implantation into silicon. Phys Rev B 54:17147–17157. https://doi.org/10.1103/PhysRevB.54.17147

Calder A, Bacon D, Barashev A, Osetsky Y (2010) On the origin of large interstitial clusters in displacement cascades. Philos Mag 90(7–8):863–884. https://doi.org/10.1080/14786430903117141

Caro A, Victoria M (1989) Ion-electron interaction in molecular-dynamics cascades. Phys Rev A 40:2287–2291. https://doi.org/10.1103/PhysRevA.40.2287

Caro A, Correa AA, Tamm A, Samolyuk GD, Stocks GM (2015) Adequacy of damped dynamics to represent the electron-phonon interaction in solids. Phys Rev B 92(144):309. https://doi.org/10.1103/PhysRevB.92.144309

Correa AA, Kohanoff J, Artacho E, Sánchez-Portal D, Caro A (2012) Nonadiabatic forces in ion-solid interactions: the initial stages of radiation damage. Phys Rev Lett 108:213201. https://doi.org/10.1103/PhysRevLett.108.213201

Diaz de la Rubia T, Averback RS, Benedek R, King WE (1987) Role of thermal spikes in energetic displacement cascades. Phys Rev Lett 59:1930–1933. https://doi.org/10.1103/PhysRevLett.59.1930

Diaz de la Rubia T, Guinan MW (1991) New mechanism of defect production in metals: a molecular-dynamics study of interstitial-dislocation-loop formation in high-energy displacement cascades. Phys Rev Lett 66:2766–2769. https://doi.org/10.1103/PhysRevLett.66.2766

Duffy DM, Rutherford AM (2007) Including the effects of electronic stopping and electronion interactions in radiation damage simulations. J Phys Condens Matter 19(1):016207. http://stacks.iop.org/0953-8984/19/i=1/a=016207

Eriksson L, Davies JA, Jespersgaard P (1967) Range measurements in oriented tungsten single crystals (0.1–1.0 Mev). I. electronic and nuclear stopping powers. Phys Rev 161:219–234. https://doi.org/10.1103/PhysRev.161.219

Fenn-Tye I, Marwick A (1986) The dependence of cascade mixing in Pd on the projectile's mass. Nucl Inst Methods Phys Res B 18(16):236–242. https://doi.org/10.1016/S0168-583X(86)80037-4

Finnis MW, Agnew P, Foreman AJE (1991) Thermal excitation of electrons in energetic displacement cascades. Phys Rev B 44:567–574. https://doi.org/10.1103/PhysRevB.44.567

Flynn CP, Averback RS (1988) Electron-phonon interactions in energetic displacement cascades. Phys Rev B 38:7118–7120. https://doi.org/10.1103/PhysRevB.38.7118

Gibson JB, Goland AN, Milgram M, Vineyard GH (1960) Dynamics of radiation damage. Phys Rev 120:1229–1253. https://doi.org/10.1103/PhysRev.120.1229

Hou Q, Hou M, Bardotti L, Prével B, Mélinon P, Perez A (2000) Deposition of Au_N clusters on Au(111) surfaces. I. atomic-scale modeling. Phys Rev B 62:2825–2834. https://doi.org/10.1103/PhysRevB.62.2825

Ismail AE, Greathouse JA, Crozier PS, Foiles SM (2010) Electron–ion coupling effects on simulations of radiation damage in pyrochlore waste forms. J Phys Condens Matter 22(22):225405. http://stacks.iop.org/0953-8984/22/i=22/a=225405

Kim SJ, Nicolet MA, Averback RS, Peak D (1988) Low-temperature ion-beam mixing in metals. Phys Rev B 37:38–49. https://doi.org/10.1103/PhysRevB.37.38

Koponen I (1993) Energy transfer between electrons and ions in dense displacement cascades. Phys Rev B 47:14011–14019. https://doi.org/10.1103/PhysRevB.47.14011

LAMMPS (1995) Lammps classical molecular dynamics computer code. http://lammps.sandia.gov

le Page J, Mason DR, Race CP, Foulkes WMC (2009) How good is damped molecular dynamics as a method to simulate radiation damage in metals? New Journal of Physics 11(1):013004. http://stacks.iop.org/1367-2630/11/i=1/a=013004

Leino AA, Daraszewicz SL, Pakarinen OH, Nordlund K, Djurabekova F (2015) Atomistic two-temperature modelling of ion track formation in silicon dioxide. Europhys Lett 110(1):16004. http://stacks.iop.org/0295-5075/110/i=1/a=16004

Lifshits I, Kaganov M, Tanatarov L (1960) On the theory of radiation-induced changes in metals. J Nucl Energy Part A 12(12):69–78. https://doi.org/10.1016/0368-3265(60)90010-4

Lim A, Foulkes WMC, Horsfield AP, Mason DR, Schleife A, Draeger EW, Correa AA (2016) Electron elevator: excitations across the band gap via a dynamical gap state. Phys Rev Lett 116:043201. https://doi.org/10.1103/PhysRevLett.116.043201

Lindhard J, Sharff M (1961) Energy dissipation by ions in the kev region. Phys Rev 124:128

Mao F, Zhang C, Zhang FS (2015) Theoretical study of the channeling effect in the electronic stopping power of silicon carbide nanocrystal for low-energy protons and helium ions. Nucl Inst Methods Phys Res B 342(Suppl C):215–220. https://doi.org/10.1016/j.nimb.2014.09.035

Markin SN, Primetzhofer D, Bauer P (2009a) Vanishing electronic energy loss of very slow light ions in insulators with large band gaps. Phys Rev Lett 103:113201. https://doi.org/10.1103/PhysRevLett.103.113201

Markin SN, Primetzhofer D, Spitz M, Bauer P (2009b) Electronic stopping of low-energy h and he in cu and au investigated by time-of-flight low-energy ion scattering. Phys Rev B 80(205):105. https://doi.org/10.1103/PhysRevB.80.205105

Mason DR, le Page J, Race CP, Foulkes WMC, Finnis MW, Sutton AP (2007) Electronic damping of atomic dynamics in irradiation damage of metals. J Phys Condens Matter 19(43):436209. http://stacks.iop.org/0953-8984/19/i=43/a=436209

Mason DR, Yi X, Kirk MA, Dudarev SL (2014) Elastic trapping of dislocation loops in cascades in ion-irradiated tungsten foils. J Phys Condens Matter 26(37):375701. http://stacks.iop.org/0953-8984/26/i=37/a=375701

Mason D, Sand A, Yi X, Dudarev S (2018) Direct observation of the spatial distribution of primary cascade damage in tungsten. Acta Mater 144(Suppl C):905–917. https://doi.org/10.1016/j.actamat.2017.10.031

Nordlund K, Dudarev SL (2008) Interatomic potentials for simulating radiation damage effects in metals. C R Phys 9(34):343352. https://doi.org/10.1016/j.crhy.2007.10.012

Nordlund K, Ghaly M, Averback RS, Caturla M, Diaz de la Rubia T, Tarus J (1998a) Defect production in collision cascades in elemental semiconductors and fcc metals. Phys Rev B 57:7556–7570. https://doi.org/10.1103/PhysRevB.57.7556

Nordlund K, Wei L, Zhong Y, Averback RS (1998b) Role of electron-phonon coupling on collision cascade development in Ni, Pd, and Pt. Phys Rev B 57:R13965–R13968. https://doi.org/10.1103/PhysRevB.57.R13965

Nordlund K, Zinkle SJ, Suzudo T, Averback RS, Sand AE, Granberg F, Malerba L, Stoller R, Banhart F, Weber B, Willaime F, Dudarev S, Simeone D (2015) Primary radiation damage in materials: review of current understanding and proposed new standard displacement damage model to incorporate in-cascade mixing and defect production efficiency effects. Technical Report No. NEA/NSC/DOC (2015)9, OECD Nuclear Energy Agency, Paris

Norgett MJ, Robinson MT, Torrens IM (1975) A proposed method of calculating displacement dose rates. Nucl Eng Des 33:50–54

Paine B, Averback R (1985) Ion beam mixing: basic experiments. Nucl Instr Meth B 78, Part 2(0):666675. https://doi.org/10.1016/0168-583X(85)90451-3

Phillips CL, Crozier PS (2009) An energy-conserving two-temperature model of radiation damage in single-component and binary Lennard-Jones crystals. J Chem Phys 131(7):074701. https://doi.org/10.1063/1.3204030

Phillips CL, Magyar RJ, Crozier PS (2010) A two-temperature model of radiation damage in – quartz. J Chem Phys 133(14):144711. https://doi.org/10.1063/1.3481356

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117:1–19

Pruneda JM, Sánchez-Portal D, Arnau A, Juaristi JI, Artacho E (2007) Electronic stopping power in LiF from first principles. Phys Rev Lett 99:235501. https://doi.org/10.1103/PhysRevLett.99.235501

Race CP, Mason DR, Finnis MW, Foulkes WMC, Horsfield AP, Sutton AP (2010) The treatment of electronic excitations in atomistic models of radiation damage in metals. Rep Prog Phys 73(11):116501. http://stacks.iop.org/0034-4885/73/i=11/a=116501

Race CP, Mason DR, Sutton AP (2012) A new directional model for the electronic frictional forces in molecular dynamics simulations of radiation damage in metals. J Nucl Mater 425(1):33–40. https://doi.org/10.1016/j.jnucmat.2011.10.054

Rutherford AM, Duffy DM (2007) The effect of electron–ion interactions on radiation damage simulations. J Phys Condens Matter 19(49):496201. http://stacks.iop.org/0953-8984/19/i=49/a=496201

Sand AE, Nordlund K (2015) On the lower energy limit of electronic stopping in simulated collision cascades in Ni, Pd and Pt. J Nucl Mater 456(0):99105. https://doi.org/10.1016/j.jnucmat.2014.09.029

Sand AE, Dudarev SL, Nordlund K (2013) High energy collision cascades in tungsten: dislocation loops structure and clustering scaling laws. EPL 103:46003

Sand AE, Nordlund K, Dudarev S (2014) Radiation damage production in massive cascades initiated by fusion neutrons in tungsten. J Nucl Mater 455(13):207–211. https://doi.org/10.1016/j.jnucmat.2014.06.007. Proceedings of the 16th international conference on fusion reactor materials (ICFRM-16), Beijing, 20–26 Oct 2013

Sand A, Dequeker J, Becquart C, Domain C, Nordlund K (2016a) Non-equilibrium properties of interatomic potentials in cascade simulations in tungsten. J Nucl Mater 470:119–127. https://doi.org/10.1016/j.jnucmat.2015.12.012

Sand AE, Aliaga MJ, Caturla MJ, Nordlund K (2016b) Surface effects and statistical laws of defects in primary radiation damage: tungsten vs. iron. Europhys Lett 115:36001

Sand AE, Mason DR, Backer AD, Yi X, Dudarev SL, Nordlund K (2017) Cascade fragmentation: deviation from power law in primary radiation damage. Mater Res Lett 5(5):357–363. https://doi.org/10.1080/21663831.2017.1294117

Semrad D (1986) Coulomb effect and threshold effect in electronic stopping power for slow protons. Phys Rev A 33:1646–1652. https://doi.org/10.1103/PhysRevA.33.1646

Sillanpää J (2000) Electronic stopping of silicon from a 3D charge distribution. Nucl Instrum Methods Phys Res Sect B 164–165(Suppl C):302–309. https://doi.org/10.1016/S0168-583X(99)01030-7

Stoller R, Odette G, Wirth B (1997) Primary damage formation in bcc iron. J Nucl Mater 251(Suppl C):49–60. https://doi.org/10.1016/S0022-3115(97)00256-0. Proceedings of the international workshop on defect production, accumulation and materials performance in an irradiation environment

Todorov IT, Smith W, Trachenko K, Dove MT (2006) DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. J Mater Chem 16:19111918. https://doi.org/10.1039/B517931A

Toulemonde M, Dufour C, Paumier E (1992) Transient thermal process after a high-energy heavy-ion irradiation of amorphous metals and semiconductors. Phys Rev B 46:14362–14369. https://doi.org/10.1103/PhysRevB.46.14362

Was GS (2007) Fundamentals of radiation materials science: metals and alloys. Springer, Berlin Heidelberg. http://books.google.co.uk/books?id=0N06swAJI0AC

Wilson RG, Jamba DM, Chu PK, Hopkins CG, Hitzman CJ (1986) Atom and acceptor depth distributions for aluminum channeled in silicon as a function of ion energy and crystal orientation. J Appl Phys 60(8):2806–2809. https://doi.org/10.1063/1.337061

Yi X, Jenkins ML, Roberts SG, Zhou Z, Kirk MA (2013) *In situ* study of self-ion irradiation damage in W and W-5Re at 500 C. Philos Mag A 93(14):1715–1738. https://doi.org/10.1080/14786435.2012.754110

Yi X, Sand AE, Mason DR, Kirk MA, Roberts SG, Nordlund K, Dudarev SL (2015) Direct observation of size scaling and elastic interaction between nano-scale defects in collision cascades. EPL 110(3):36001. http://stacks.iop.org/0295-5075/110/i=3/a=36001

Zarkadoula E, Daraszewicz SL, Duffy DM, Seaton MA, Todorov IT, Nordlund K, Dove MT, Trachenko K (2014) Electronic effects in high-energy radiation damage in iron. J Phys Condens Matter 26(8):085401. http://stacks.iop.org/0953-8984/26/i=8/a=085401

Zarkadoula E, Duffy DM, Nordlund K, Seaton M, Todorov IT, Weber WJ, Trachenko K (2015) Electronic effects in high-energy radiation damage in tungsten. J Phys Condens Matter 27:135401

Zarkadoula E, Samolyuk G, Weber WJ (2017) Effects of the electron-phonon coupling activation in collision cascades. J Nucl Mater 490(Suppl C):317–322. https://doi.org/10.1016/j.jnucmat.2017.04.020

Zeb MA, Kohanoff J, Sánchez-Portal D, Arnau A, Juaristi JI, Artacho E (2012) Electronic stopping power in gold: the role of *d* electrons and the **H/**He anomaly. Phys Rev Lett 108:225504. https://doi.org/10.1103/PhysRevLett.108.225504

Zhang C, Mao F, Zhang FS (2013) Electron–ion coupling effects on radiation damage in cubic silicon carbide. J Phys Condens Matter 25(23):235402

Zhong Y, Nordlund K, Ghaly M, Averback RS (1998) Defect production in tungsten: a comparison between field-ion microscopy and molecular-dynamics simulations. Phys Rev B 58:2361–2364. https://doi.org/10.1103/PhysRevB.58.2361

Ziegler JF (2008) SRIM-2008.04 software package. http://www.srim.org

Ziegler JF, Ziegler MD, Biersack J (2010) SRIM the stopping and range of ions in matter (2010). Nucl Instr Meth B 268(1112):1818–1823. https://doi.org/10.1016/j.nimb.2010.02.091

Atomistic Kinetic Monte Carlo and Solute Effects

99

Charlotte S. Becquart, Normand Mousseau,
and Christophe Domain

Contents

Abstract

The atomistic approach of the kinetic Monte Carlo methods allows one to explicitly take into account solute atoms. In this chapter, we present and discuss the different pathways available at this point to go behind nearest neighbor pair interaction for binary alloys on rigid lattices as well as their perspectives. Different strategies to treat complex alloys with several solutes with improved

C. S. Becquart (✉)
Univ.Lille, CNRS, INRA, ENSCL, UMR 8207, UMET, Unité Matériaux et Transformations, Lille, France
e-mail: charlotte.becquart@univ-lille.fr

N. Mousseau
Département de physique and Regroupement québécois sur les matériaux de pointe, Université de Montréal, Montréal, QC, Canada

C. Domain
Département MMC, Les Renardières, Moret sur Loing, France

2437

cohesive models are exposed and illustrated as well as the modeling of self-interstitial diffusion under irradiation and its complexity compared to vacancy diffusion.

Abbreviations

AKMC	Atomic kinetic Monte Carlo
BKL	Bortz, Kalos, and Lebowitz
CE	Cluster expansion
DFT	Density functional theory
FIA	Foreign interstitial atom
FISE	Final initial system energy
GAP	Gaussian approximation potential
KMC	Kinetic Monte Carlo
KRA	Kinetically resolved activation
LAE	Local atomic environment
NEB	Nudged elastic band
PD	Point defect
RPV	Reactor pressure vessel
RTA	Residence time algorithm
SFT	Stacking fault tetrahedra
SIA	Self interstitial atom
SNAP	Spectral neighbor analysis potential

1 Introduction

The kinetics of alloy microstructures under nonequilibrium conditions such as irradiation is driven by the evolution of the population of point defects introduced by the energetic particles impinging the materials. Being capable of predicting such evolution is an important academic as well as industrial issue that needs to be tackled starting at the atomistic level. For this purpose, Atomistic Kinetic Monte Carlo (AKMC) is a very versatile method that can be used to simulate the evolution of complex microstructures at the atomic scale, dealing with elementary atomic mechanisms. It was developed more than 40 years ago to investigate diffusion events triggered by the motion of a single vacancy (Young and Elcock 1966) in a binary alloy. Since then, hetero-interstitials (Clouet et al. 2006; Hin et al. 2008) and self interstitials (Ngayam-Happy et al. 2010) have been introduced in the models and alloys of more complex nature have been investigated (Cerezo et al. 2003; Liu et al. 1997). The purpose of this chapter is to discuss how the effect of solutes can be taken into account in this approach. We will first start by describing briefly the basis of the AKMC method and its key ingredients. In the following sections, we present and discuss the different strategies available to build Hamiltonians that take into account solute atoms as well as the techniques used to estimate the activation barriers. We conclude by discussing the outlooks and perspectives of the different approaches described in the previous sections.

2   AKMC Method

Kinetic Monte Carlo (KMC) methods are techniques intended to simulate the time evolution of processes that occur with a given known rate, used as inputs to the algorithm (Voter 2007). It is usually applied at two scales: the atomistic scale, where processes that can take place are the ones atoms undergo during the evolution of a microstructure, and the mesoscopic scale, where the microstructure is coarse grained into objects that can evolve on a lattice or not. The first papers presenting a kinetic Monte Carlo approach were published by Young and Elcock (1966) and by Bortz et al. (1975), the first dealing with the kinetic of the vacancy diffusion in ordered alloys, the latter dealing with the kinetic evolution of an Ising spin system. A full theoretical description and formalization of the AKMC algorithm was proposed by Fichthorn (Fichthorn and Weinberg 1991). The most important feature of the AKMC technique – as compared to atomistic Metropolis Monte Carlo (MMC) simulations (Metropolis 1953) – is the fact that a time step can be related to each simulation step, i.e., for instance, the jump of a point defect. In AKMC simulations, configurations are not generated as a Markov chain and sampled respecting a given probability distribution (usually a thermodynamic equilibrium distribution) – as it is done in the MMC approach; rather, the configuration at step i is obtained from configuration at step $i\text{-}1$ by realizing a point defect or an atom move, i.e., the Markov chain is constructed taking into account the kinetics of the system.

We will not give an extensive description of the KMC algorithm which can be found for instance in (Voter 2007) but rather focus on its applications to model alloys and solute effects. We will remain in the context of diffusion controlled phase transformation under classical conditions (for instance under thermal aging) or in the presence of external forcing: here irradiation. A review of AKMC applied to precipitation can be found in Becquart and Soisson (2018). Atomistic kinetic Monte Carlo models are based on the residence time algorithm (Young and Elcock 1966) sometimes referred to as the BKL approach (Bortz et al. 1975). The elementary mechanisms are the point defect jumps or the interstitial jumps in the case of C or other hetero-interstitials. In the rigid lattice approach, which is the approach most used so far, vacancies and interstitials (dumbbells or mixed dumbbells) can jump from one lattice site to another lattice site (typically first nearest neighbor sites). If hetero-interstitial atoms are included in the model, they lie on an interstitial sublattice and jump on this sublattice. Typically one atom is assumed to move and the rates of possible transitions are determined from the local environment around the moving atom. According to harmonic standard transition state theory (Vineyard 1957), the frequency of a thermally activated event such as a vacancy jump in an alloy can be expressed as:

$$\Gamma = \nu e^{-\frac{E_a}{k_B T}} \tag{1}$$

where ν denotes the attempt frequency and E_a corresponds to the activation energy of the jump or the migration energy obtained as the difference between the energies

of the system at the saddle-point and in its initial configuration. Such a description is derived from the theory of thermally activated processes and is justified when the thermal fluctuations are smaller than the activation energies: $k_B T \ll E_a$. The attempt frequency can be expressed as:

$$\nu = \frac{\prod_{j=1}^{3N-3} \nu_j}{\prod_{j=1}^{3N-4} \nu_j^*} \qquad (2)$$

where ν_j and ν_j^* are the normal frequencies for vibrations at the local minimum and saddle states, respectively, and N is the number of atoms. Note that, at constant pressure, one should consider the enthalpies and free enthalpies of migration, but the difference is often negligible in solids. The jump frequencies obtained using this theory appear to be a very good approximation for the real jump frequencies up to at least half the melting point in most solid materials (Voter 2007).

For sake of simplicity, constant prefactors of the order of Debye's frequency are used most of the time (typical values for transition metals are 10^{12}–10^{13} s^{-1}). This commonly chosen assumption is based on the fact that small variations of the activation energy are more likely to have a greater impact on the transition probability than variations of the attempt frequency which are second-order effects (because of the exponential dependency of the former). This assumption is supposed to be valid especially for systems where all species have similar sizes such as 3d transition metal typical alloying elements in steels (Fe, Ni, Cr, Mn). However, it has been shown recently (Lazauskas et al. 2014) using Vineyard approach (Vineyard 1957) that the prefactor has an impact on the primary damage evolution and that the use of a constant prefactor, as is done in many lattice KMC simulations, may change the self interstitial atom (SIA) migration mechanisms, the amount of vacancy SIA recombinations, and enhance the difference between the diffusion rates of vacancies and interstitials. This work also concluded that the vibrational internal energy contributes little to the attempt frequencies of small point defect cluster migration mechanisms in Fe and that Vineyard approach was thus accurate enough to obtain them for each of the defects simulated.

At each KMC step, the algorithm determines the jump frequencies of all possible jumps in the system, chooses one of them according to its probability, and evaluates the time that the system would have taken to do the jump. The associated time-step length δt and average time-step length Δt is given by:

$$\delta t = \frac{-\ln r}{\sum_n \Gamma_X} \quad \Delta t = \frac{1}{\sum_n \Gamma_X} \qquad (3)$$

where r is a random number between 0 and 1.

The key issue of this method is the estimation of the jump frequency and the activation barrier, E_a, which, when the local chemical environment is complex, i.e., when the system is composed of several atomic elements, is not straightforward to calculate.

Three kinds of methods have been used, so far, to obtain E_a:

(i) **Direct calculation of E_a**: This approach is possible for simple cases, for
example, in the study of solute transport mechanisms where only one solute and
one point defect (vacancy or interstitial) are considered, i.e., when the number
of different barriers to determine is small. Such a study was performed by Costa
(2012), using an empirical potential and assuming a limited range of interaction,
e.g., first and second nearest neighbor interactions, to study thermal aging and
spinodal decomposition in the FeCr system, or, more recently, Messina et al.
(2015) using ab initio-computed binding energies and migration barriers in the
FeMnNi system. Usually, the activation energies are tabulated, to speed up
the calculations; however, it is also possible to obtain them on-the-fly, using
approaches such as k-ART (El-Mellouhi et al. 2008) or SEAKMC (Xu et al.
2015) discussed elsewhere in this volume (Chaps. 34, "Off-Lattice Kinetic
Monte Carlo Methods" and 32, "Computational Methods for Long-Timescale
Atomistic Simulations").

Another possibility to obtain E_a is to perform a cluster expansion (CE) at
the saddle point, as proposed in Rehman et al. (2013) again when the system
is not too complex or to use machine learning method based methods, such as
artificial neural network, that can be trained to estimate the activation energy as
a function of the local environment as proposed 10 years ago by Djurabekova
(Djurabekova et al. 2007), see Behler (2016) for a review.

(ii) **Broken bond or cut bond models** (Soisson et al. 1996): In these models, the
activation energy is given by the difference between the energy at the saddle
point position and the energy due to the interactions of the bonds broken during
the jump. The activation/migration energy is obtained as:

$$E_a = e_k^{sp} - \sum_j \varepsilon_{k-j} - \sum_{j \neq k} \varepsilon_{PD-j} \qquad (4)$$

where e_k^{sp} is the binding energy of atom k at the saddle point position and where
the summations over j extend over all the broken bonds of the migrating atom
k and of the moving point defect (PD). This scheme is illustrated in Fig. 1.
Another form of cut bond model has been developed in Vincent et al. (2008a).
In this method, it is considered that the migrating atom jump does not lead to
the destruction of the bonds with the point defect but that the bonds between
the migrating atom and its neighbors are replaced by bonds with the PD and the
same neighbors. The activation energy is then given by:

$$E_a = \sum_j \varepsilon_{k-j}^{sp} + \sum_j \varepsilon_{PD-j} - \sum_j \varepsilon_{k-j} \qquad (5)$$

(iii) **Methods based on the energy of the final and initial states:** They require
that a Hamiltonian is available to obtain the energy of any equilibrium
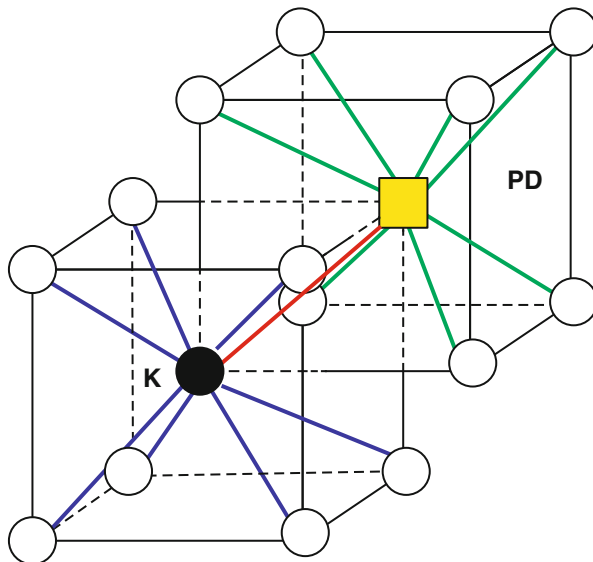
Fig. 1 Illustration of the broken bond model. The green lines represent the bonds around the jumping point defect and the blue lines the bonds around its final position

configurations and can be used with any Hamiltonian (simple ones like pair interactions or more complex ones). The kinetically resolved activation (KRA) according to van der Ven (Van der Ven et al. 2001), or final initial state energy (FISE) approximation, according to the terminology adopted by Vincent et al.(2008a) in the context of Fe-cu alloys, is based on the Kang and Weinberg decomposition of migration energy barriers (Kang and Weinberg 1989), which consists in computing the migration energy as follows:

$$E_a = E_a^0 + \frac{\Delta E}{2} \tag{6}$$

where E_a^0 depends usually on the migrating atom type but not on the chemical environment. The method, originally applied to vacancies, can be easily extrapolated to the migration of SIAs or foreign interstitial atoms (FIAs).

Whatever the method chosen to determine E_a, it must fulfill the detailed balance. Broken bond and FISE/KRA approaches have been compared in a study of FeCu alloys thermal aging (Vincent et al. 2008a), and it was found, not very surprisingly, that it is the parameterization rather than the method which impacts the most on the results, i.e., the density and mean precipitate radii.

The next section describes the different approaches used to determine the activation energies as well as build the energetic model of the system. The energetic model has to be built on a compromise between the complexity of the system (e.g., the number of solutes to treat) and the combinatorial increase of parameters to

determine and adjust, as well as a computational simplicity in order to be able to perform a very large number of KMC steps to simulate the target microstructure evolution.

3   Hamiltonians for AKMC Simulations

3.1   Exact Hamiltonians Based on DFT and Empirical Potentials

The evolution of the microstructure is driven by the point defects and the hetero-interstitials motions in the lattice. The key issue in these models is, as already mentioned in the previous section, the calculation of the jump probabilities. In the case of perfect and pure elements with simple defects, jump probabilities are not difficult to determine as it is easy to calculate the migration barrier and the attempt frequency using density functional theory (DFT) (i.e. *ab initio* or first principles approaches) calculations. However, in reals systems, solute atoms and strains will impact the activation barriers as described for instance in Tchitchekova et al. (2014) which can lead to incorrect predictions for the barriers even with a highly accurate energy calculation. The ideal way to compute jump frequencies would be to use a first principles approach on a box sufficiently large to limit size effects. Even though first principles calculations are not strictly exact, since they are based on a certain number of approximations which are more or less well controlled according to the different cases, its use represents, at the moment, the most reliable approach to obtain the potential energy for a given atomic distribution. Many methods for finding barriers exist and are implemented in DFT-based codes to determine activation barriers or migration energies. They include the nudged elastic band (NEB) (Henkelman et al. 2000), the growing spring method (Peters et al. 2004), the dimer (Henkelman and Jónsson 1999), autonomous basin climbing (ABC) (Fan et al. 2010) or eigen-vector following methods such as ART-nouveau (Barkema and Mousseau 1996; El-Mellouhi et al. 2008). However, these methods are computationally very demanding when applied to DFT calculations, and they have been used so far only in simple cases such as the estimation of tracer diffusion coefficients that require the calculations of only a few jump frequencies (Mantina et al. 2008; Messina et al. 2014). This is why it is often necessary to turn to empirical potential or force field approach to assess energy barriers. While these potentials are cheap enough to avoid strain artefacts due to small boxes, one of their recurring problems is their transferability, and more precisely their ability to correctly model the desired material under conditions far from the conditions used when adjusting the parameters. Even with empirical potentials, methods such as "on-the-fly" AKMC simulations that can require billions of point defects, jumps remain time consuming without additional tricks (Athènes and Bulatov 2014). Furthermore, the empirical potentials available at the moment are mostly limited to binaries, a few ternaries and even less quaternaries. Often, therefore, simpler models based on strong approximations typically associated with rigid lattices that are described in the following subsections, and illustrated with a few applications, must be used.

3.2 Effective Hamiltonians Based on Cluster Expansions

Cluster expansion methods provide a formalism to derive Hamiltonian, based on a decomposition of the interaction into pairs (that can be between different nearest neighbor distances), triplets (ε_{ijk}), quadruplets (ε_{ijkl}), quintuplet, and so on (Sanchez et al. 1984). The total energy of the system is obtained as:

$$E_{tot} = \sum_i \varepsilon_i + \sum_i \sum_{j<i} \varepsilon_{ij} + \sum_i \sum_{j<i} \sum_{k<j} \varepsilon_{ijk} + \sum_i \sum_{j<i} \sum_{k<j} \sum_{l<k} \varepsilon_{ijkl} + \cdots \quad (7)$$

3.2.1 Pair Models

The simplest approach is to use pair interaction models (Fig. 2). These models, based simply on the presence or absence of a given atom, require a rigid lattice and are often parameterized on experimental or DFT thermodynamical data.

In the case of an Fe alloy containing vacancies, the total energy of the system can be obtained as:

$$
\begin{aligned}
E = &\sum_j \varepsilon_{(Fe-Fe)}^{(i)} + \sum_k \varepsilon_{(V-V)}^{(i)} + \sum_l \varepsilon_{(Fe-V)}^{(i)} + \sum_m \varepsilon_{(Fe-X)}^{(i)} \\
&+ \sum_n \varepsilon_{(V-X)}^{(i)} + \sum_p \varepsilon_{(X-Y)}^{(i)}
\end{aligned}
\quad (8)
$$

where V stands for the vacancy, X and Y are solute atoms.

When interactions are limited to first nearest pairs, the thermodynamical properties of the system are easily determined using the Bragg Williams approximation. This approach has been applied for instance to study the precipitation of Cu in Fe, either using a broken bond model (Soisson et al. 1996) or FISE model (Vincent et al. 2008a) to determine E_a. This is the approach also chosen by Liu et al. to investigate complex Fe-CuNiMnSi alloys (Liu et al. 1997) using only first nearest neighbor parameters adjusted on binary mixing enthalpies. It has since then been shown that in bcc metals, one needs to go beyond the first nearest neighbor interaction to correctly reproduce the interaction of solutes with vacancies or to describe properly solute transport mechanism predicted by DFT calculations (Olsson et al. 2010) (Vincent et al. 2006; Messina et al. 2014).

Fig. 2 Illustration of a pair interaction model
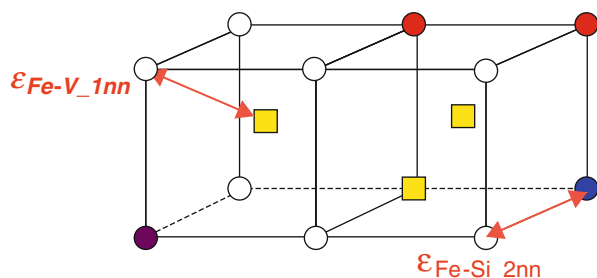
Hamiltonian-based pair models have also been derived for SIAs. These defects are more complex than vacancies: different dumbbells form in fcc or bcc metals, with off-lattice positions. The models have to be simplified as in Soisson (2006) or Soisson and Jourdan (2016) for which only the elements of the dumbbell X–Y are stored without considering the detailed migration mechanism and the dumbbell orientation or adapted on the main diffusion mechanisms as in Vincent et al. (2008b). In bcc Fe, most of the time, only the ⟨110⟩ dumbbell is considered, with eight possible jumps (Fig. 3a). It migrates in general according to the Johnson mechanism (Johnson 1964; Vincent et al. 2008b; Ngayam-Happy et al. 2010). In fcc
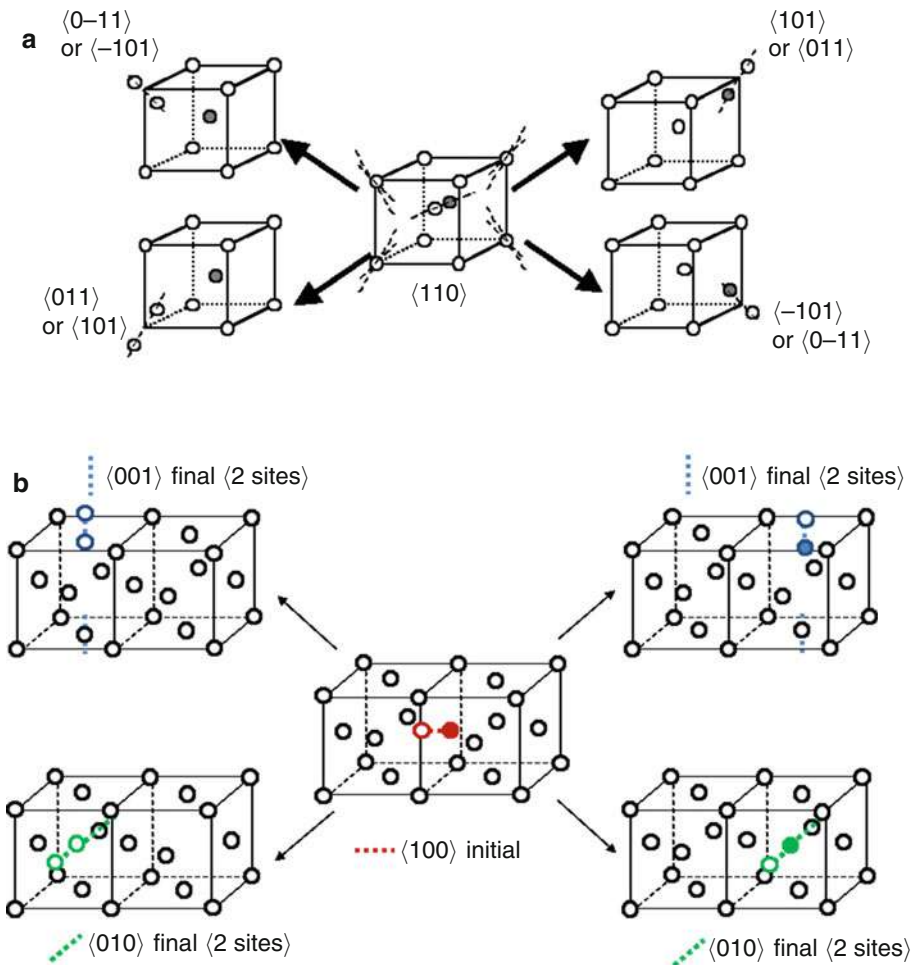
Fig. 3 SIA migration mechanisms considered (a) in bcc Fe, (b) in fcc Ni, Ni alloys, FeNiCr model fcc alloys

solute - dumbbell
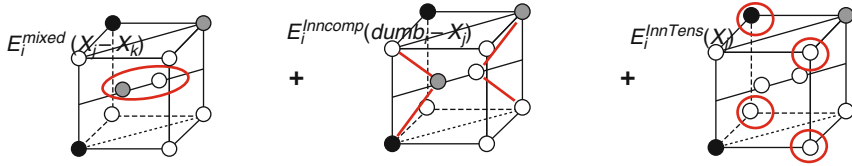
dumbbell - dumbbell
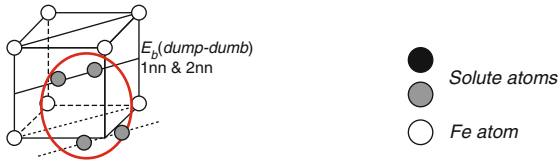
Fig. 4 SIA-solute pair interaction model proposed by Vincent et al. (Vincent et al. 2008b). Three kinds of interactions are considered: dumbbell-dumbbell, dumbbell-compression, and dumbbell-tension sites interactions

alloys such as Ni, Ni alloys or FeNiCr model fcc alloys used to simulate austenitic materials, the $\langle 100 \rangle$ dumbbell is considered with eight possible translation rotation jumps (Fig. 3b).

One possible approach decomposes the SIA interaction into pairs that take into account the different nearest neighbor sites and the sign of the strains of these sites as illustrated on Fig. 4. Using this approach, the dumbbell contribution to the total energy of the system can be obtained as (Vincent et al. 2008b):

$$
E_{\text{dumb}} = \sum_i \left(E_f + \sum_j E_l^{1nn\text{Comp}} \left(\text{dumb}_i - X_j \right) + \sum_j E_l^{1nn\text{Tens}} \left(X_j \right) \right.
$$

$$
\left. + \sum_{i,j} E_l^{\text{mixte}} \left(X_j - X_k \right) + \sum_l E_l \left(\text{dumb} - \text{dumb} \right) \right)
$$

where three kinds of interactions are considered: dumbbell-dumbbell, dumbbell-compression, and dumbbell-tension sites.

FIAs can also be introduced on a sublattice as shown in Hin et al. (2008), where the precipitation of carbides in Fe and FeNb alloys has been modeled under thermal aging using a broken bond model.

Pair interactions and a broken bond model have been used, for instance, to study radiation induced segregation by electrons and neutrons in Fe-15 at %Cr alloys (Soisson and Jourdan 2016) taking into account replacement collision sequences and ballistic mixing as well as the migration of both vacancies and $\langle 110 \rangle$ dumbbells. Hocker et al. (Hocker et al. 2014) used a similar approach to study precipitation in FeCuNiMn alloys whereas Vincent and coworkers (Vincent et al. 2008b) chose the FISE/KRA approach and pair interactions to model FeCuNiMnSi alloys under irradiation.

Fig. 5 Most compact
structure for a triplet and a
quadruplet in a bcc lattice
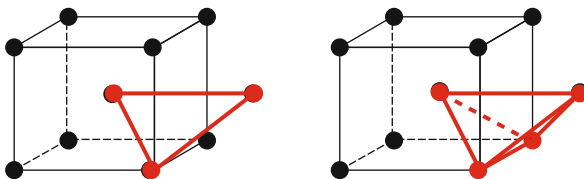
3.2.2   Higher Order Cluster Expansion Models

When expanding the model beyond pair interactions, complexity increases rapidly
as different geometries must be defined for all n-uplets. The most compact ones, as
illustrated for bcc in Fig. 5, can be privileged in order to limit the possible choices
to the most significant interactions – with the selection, in general of solute atoms
close to each other.

Since each triplet has a different energy contribution, function of the three species
considered, the number of energy parameters (i.e., n-uplets containing combinations
of all the solutes) that need to be defined increases exponentially as the number
of solute species increases (Fig. 6). As in the case of pair interaction models,
parameters can be adjusted on experimental thermodynamical data and/or DFT
calculations. This method can be easily used in conjunction with the FISE/KRA
method as done in (Nguyen-Manh et al. 2008) to model thermal aging in FeCr
alloys or more recently to simulate thermal annealing in Fe–20Cr–xW alloys at
773 K (Bonny et al. 2017). The same approach was used to study ordering in
NiCr fcc alloys by Barnard et al. (2014) who used the CASM package (CASMcode
2017; CASM Developers 2016) to build their cluster expansion Hamiltonian which
was then used with the FISE/KRA approach to follow the formation of the Ni_2Cr
ordered phase under thermal aging. Van de Walle has developed the Alloy Theo-
retic Automated Toolkit (ATAT) (2017) that builds cluster expansion Hamiltonian
interfaced with DFT. ATAT includes tools for thermodynamical analysis and Monte
Carlo simulation but cannot perform kinetic Monte Carlo.

Note that even with simple pair models, the number of parameters required
increases very quickly with the amount of species that needs to be modeled as
can be seen from Fig. 6. Modelling realistic steels for instance remains a not
straightforward task, especially if the adjustment of the parameters is done "by
hand."

3.3   Neural Network and Machine Learning Methods

One promising approach by taking into account the local environment influence on
the migration energies is the use of artificial neural network (ANN) as proposed
by Djurabekova et al. 10 years ago (Djurabekova et al. 2007). The more accurate
the training sets, the better the predictions, thus building training sets from DFT
calculations appears to be the direction to take, as long as elastic effects are
correctly accounted for. For accuracy the training set should cover a large set of
configurations. This starts now to be possible with DFT, despite requiring large
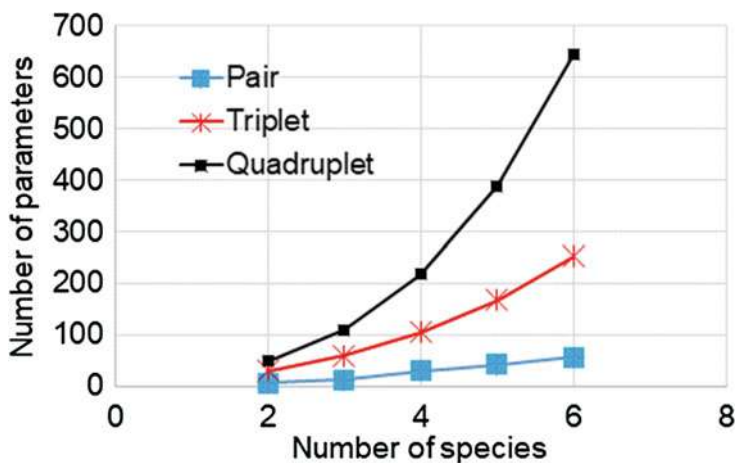
Fig. 6 Evolution of the number of parameters versus the number of species for different approaches

computing resources. Castin et al. (2017) and Messina et al. (2017) combining DFT and ANN proved this approach to be manageable to accurately model the thermal aging of two very well-known systems: FeCr and FeCu. The training sets have been limited to 2,000–10,000 configurations, with extra care taken to choose different environments in order to reproduce both the solid solution and the vacancy environment close to a precipitate. Another possible approach is genetic algorithm implemented for instance in the open source code *potfit* that constructs force matching empirical potentials (Brommer et al. 2015).

4 Taking into Account Solute Effects in the Activation Barriers

The effect on the local atomic environment on the migration barrier can be introduced using different approaches that are discussed in that section.

A first approach is to have E_{sp} or E_a^0 depend on the local environment, and in a first-order approximation on the number of solute atoms in the vicinity of the jumping atom. In a typical FISE/KRA approach, such as the one employed in (Vincent et al. 2008b), all the effects of the environment on the migration barrier are carried by the ΔE term. This simple model can be improved by introducing a dependence of the E_a^0 term on the local atomic environment. In particular, an explicit dependence of E_a^0 on the migrating atom's saddle point local atomic environment – whose effect on the vacancy migration energy is particularly strong (Nguyen-Manh et al. 2008; Costa et al. 2014; Bouar and Soisson 2002), can be easily introduced. This has been done, with the FISE/KRA model, to study Cu precipitation in FeCu (Bouar and Soisson 2002) with a broken-bond model, thermal

aging in FeCr (Nguyen-Manh et al. 2008) or the spinodal decomposition in the same system (Costa 2012). All studies show, however, that saddle point energies are very sensitive to both the nature of the jumping atom and that of the first neighbors of the saddle point. As a result, the kinetics pathways are changed and, for instance, in the FeCu study, Le Bouar et al. observed that the dependence of the saddle-point binding energies on the local atomic configurations modifies the relative mobility of small Cu clusters and Cu monomers.

A more sophisticated model to take into account the LAE in a pair interaction model has been proposed by Martinez et al. (Martínez et al. 2012) to study FeCr alloys. In this approach, the pair interactions are fitted on DFT data and depend on the local concentration as well as temperature. This allows a much better description of the FeCr phase diagram and its miscibility gap.

Recently, this method has been adapted for multi-component dilute alloys and, more precisely, to model the Fe-CuNiMnSi system representative of Reactor Pressure Vessel (RPV) steels (Pannier 2017). In a first and second nearest neighbor pair interaction model, the pair interaction values depend on whether the vacancy is in a solid solution or in a precipitate. This method introduces some additional degrees of freedom as compared to the original method proposed in Vincent et al. (2008b) with a limited increase of the number of parameters that need to be adjusted ($2\times$ more compared to the original pair model).

The cluster expansion method can also be applied to determine the activation barrier. In their study of Li diffusion in Li_xCoO_2 oxides, Van der Ven et al. (Van der Ven et al. 2001) added a cluster expansion contribution to the first term (i.e., E_a^0 in Eq. 6) of the FISE /KRA equation, introducing a dependence of E_a^0 on the local environment, independently of the jump direction. Rehman et al. (Rehman et al. 2013) proposed a procedure based on cluster expansion to obtain directly the migration barriers encountered during diffusion of Ag on Ag(100) surfaces. They trained the cluster expansion on activation barriers obtained using the NEB method and found that they were able to generate an accurate process rate catalogue using this approach.

A mean field approach to include the influence of the local environment in the migration barrier has been proposed recently to study the diffusion of H in zircaloy (Zhang et al. 2017). At each KMC step, the migration barrier of each jump, E_a, is modified as:

$$E_a = E_a^0 + E_t^i \text{ if } R < c_t^i$$

where E_a^0 is the barrier without trapping, i.e., the activation energy in pure Zr. E_t^i is the binding energy of H at a trapping site t with a concentration c_t^i induced by solute i, the concentration of which is c_i. R is a random number drawn each time the jump rate is evaluated.

A last approach uses many body potentials or DFT calculations, recognizing that they take into account naturally the effects of mean local solute concentrations. Potentials based on the embedded atom method (EAM) (Daw and Baskes 1984) are

usually a good compromise to describe metals and their alloys. Such potentials, with additional refinements, have been used to describe the evolution of FeCr systems during short range ordering (Pareige et al. 2009) and spinodal decomposition (Pareige et al. 2011). In Kushima and Yildiz (2010), a database of migration barrier energies as a function of lattice strain for a set of representative defect distributions in the vicinity of the migration path in yttria stabilized zirconia was constructed using DFT and the NEB.

5 Specific Issues Linked to Irradiated Microstructures

Real microstructures contain interfaces such as grain boundaries, phase boundaries, and so on. They are important to take into account as segregation or depletion of solutes can take place in their vicinity because they can act as sinks or traps.

In the AKMC approach, grain boundaries can be modeled using different methods. They can simply be introduced as an atomic plane in the simulation boxes that has perfect sink properties for vacancies and SIAs (Soisson and Jourdan 2016) or with specific trapping properties in the energetic model (Soisson and Jourdan 2016; Piochaud 2013).

Another issue is the impact of energetic particles impacting alloys as in the case of irradiation. Depending on the energy of the incident particle, isolated Frenkel pairs (e.g., with electron irradiation) or displacement cascades with the formation of vacancies and self interstitials isolated and in clusters in a localized zone, few nanometers large (e.g., for high energy ions and neutrons) are created in a few pisoseconds. The associated time scale being much smaller than the atomic migration time scale, the point defects associated to one energetic particle interaction can be introduced between two atomic migration kinetic Monte Carlo steps.

Different approaches are possible to model such events which, in the case of a pure metal, can lead to similar microstructures; however, in alloys, they are not equivalent. In the case of Frenkel pairs, one can either introduce a vacancy and a SIA, at with or without a correlation distance between them, randomly in the simulation box as in Ngayam-Happy et al. (2010) or introduce a series of replacements ending by one vacancy and one SIA at the other end as in Soisson and Jourdan (2016). In the case of displacement cascades, they can be simulated by introducing, in the simulation box, the positions of vacancies and SIAs from a cascade database debris obtained by MD (Vincent et al. 2008b) or they can be obtained as a superposition of channeling sequences (along the <111> direction in bcc crystals for instance), replacement collision sequences (with a specific length), and replacement cascades, i.e., the introduction of vacancies and SIAs within the volume of a sphere (Soisson and Jourdan 2016). When replacements are made in alloys, solute mixing can occur, which may be important in concentrated alloys as they can lead to precipitation or ordering (Lear et al. 2017).

6   Future Directions and Perspectives

Modeling the microstructure evolution of alloys, during thermal aging, irradiation, or under any external constrain, is a difficult task as one has to deal with multiple solutes and/or several point defects. A balance between the computational cost, the cohesive model accuracy and the number of input parameters to build it, the chemical complexity (e.g., multi solute treatment) has to be found. This is why, even as off-lattice AKMC methods are introduced, the need will remain for lattice-based approaches as these are much faster and can achieve the billions of time steps required to understand the physics of defects and alloys in complex environments.

There is still considerable place for improvements, and we consider, in this section, a few directions that must be considered.

For dilute alloys, methods that estimate the activation energies in a specific local environment as a correction of the activation energy in a perfect unary matrix as in Zhang et al. (2017) can be used as long as the local environment remains dilute. These methods are thus appropriate to determine for instance the diffusion coefficients in dilute alloys but not precipitation or solute segregation.

Without a doubt, the most urgent advance must come from force field developers particularly when it comes to complex alloys and solutes. Since, *ab initio* approaches will remain confined to relatively small systems for the foreseeable future, we must turn to approximation. As far as the cohesive model is concerned, machine learning methods that are under development to build improved empirical potentials or cohesive models are currently the most promising path to provide a satisfactory solution both in terms of computational costs and precision. "Quantum accurate interatomic potentials" such as the Gaussian approximation potentials (GAP) (Bartók and Csányi 2015) or the Spectral Neighbor Analysis Potential (SNAP) (Thompson et al. 2015) can now be built, by constituting a very large DFT database and the use of machine learning tools. Fitting empirical potentials to functional forms based on the knowledge of the chemical interactions responsible for the material cohesion has limitations and the development of methods that uses very generic functional forms, not based on physical considerations, is well under way (Handley and Behler 2014). This is where machine learning algorithms become even more necessary. A recent review of these methods that discusses the central ideas behind machine learning potentials as well as their current applicability and limitations can be found in Behler (2016). Among the latter, the size of the database required to fit multicompound systems remains, nevertheless, a major hurdle that has not yet been solved. It is, however, only a matter of time before more complex systems become treatable, especially in the context of AKMC, where only local minima and transition require accuracy.

An accurate description of the energetics is not always sufficient, however. In many cases, variations in the entropy associated with diffusion mechanisms can completely change the kinetics. Yet, evaluating the contribution of entropy on the migration barriers is very computational demanding and will depend significantly,

again, on the cohesive model used. Theoretical methods exist to do so, see a review in Van der Ven et al. (2010); however, there require a considerable amount of additional calculations as one should calculate the configurational, vibrational, electronic, and magnetic entropic contributions as it has been done for defect and defect clusters formation energies (Marinica and Willaime 2007; Murali et al. 2015; Tucker et al. 2010; Posselt et al. 2017). This becomes especially important for inhomogeneous conditions such as grain boundaries, high entropy alloys, and surfaces. For example, the tetrahedral vacancy cluster, formed by a three-vacancy cluster with a displaced atom in the center (Aidhy et al. 2016) is very unstable at low temperature in Ni, with an energy 0.4 eV above ground state, separated by a barrier of only 0.08 eV, but is found to be stable above 400 K, due to entropic effects (Mahmoud et al. 2018).

Incorporating these effects in lattice-based AKMC will require a better approach to evaluate barriers, including energy and entropy, without exploding the event catalog and the number of local environments increases with solutes. This could be done by expending on cluster expansion methods. One promising improvement in this approach towards modeling more realistic microstructures is the methods derived by Yuge and co-worker (Yuge and Okawa 2014) to take into account strain effects which can be non negligible during alloys phase transitions, and handle multiple lattices (Yuge 2012). The performance of three CE fitting algorithms have been investigated in Herder et al. (2015) who find that performance depends on system details and complexity and becomes a trade-off between accuracy and computational cost. They find that quality of a CE depends on the data that it is fit to and the algorithm used to identify terms to introduce in the CE.

Clearly, CE methods will have to consider more complex and larger objects containing 10 or 20 atoms. To do that requires a better description and classification of local environments. Topological analysis tools such as NAUTY (McKay and Piperno 2014) or the graph approach proposed in Yuge (2017), for example, while first used with off-lattice KMC approaches, provide a very efficient way to differentiate between local environments, even with a large number of solutes, while automatically taking symmetry operations into account.

Even if this is not the purpose of this chapter, one important issue is the speeding up of the simulations as the calculation speed has to be taken into account in the choice of the cohesive models. In particular, when both low and high migration barriers can be encountered during the point defect motion, the defect can be trapped and many KMC steps are necessary to change configuration which can drastically slow down the evolution of the microstructure. Different methods are being developed to boost the simulation, see for instance Athènes and Bulatov (2014) or Danielson et al. (2017).

References

Aidhy DS, Lu C, Jin K, Bei H, Zhang Y, Wang L, Weber WJ (2016) Formation and growth of stacking fault tetrahedra in Ni via vacancy aggregation mechanism. Scr Mater 114:137–141

Alloy Theoretic Automated Toolkit (ATAT) Home Page. https://www.brown.edu/Departments/Engineering/Labs/avdw/atat/. Accessed 7 Nov 2017

Athènes M, Bulatov VV (2014) Path factorization approach to stochastic simulations. Phys Rev Lett 113:230601. https://doi.org/10.1103/PhysRevLett.113.230601

Barkema GT, Mousseau N (1996) Event-based relaxation of continuous disordered systems. Phys Rev Lett 77:4358–4361

Barnard L, Young GA, Swoboda B, Choudhury S, Van der Ven A, Morgan D, Tucker JD (2014) Atomistic modeling of the order–disorder phase transformation in the Ni_2Cr model alloy. Acta Mater 81:258–271

Bartók AP, Csányi G (2015) Gaussian approximation potentials: a brief tutorial introduction. Int J Quantum Chem 115:1051–1057

Becquart CS, Soisson F (2018) Monte Carlo simulations of precipitation under irradiation. In: Hsueh CH et al (eds) Handbook of mechanics of materials. Springer, Singapore

Behler J (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145:170901

Bonny G, Castin N, Domain C, Olsson P, Verreyken B, Pascuet MI, Terentyev D (2017) Density functional theory-based cluster expansion to simulate thermal annealing in FeCrW alloys. Philos Mag 97:299–317

Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. J Comput Phys 17:10–18

Bouar YL, Soisson F (2002) Kinetic pathways from embedded-atom-method potentials: influence of the activation barriers. Phys Rev B 65:094103. https://doi.org/10.1103/PhysRevB.65.094103

Brommer P, Kiselev A, Schopf D, Beck P, Roth J, Trebin H-R (2015) Classical interaction potentials for diverse materials from *ab initio* data: a review of *profit*. Model Simul Mater Sci Eng 23:074002

CASM Developers (2016) CASMcode: V0.2.0. https://doi.org/10.5281/zenodo.60142

CASMcode (2017) First-principles statistical mechanical software for the study of multi-component crystalline solids. PRISMS Center

Castin N, Messina L, Domain C, Pasianot RC, Olsson P (2017) Improved atomistic Monte Carlo models based on *ab-initio* -trained neural networks: application to FeCu and FeCr alloys. Phys Rev B 95:214117. https://doi.org/10.1103/PhysRevB.95.214117

Cerezo A, Hirosawa S, Rozdilsky I, Smith GDW (2003) Combined atomic-scale modelling and experimental studies of nucleation in the solid state. Philos Trans R Soc Math Phys Eng Sci 361:463–477

Clouet E, Hin C, Gendt D, Nastar M, Soisson F (2006) Kinetic Monte Carlo simulations of precipitation. Adv Eng Mater 8:1210–1214

Costa D (2012) Modelling the thermal ageing evolution of Fe-Cr alloys using a lattice based kinetic Monte Carlo approach based on DFT calculations. PhD dissertation, Université Lille

Costa D, Adjanor G, Becquart CS, Olsson P, Domain C (2014) Vacancy migration energy dependence on local chemical environment in Fe–Cr alloys: a density functional theory study. J Nucl Mater 452:425–433

Danielson T, Sutton JE, Hin C, Savara A (2017) SQERTSS: dynamic rank based throttling of transition probabilities in kinetic Monte Carlo simulations. Comput Phys Commun 219: 149–163

Daw MS, Baskes MI (1984) Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B 29:6443–6453

Djurabekova FG, Domingos R, Cerchiara G, Castin N, Vincent E, Malerba L (2007) Artificial intelligence applied to atomistic kinetic Monte Carlo simulations in Fe–Cu alloys. Nucl Instrum Methods Phys Res Sect B Beam Interact Mater At 255:8–12

El-Mellouhi F, Mousseau N, Lewis LJ (2008) Kinetic activation-relaxation technique: an off-lattice self-learning kinetic Monte Carlo algorithm. Phys Rev B 78:153202

Fan Y, Kushima A, Yildiz B (2010) Unfaulting mechanism of trapped self-interstitial atom clusters in bcc Fe: a kinetic study based on the potential energy landscape. Phys Rev B 81:104102

Fichthorn KA, Weinberg WH (1991) Theoretical foundations of dynamical Monte Carlo simulations. J Chem Phys 95:1090

Handley CM, Behler J (2014) Next generation interatomic potentials for condensed systems. Eur Phys J B 87:152. https://doi.org/10.1140/epjb/e2014-50070-0

Henkelman G, Jónsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111:7010

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113:9901

Herder LM, Bray JM, Schneider WF (2015) Comparison of cluster expansion fitting algorithms for interactions at surfaces. Surf Sci 640:104–111

Hin C, Bréchet Y, Maugis P, Soisson F (2008) Kinetics of heterogeneous grain boundary precipitation of NbC in α-iron: a Monte Carlo study. Acta Mater 56:5653–5667

Hocker S, Binkele P, Schmauder S (2014) Precipitation in α-Fe based Fe-cu-Ni-Mn-alloys: behaviour of Ni and Mn modelled by ab initio and kinetic Monte Carlo simulations. Appl Phys A Mater Sci Process 115:679–687

Johnson RA (1964) Interstitials and vacancies in α iron. Phys Rev 134:A1329–A1336

Kang HC, Weinberg WH (1989) Dynamic Monte Carlo with a proper energy barrier: surface diffusion and two-dimensional domain ordering. J Chem Phys 90:2824

Kushima A, Yildiz B (2010) Oxygen ion diffusivity in strained yttria stabilized zirconia: where is the fastest strain? J Mater Chem 20:4809

Lazauskas T, Kenny SD, Smith R (2014) Influence of the prefactor to defect motion in α-Iron during long time scale simulations. J Phys Condens Matter 26:395007

Lear CR, Bellon P, Averback RS (2017) Novel mechanism for order patterning in alloys driven by irradiation. Phys Rev B 96:104108

Liu CL, Odette GR, Wirth BD, Lucas GE (1997) A lattice Monte Carlo simulation of nanophase compositions and structures in irradiated pressure vessel Fe-Cu-Ni-Mn-Si steels. Mater Sci Eng A 238:202–209

Mahmoud S, Trochet M, Restrepo OA, Mousseau N (2018) Study of point defects diffusion in nickel using kinetic activation-relaxation technique. Acta Mater 144:679–690. https://doi.org/10.1016/j.actamat.2017.11.021

Mantina M, Wang Y, Arroyave R, Chen LQ, Liu ZK, Wolverton C (2008) First-principles calculation of self-diffusion coefficients. Phys Rev Lett 100:215901

Marinica MC, Willaime F (2007) Orientation of interstitials in clusters in α-Fe: a comparison between empirical potentials. Solid State Phenom 129:67–74

Martínez E, Senninger O, Fu C-C, Soisson F (2012) Decomposition kinetics of Fe-Cr solid solutions during thermal aging. Phys Rev B 86:224109. https://doi.org/10.1103/PhysRevB.86.224109

McKay BD, Piperno A (2014) Practical graph isomorphism, II. J Symb Comput 60:94–112

Messina L, Nastar M, Garnier T, Domain C, Olsson P (2014) Exact ab initio transport coefficients in bcc Fe-X (X=Cr,Cu,Mn,Ni,P,Si) dilute alloys. Phys Rev B 90:104203

Messina L, Malerba L, Olsson P (2015) Stability and mobility of small vacancy–solute complexes in Fe–MnNi and dilute Fe–X alloys: a kinetic Monte Carlo study. Nucl Instrum Methods Phys Res Sect B Beam Interact Mater At 352:61–66

Messina L, Castin N, Domain C, Olsson P (2017) Introducing *ab initio* based neural networks for transition-rate prediction in kinetic Monte Carlo simulations. Phys Rev B 95:064112. https://doi.org/10.1103/PhysRevB.95.064112

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092. https://doi.org/10.1063/1.1699114

Murali D, Posselt M, Schiwarth M (2015) First-principles calculation of defect free energies: general aspects illustrated in the case of bcc Fe. Phys Rev B 92:064103. https://doi.org/10.1103/PhysRevB.92.064103

Ngayam-Happy R, Olsson P, Becquart CS, Domain C (2010) Isochronal annealing of electron-irradiated dilute Fe alloys modelled by an ab initio based AKMC method: influence of solute–interstitial cluster properties. J Nucl Mater 407:16–28

Nguyen-Manh D, Lavrentiev MY, Dudarev SL (2008) The Fe–Cr system: atomistic modelling of thermodynamics and kinetics of phase transformations. Comptes Rendus Phys 9: 379–388

Olsson P, Klaver TPC, Domain C (2010) *Ab initio* study of solute transition-metal interactions with point defects in bcc Fe. Phys Rev B 81:054102. https://doi.org/10.1103/PhysRevB.81.054102

Pannier B (2017) Towards the prediction of microstructure evolution under irradiation of model ferritic alloys with an hybrid AKMC-OKMC approach. PhD dissertation, Université Lille

Pareige C, Domain C, Olsson P (2009) Short- and long-range orders in Fe–Cr: a Monte Carlo study. J Appl Phys 106:104906

Pareige C, Roussel M, Novy S, Kuksenko V, Olsson P, Domain C, Pareige P (2011) Kinetic study of phase transformation in a highly concentrated Fe–Cr alloy: Monte Carlo simulation versus experiments. Acta Mater 59:2404–2411

Peters B, Heyden A, Bell AT, Chakraborty A (2004) A growing string method for determining transition states: comparison to the nudged elastic band and string methods. J Chem Phys 120:7877–7886

Piochaud JB (2013) Modelling of radiation induced segregation in austenitic Fe alloys at the atomistic level. PhD dissertation, Université Lille

Posselt M, Murali D, Schiwarth M (2017) Influence of phonon and electron excitations on the free energy of defect clusters in solids: a first-principles study. Comput Mater Sci 127:284–294

Rehman T, Jaipal M, Chatterjee A (2013) A cluster expansion model for predicting activation barrier of atomic processes. J Comput Phys 243:244–259

Sanchez JM, Ducastelle F, Gratias D (1984) Generalized cluster description of multicomponent systems. Physica A Stat Mech Appl 128:334–350

Soisson F (2006) Kinetic Monte Carlo simulations of radiation induced segregation and precipitation. J Nucl Mater 349:235–250

Soisson F, Jourdan T (2016) Radiation-accelerated precipitation in Fe–Cr alloys. Acta Mater 103:870–881

Soisson F, Barbu A, Martin G (1996) Monte Carlo simulations of copper precipitation in dilute iron-copper alloys during thermal ageing and under electron irradiation. Acta Mater 44:3789–3800

Tchitchekova DS, Morthomas J, Ribeiro F, Ducher R, Perez M (2014) A novel method for calculating the energy barriers for carbon diffusion in ferrite under heterogeneous stress. J Chem Phys 141:034118

Thompson AP, Swiler LP, Trott CR, Foiles SM, Tucker GJ (2015) Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. J Comput Phys 285:316–330

Tucker JD, Najafabadi R, Allen TR, Morgan D (2010) Ab initio-based diffusion theory and tracer diffusion in Ni–Cr and Ni–Fe alloys. J Nucl Mater 405:216–234

Van der Ven A, Ceder G, Asta M, Tepesch PD (2001) First-principles theory of ionic diffusion with nondilute carriers. Phys Rev B 64:184307. https://doi.org/10.1103/PhysRevB.64.184307

Van der Ven A, Thomas JC, Xu Q, Bhattacharya J (2010) Linking the electronic structure of solids to their thermodynamic and kinetic properties. Math Comput Simul 80:1393–1410

Vincent E, Becquart CS, Domain C (2006) Solute interaction with point defects in α Fe during thermal ageing: a combined ab initio and atomic kinetic Monte Carlo approach. J Nucl Mater 351:88–99

Vincent E, Becquart CS, Pareige C, Pareige P, Domain C (2008a) Precipitation of the FeCu system: a critical review of atomic kinetic Monte Carlo simulations. J Nucl Mater 373:387–401

Vincent E, Becquart CS, Domain C (2008b) Microstructural evolution under high flux irradiation of dilute Fe–CuNiMnSi alloys studied by an atomic kinetic Monte Carlo model accounting for both vacancies and self interstitials. J Nucl Mater 382:154–159

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3:121–127

Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Sickafus KE, Kotomin EA, Uberuaga BP (eds) Radiat. Eff. Solids. Springer, Dordrecht, pp 1–23

Xu H, Stoller RE, Béland LK, Osetsky YN (2015) Self-evolving atomistic kinetic Monte Carlo simulations of defects in materials. Comput Mater Sci 100, Part B:135–143

Young WM, Elcock EW (1966) Monte Carlo studies of vacancy migration in binary ordered alloys: I. Proc Phys Soc 89:735

Yuge K (2012) Modeling configurational energetics on multiple lattices through extended cluster expansion. Phys Rev B 85:144105. https://doi.org/10.1103/PhysRevB.85.144105

Yuge K (2017) Graph representation for configurational properties of crystalline solids. J Phys Soc Jpn 86:024802

Yuge K, Okawa R (2014) Cluster expansion approach for modeling strain effects on alloy phase stability. Intermetallics 44:60–63

Zhang Y, Jiang C, Bai X (2017) Anisotropic hydrogen diffusion in α-Zr and Zircaloy predicted by accelerated kinetic Monte Carlo simulations. Sci Rep 7:41033

DFT-Parameterized Object Kinetic Monte Carlo Simulations of Radiation Damage

100

Enrique Martínez, María José Caturla, and Jaime Marian

Contents

E. Martínez (✉)
Material Science and Technology Division, MST-8, Los Alamos National Laboratory,
Los Alamos, NM, USA
e-mail: enriquem@lanl.gov

M. J. Caturla
Departamento de Física Aplicada, Facultad de Ciencias, Fase II, Universidad de Alicante,
Alicante, Spain

J. Marian
Department of Materials Science and Engineering and Department of Mechanical and Aerospace
Engineering, University of California, Los Angeles, CA, USA

Department of Mechanical and Aerospace Engineering, University of California,
Los Angeles, CA, USA

Abstract

Materials for nuclear applications are subjected to extremely stringent conditions. The incoming energetic particles create different types of defects in the material that modify the system microstructure. These defects diffuse and interact with each other and pre-existing features in the material, leading to alterations of the material properties, and even to failure. To deploy reliable materials for such extreme applications, a deep understanding of the microstructural changes and their relation to material properties is critically required. Synergistic experimental and theoretical studies are paramount to gain such crucial knowledge. In this work we review one theoretical venue developed over the years to first understand and then predict the material response upon irradiation: an object kinetic Monte Carlo (OKMC) approach parameterized to first-principles data. We review the theory behind the kinetic Monte Carlo (KMC) algorithm and the specifics of the OKMC as a mesoscale methodology. We describe density functional theory (DFT) as an ab initio approach that can accurately calculate parameters required by the OKMC as input data to be able to analyze the microstructure evolution of the system. Finally, we show two applications lengthily studied in the literature: the microstructural evolution of both ferritic steels and tungsten under diverse irradiation conditions.

1   Introduction

Irradiation has the potential to significantly alter the atomic structure of the material being irradiated. As energetic particles collide with the underlying atomic structure, extra energy is deposited in the material generating a sequence of structural modifications that are commonly denoted as defects. It is the evolution of these defects and their coupling with the material that dictate the resilience of the material to such irradiation conditions. Defects at the electronic level are known to vary the cohesive properties of the material (Correa et al. 2012; Draeger et al. 2017) and are usually characterized by short existence time on the order of femtoseconds. Defects at the atomic level, such as vacancies and interstitials, may also be created during the irradiation process. These structural defects are characterized by much longer lifetimes compared to electronic defects, during which these defects can migrate and interact with one another, generating new microstructural features that modify the material properties, and may eventually lead to failure. The time and length scales required to study the evolution of the irradiation-created defects exceed that attainable by molecular dynamics (MD) calculations. This is because MD solves Newton's equations of motion for an ensemble of atoms using an integration time step on the order of femtoseconds. This implies that the total number of steps

needed to reach simulated times of hours is around 3.6×10^{18}, far beyond the
current computational capabilities. Accelerated MD (AMD) algorithms (Voter 1997,
1998; Sørensen and Voter 2000) extend the physical time that can be studied using
MD approaches and provide good insights into complex mechanisms, but, because
they also rely on a full description of the interatomic forces, their application is
still limited in the system size as well as the simulated time. The kinetic Monte
Carlo (KMC) algorithm (Elcock and McCombie 1958; Elcock 1959; Young and
Elcock 1966; Bortz et al. 1975; Gillespie 1976, 1977; Voter 2007) also called the
Stochastic Simulation Algorithm provides a methodology to overcome the MD time
limitation. The scope of applications for KMC is extraordinarily wide, ranging
from epidemiology and population kinetics to surface growth or radiation damage.
This chapter reviews the time evolution of this last type of defects generated under
irradiation as analyzed theoretically with a computational tool called object kinetic
Monte Carlo (OKMC).

   The specific type of defects generated under irradiation and the spatial cor-
relations between them depend on the irradiation conditions and the material
subjected to irradiation. Under electronic irradiation, usually Frenkel pairs, i.e.,
a vacancy and a self-interstitial, are formed (Was 2007). On the other hand, if
irradiation is performed with neutrons, protons, or heavier ions with high enough
energies, a thermal spike occurs in the material leading to a displacement cascade.
Displacement cascades are commonly characterized by a low density core and a
denser outer shell, i.e., the core is vacancy rich while the periphery is self-interstitial
rich (Was 2007). The main goal of the OKMC is to study how these long-standing
defects diffuse, react, and couple with pre-existing material's features to be able
to first understand and then predict the microstructural changes taking place upon
specific irradiation conditions. For such end, the OKMC needs the probabilities per
unit time of all the possible events in the system. These probabilities are usually
calculated in the context of transition state theory (TST) (Eyring et al. 1944; Voter
and Doll 1984, 1985). According to TST, the rate for an event to happen follows an
Arrhenius expression

$$\Gamma_{\alpha \to \beta}^{TST} = \nu^* \exp\left(-\frac{\Delta G_{\alpha \to \beta}}{k_B T}\right) \tag{1}$$

where ν^* is a pre-exponential factor, $\Delta G_{\alpha \to \beta}$ the activation free energy for the
event that takes the system from state α to state β, k_B the Boltzmann constant,
and T the temperature. This model relies on several assumptions: (i) the adiabatic
approximation (Born-Oppenheimer) holds; (ii) the dynamics of the nuclei can be
described with the classical equations of motion; (iii) the configuration at the initial
minimum is at equilibrium, i.e., the system follows a Boltzmann distribution; and
(iv) recrossing events at the transition state are negligible. The TST rate represents
an upper bound to the true rate, $\Gamma_{\alpha \to \beta}^{TST} > \Gamma_{\alpha \to \beta}$, and can be adjusted considering
dynamical corrections (Voter and Doll 1984, 1985). In the crystalline systems of
most interest, at intermediate-to-low homologous temperatures, TST is commonly

approximated considering that the potential energy landscape is mostly harmonic (Vineyard 1957). In this approach, termed as harmonic TST (HTST), the rate is obtained through the expression

$$\Gamma_{\alpha \to \beta}^{HTST} = \frac{\prod_k^{3N} \omega_k^{\alpha}}{\prod_k^{3N-1} \omega_k^{ts}} \exp\left(-\frac{\Delta E_{\alpha \to \beta}}{k_B T}\right) \qquad (2)$$

where ω_k^{α} are the normal mode frequencies in the initial minimum configuration (α), ω_k^{ts} are the normal mode frequencies when the system is at the transition state (ts), and $\Delta E_{\alpha \to \beta}$ is the activation potential energy for the particular event. Assuming that HTST holds, the big question is then how to obtain the activation energies and the vibrational frequencies to calculate the rate for the specific event. As we shall see in the following, for small defects and defect clusters containing less than around six individual defects, the most accurate tool that is commonly used is density functional theory (DFT) (Sholl and Steckel 2009), which considers electronic degrees of freedom and relies on fewer assumptions than classical empirical potentials. It is worth noting that free energies can also be computed within the DFT approach, which sometimes allows for the harmonicity constrain to be relaxed. Although more accurate, the computational cost constrains the calculations to fairly small samples and thus the size of the defects to be studied. The activation energy might depend on both the migration energy and a binding energy of the defect to a microstructural feature. For larger defects, the calculation of the rates usually relies on empirical potential approaches, such that interactions between defects are minimized and accurate parameters are obtained.

In this chapter we first review the theory behind the kinetic Monte Carlo algorithm to then particularize into the object-like KMC methods. Subsequently, we review the density functional theory approach and the calculations that can be done to obtain the mechanisms and the rates for such events to finally give some examples of OKMC simulations of microstructure evolution under irradiation.

2 Formulation and Mathematical Aspects of the Kinetic Monte Carlo Algorithm

The KMC algorithm gives one realization of the *Markovian master equation*. In this section we will review the definition and derivation of the master equation and its relation to the KMC method.

2.1 The Master Equation

2.1.1 Markov Processes

A Markov process (Kampen 1992) is a stochastic process with the property that the outcome of one step depends only on the outcome of the previous step. Formally, this can be written for n successive measurements as

$$P(x_n, t_n | x_1, t_1; \cdots ; x_{n-1}, t_{n-1}) = P(x_n, t_n | x_{n-1}, t_{n-1}) \tag{3}$$

That is, the conditional probability density P to find the system in a state defined by x_n at time t_n, given by the value x_{n-1} at t_{n-1}, is uniquely determined and is not affected by any knowledge of the values at earlier times. P is called the transition probability, and x_n and t_n are two stochastic variables representing the measure and the time of the process.

A Markov process is fully determined by the two functions $P(x_1, t_1)$ and $P(x_2, t_2 | x_1, t_1)$; the whole process can be constructed from them. For instance, taking $t_1 < t_2 < t_3$

$$
\begin{aligned}
P(x_1, t_1; x_2, t_2; x_3, t_3) &= P(x_1, t_1; x_2, t_2) P(x_3, t_3 | x_1, t_1; x_2, t_2) \\
&= P(x_1, t_1) P(x_2, t_2 | x_1, t_1) P(x_3, t_3 | x_2, t_2) \tag{4}
\end{aligned}
$$

Continuing with this logic, one finds successively all P_n. This property makes Markov processes manageable, which is the reason why they are so useful in applications.

Integrating Eq. 4 over x_2 and dividing by $P(x_1, t_1)$, one obtains for $t_1 < t_2 < t_3$

$$P(x_3, t_3 | x_1, t_1) = \int P(x_3, t_3 | x_2, t_2) P(x_2, t_2 | x_1, t_1) dx_2 \tag{5}$$

This is the Chapman-Kolmogorov equation. It is an identity which must be obeyed by the transition probability of any Markov process. The time ordering is essential: t_2 lies between t_1 and t_3. Two important remarks are as follows:

1. Two different Markov processes can describe the same physical system, depending on the level of coarseness of the description.
2. A Markov process is not restricted to one-component processes. For any r-component stochastic process, one may ignore a number of components, and the remaining s components again constitute a stochastic process. But, if the r-component process is Markovian, the process formed by the $s < r$ components in general is not. Vice versa, if a certain physically given process is not Markovian, it is sometimes possible by introducing additional components, to embed it in a Markov process. These additional components serve to describe explicitly information that otherwise would be contained implicitly in the past values of the variables.

As a matter of fact, in general, any closed isolated physical system can be described as a Markov process introducing all microscopic variables as components of the process. The key problem is to find a small set of variables which evolution can be described as a multicomponent Markov process. This reduction to a smaller number of variables is called "contraction" or "projection" (Kampen 1992).

2.1.2 Derivation of the Master Equation

Consider a Markov process in which the transition probabilities only depend on the time difference $\tau = t_2 - t_1$ such that $P(x_2, t_2 | x_1, t_1) = T_\tau(x_2 | x_1)$. The Chapman-Kolmogorov Eq. 5 for T_τ is a functional relation, which is not easy to handle in actual applications. The master equation is a more convenient version of the same equation: it is a differential equation obtained by going to the limit of vanishing time difference τ. It can be shown (Kampen 1992) that $T_\tau(x_2 | x_1)$ for small τ has the form

$$T_\tau(x_2 | x_1) = (1 - a_0 \tau)\delta(x_2 - x_1) + \tau W(x_2 | x_1) + \mathcal{O}(\tau) \tag{6}$$

Here $W(x_2 | x_1)$ is the transition probability per unit time from x_1 to x_2, and hence

$$W(x_2 | x_1) \geq 0 \tag{7}$$

The coefficient $(1 - a_0 \tau)$ in front of the delta function is the probability that no transition takes place during τ; hence

$$a_0(x_1) = \int W(x_2 | x_1) dx_2 \tag{8}$$

The Chapman-Kolmogorov Eq. 5 can be written as

$$T_{\tau + \tau'}(x_3 | x_1) = [1 - a_0(x_3)\tau']T_\tau(x_3 | x_1) + \tau' \int W(x_3 | x_2)T_\tau(x_2 | x_1) dx_2 \tag{9}$$

Dividing by τ', go to the limit $\tau' \to 0$, and use Eq. 8

$$\frac{\partial}{\partial \tau} T_\tau(x_3 | x_1) = \int \{W(x_3 | x_2)T_\tau(x_2 | x_1) - W(x_2 | x_3)T_\tau(x_3 | x_1)\} dx_2 \tag{10}$$

which is a differential form of the Chapman-Kolmogorov equation, known as the master equation. For a discrete Markov chain with index n, this equation reduces to

$$\frac{\partial P(x, t)}{\partial t} = \int \{W(x | x')P(x', t) - W(x' | x)P(x, t)\} dx' \tag{11}$$

If the range of $X = (x_1; x_2; \cdots; x_n)$ is a discrete set of states, the equation reduces to

$$\frac{\partial p_n(t)}{\partial t} = \sum_n \{W_{nn'} p_{n'}(t) - W_{n'n} p_n(t)\} \tag{12}$$

which represents a rate equation (gain-loss equation) for the probabilities of the separate states n. The first term is the gain of state n due to transitions from other

states n', and the second term is the loss due to transitions from n into other states. Remember that $W_{nn'} \geq 0$ when $n \neq n'$ and that the term with $n = n'$ does not contribute to the sum (Kampen 1992).

Equation 12 can be written in a more compact form when the following matrix $\mathbb{W}$ is defined

$$\mathbb{W}_{nn'} = W_{nn'} - \delta_{nn'} \left(\sum_{n''} W_{n''n} \right) \tag{13}$$

Then Eq. 12 can be written

$$\dot{p}_n(t) = \sum_{n'} \mathbb{W}_{nn'} p_{n'}(t) \tag{14}$$

Formally the solution of Eq. 14 with given initial $p_n(0)$ may be written

$$p_n(t) = \exp\{t \cdot \mathbb{W}_{nn'}\} p_n(0) \tag{15}$$

$\mathbb{W}_{nn'}$ needs to follow

$$\mathbb{W}_{nn'} \geq 0 \quad \text{for n} \neq \text{n'}; \tag{16}$$

$$\sum_n \mathbb{W}_{nn'} = 0 \quad \text{for each n'.} \tag{17}$$

Equation 16 shows the fact that the probability of staying in state n' is simply 1 minus the probability of escaping. For a complete reference on the properties of these matrices, see, e.g., Kampen (1992).

A fundamental property of the master equation is that as $t \rightarrow \infty$, all solutions tend to the stationary solution. This statement is strictly true only for a finite number of discrete states. For an infinite number of states, there are exceptions, such as the random walk.

3   The Kinetic Monte Carlo Algorithm

In any particular case, the master equation is fairly easy to write; however, solving it is quite another matter. The number of problems for which the master equation can be solved analytically is very limited. In addition, the master equation does not readily lend itself to numerical solutions due to the number and nature of its independent variables.

In 1976 Daniel T. Gillespie published his seminal paper *A General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical Reactions* (Gillespie 1976). In this work, Gillespie developed the main theory behind what is nowadays known as the kinetic Monte Carlo algorithm. We will follow his

derivation here since it is concise and very clear. In Gillespie's words: "The method is a systematic, computer-oriented procedure in which rigorously derived Monte Carlo techniques are employed to *numerically simulate* the very Markov process that the Master Equation describes analytically." That is, the method circumvents the need to solve analytically or numerically the master equation and gives what is usually called "one realization of the master equation" since one dynamic trajectory in phase space is obtained from it. If many trajectories are simulated, the obtained probability distribution complies with the master equation, and thus, in essence, the algorithm is a method to numerically solve the master equation that we have described in previous sections.

Gillespie developed this method to study chemical reactions. In general, the method is applicable to every process that can be modeled with the master equation, and hence, the constraints imposed to its derivation apply to the problems that can be solved with this algorithm; mainly, the process has to be Markovian (remember that every deterministic process is Markovian) and first order in time.

We want to solve the dynamic evolution of a system characterized by a set of states in phase space α, β, γ, etc. As mentioned above, it is the researcher's task to define phase space so to comply with the necessary constraints and, therefore, to be able to solve the problem at hand using the kinetic Monte Carlo methodology. The fundamental hypothesis of the stochastic formulation, and the only assumption to be made by the computational method, is that the probability that a particular transition occurs in the next time interval δt is equal to $c_{\mu\nu}\delta t + o(\delta t)$, where $o(\delta t)$ denotes unspecified terms which satisfy $o(\delta t)/\delta t \to 0$ as $\delta t \to 0$. Therefore

$$c_{\mu\nu}\delta t \equiv \text{average probability, to first order in } \delta t,$$

$$\text{that a particular transition } (\mu \to \nu) \text{ occurs.} \tag{18}$$

We want to solve the probability for the system to be in state μ at time t, $\forall t$, and $P(\mu; t)$, knowing the initial probability distribution and the transition probabilities per unit time $c_{\mu\nu}$. It is worth mentioning that the master equation can be derived from Eq. 18 by using the addition and multiplication laws of probability theory to write $P(\mu; t+dt)$ as the sum of the probabilities of the different ways for the system to arrive at state μ at time $t + dt$:

$$P(\mu, t + dt) = P(\mu, t)\left[1 - \sum_{\mu}^{M} c_{\mu\nu}dt\right] + \sum_{\nu}^{M} c_{\nu\mu}P(\nu; t)dt \tag{19}$$

and therefore

$$\frac{\partial P(\mu, t)}{\partial t} = \sum_{\nu}^{M} c_{\nu\mu}P(\nu; t) - \sum_{\mu}^{M} c_{\mu\nu}P(\mu, t) \tag{20}$$

which is the master equation, with M as the total number of states.

The principal theoretical construct upon which the procedure is based is not the grand probability function $P(\mu, t)$, nor any of its derived quantities, but rather an entity which is called the *reaction probability density function*, $P(\tau, \mu\nu)$, which is defined as

$$P(\tau, \mu\nu)d\tau \equiv \text{probability at time t that the next transition}$$

$$\text{in volume V will occur in the differential}$$

$$\text{time interval } (t + \tau, t + \tau + d\tau)$$

$$\text{and will be transition } \mu \rightarrow \nu. \tag{21}$$

$P(\tau, \mu\nu)$ is a joint probability density function on the space of the continuous variable $\tau (0 \leq \tau < \infty)$ and the discrete variable $\mu\nu(\mu\nu = \alpha \rightarrow \beta, \alpha \rightarrow \gamma, \ldots, \beta \rightarrow \alpha, \ldots)$. This joint probability is the product of $P_0(\tau)$, the probability at time t that no transition will occur in the time interval $(t, t + \tau)$, times $c_{\mu\nu}d\tau$, and the subsequent probability that transition $\mu \rightarrow \nu$ will occur in the next differential time interval $(t + \tau, t + \tau + d\tau)$:

$$P(\tau, \mu\nu)d\tau = P_0(\tau) \cdot c_{\mu\nu}d\tau \tag{22}$$

Note that we need not worry about more than one reaction occurring in $(t + \tau, t + \tau + d\tau)$, since the probability for this to happen is $o(d\tau)$. $P_0(\tau)$ can be obtained if the interval $(t, t + \tau)$ is divided into K subintervals of equal length $\epsilon = \tau/K$. The probability if nothing happened in the first subinterval ϵ is

$$\prod_{\mu\nu}^{M} \left[1 - c_{\mu\nu}\epsilon + o(\epsilon) \right] = 1 - \sum_{\mu\nu}^{M} c_{\mu\nu}\epsilon + o(\epsilon) \tag{23}$$

For the whole interval, we can then write

$$P_0(\tau) = \left[1 - \sum_{\mu\nu}^{M} c_{\mu\nu}\epsilon + o(\epsilon) \right]^K = \left[1 - \sum_{\mu\nu}^{M} c_{\mu\nu}\tau/K + o(K^{-1}) \right]^K \tag{24}$$

In the limit of infinitely large K

$$P_0(\tau) = \exp\left[-\sum_{\mu\nu}^{M} c_{\mu\nu}\tau \right] \tag{25}$$

Substituting in Eq. 22, we obtain

$$P(\tau, \mu\nu) = c_{\mu\nu} \exp\left[-\sum_{\mu\nu}^{M} c_{\mu\nu}\tau \right] \tag{26}$$

The main goal of the KMC algorithm is to generate random points in phase space according to such probability density function. Three main methods have been developed to do so: (i) the direct method, (ii) the first reaction method (also called the first passage method), and (iii) the null event method.

3.1   The Direct Method

This method is based on the fact that any two-variable probability density function can be written as the product of two one-variable probability density functions. As such we can write $P(\tau, \mu\nu)$ as

$$P(\tau, \mu\nu) = P_1(\tau) \cdot P_2(\mu\nu|\tau) \tag{27}$$

where $P_1(\tau)d\tau$ is the probability that the next reaction takes place within the interval $[t + \tau, t + \tau + d\tau]$, independent of which reaction it might be. $P_2(\mu\nu|\tau)$ is the conditional probability that next event is $\mu\nu$ given that it occurs at time $t + \tau$. $P_1(\tau)d\tau$ is obtained summing $P(\tau, \mu\nu)d\tau$ over all possible events $\mu\nu$:

$$P_1(\tau) = \sum_{\mu\nu}^{M} P(\tau, \mu\nu) \tag{28}$$

Substituting this expression into Eq. 27 and solving for $P_2(\mu\nu|\tau)$

$$P_2(\mu\nu|\tau) = \frac{P(\tau, \mu\nu)}{\sum_{\mu\nu}^{M} P(\tau, \mu\nu)} \tag{29}$$

Substituting now the expression for $P(\tau, \mu\nu)$ given in Eq. 26 and defining $R = \sum_{\mu\nu}^{M} c_{\mu\nu}$, we obtain

$$P_1(\tau) = R \exp(-R\tau) \qquad (0 \leq \tau < \infty), \tag{30}$$

$$P_2(\mu\nu|\tau) = \frac{c_{\mu\nu}}{R} \qquad (\mu\nu = 1, 2, \ldots, M), \tag{31}$$

and we can now generate a random τ according to $P_1(\tau)$ and a random $c_{\mu\nu}$ from $P_2(\mu\nu|\tau)$ with the resulting pair $(\tau, c_{\mu\nu})$ distributed according to $P(\tau, \mu\nu)$. τ can be obtained drawing a uniformly distributed random number (ξ_1) in the interval $[0, 1)$ from the expression

$$\tau = \frac{1}{R} \ln \frac{1}{\xi_1} \tag{32}$$

while $c_{\mu\nu}$ can be obtained drawing a second random number (ξ_1) again uniformly distributed in [0, 1) and taking the event that satisfies

$$\sum_{\mu\nu}^{m-1} c_{\mu\nu} < \xi_2 \leq \sum_{\mu\nu}^{m} c_{\mu\nu} \tag{33}$$

3.2   The First Reaction Method

This method is based on the fact that the probability that at time t the event $\mu\nu$ occurs in the interval $(t + \tau, t + \tau + d\tau)$ irrespectively of all other possible events is given by

$$P_{\mu\nu}(\tau)d\tau = \exp\left(-c_{\mu\nu}\tau\right) \cdot c_{\mu\nu}d\tau \tag{34}$$

Therefore, independent times for each event might be drawn from such probability distribution using the same expression as before:

$$\tau_{\mu\nu} = \frac{1}{c_{\mu\nu}} \ln \frac{1}{\xi_{\mu\nu}} \tag{35}$$

where $\xi_{\mu\nu}$ is a random number to be drawn uniformly for each possible event. The actual next event is the one with the shortest time. This scheme generates a pair $(\mu\nu, \tau)$ that complies with the $P(\tau, \mu\nu)$ probability density function. Because the amount of random number needed is much larger than in the previous method, this algorithm is usually less efficient. However, recent implementations take advantage of spatial decompositions to accelerate the wall time for the system to evolve in time (Opplestrup et al. 2006).

3.3   The Null Event Method

algorithm is based on the introduction of a new type of event, usually denoted as a null event, which if chosen keeps the microstructure in its original state. This extra event can be added to each individual event such that all events have the same total rate (Gilmer and Bennema 1972) or can be added to a cumulative rate of a set of events (Hanusse and Blanche 1981). It can be shown that this algorithm solves the same master equation as the previous two methods and therefore can be used to study the microstructural evolution, that is, the algorithm follows the right dynamics. The advantage of this method is that the searching wall time to find the event is reduced considerably, even to zero, i.e., the CPU time scales as $O(1)$. However, the simulated time step decreases, since it is sampled from the same probability distribution with the extra rates due to the null events, $\tau = \frac{1}{R_{\text{null}}} \ln \frac{1}{\xi_1}$, with R_{null} the aggregate of all rates. This implies that the best computational efficiency will occur
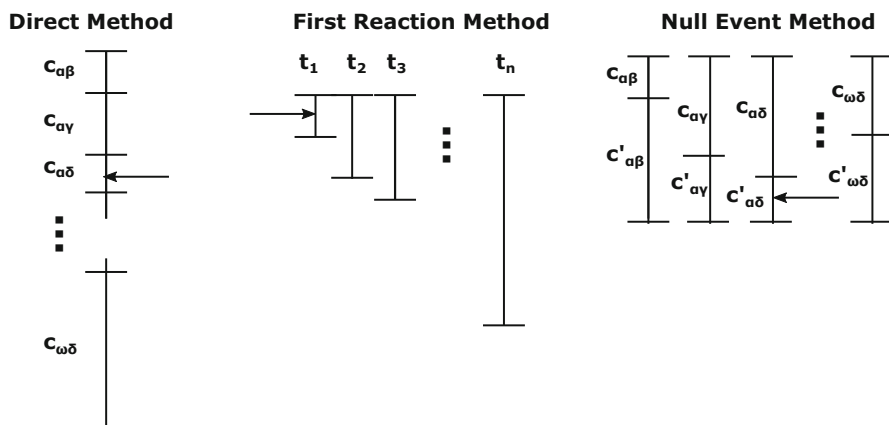
Fig. 1 Sampling approaches

for a minimum in the total null event rate (Chatterjee and Vlachos 2007). If the searching step in the direct method represents a large bottleneck in computational speed and the system is not significantly stiff (i.e., the total null event rate is not daunting), this algorithm can outperform the direct method.

Figure 1 shows schematics of the three sampling methods described above.

4 Searching and Updating Algorithms

Provided the rates are trivially calculated or are tabulated in advance, sampling Eq. 33 can represent the main computational bottleneck in KMC simulations. In general, the calculation of the rates is far from trivial, and several algorithms have been recently developed to calculate more accurately rates *on-the-fly* within a KMC scheme. This set of methods are usually denoted as adaptive KMC algorithms (Henkelman and Jonsson 2001; El-Mellouhi et al. 2008; Xu and Henkelman 2008). However, these methods are computationally demanding, and the total simulated time and the sample size remain limited. On the other hand, they are usually able (like MD or AMD) of obtaining the reaction mechanism without extra assumptions. In the context of irradiation to analyze the long-term microstructural evolution of the system under irradiation, the rates for the different events are usually pre-computed, and under these conditions, it is important to identify the possible different searching and updating algorithms that reduce the wall time in sampling Eq. 33 and updating the rates and its aggregate.

4.1 Linear Search

The linear search is the simplest algorithm to implement. It uses an array or list of possible events with an assigned rate that visits linearly, i.e., one by one in the list,

to calculate the cumulative rate, until Eq. 33 is satisfied. This method scales linearly with the number of events, i.e., $O(N)$ with N the total number of events. Grouping can be used to improve the efficiency in what has been denoted as n-level linear algorithms. In this case a set of events are grouped together and the cumulative rate computed in as many levels as wanted. Each search goes in general through a number of events that is $N_n \approx N^{1/n}$, with n the level of grouping. The wall time of this algorithms scales then as $nO(N_n)$, which can be considerably faster than the brute force linear search.

4.2   Binary Search

Binary searches rely on the generation of a tree data structure containing information about the rates (Gibson et al. 2000; Schulze 2002) to reduce the CPU time required to sample Eq. 33. In this structure, each node represents a cumulative propensity of the branches below it, such that the bottom nodes represent individual propensities for specific events. The number of levels in a tree will be at least $N_{level} = floor(\log_2(N))$. If this is the case, the tree is considered equilibrated, and the computational efficiency is maximum. The algorithm starts at the head node that contains the cumulative rate of the left branch. If the target rate is lower, the next search is performed in the children to the left of the head node, to the right otherwise. This process is repeated until the bottom of the tree is reached and the event identified. In this algorithm the CPU time scales as $O(\log_2(N))$; however, the pre-factor to keep the tree equilibrated might be significant, and, not in every situation, this algorithm is faster than the linear searches.

4.3   Constant Time Search

The main idea of this method is based on the null event (or rejection) algorithm, in which the need for searching is reduced to drawing an integer random number from a uniform distribution between $[1, N]$, with N the total number of possible events, which will directly point at the event to be performed, including a null event. A second random number must be generated to pick between the actual or the null events. The search does not depend on the number of events, N, and therefore scales as $O(1)$. In a stiff problem with significantly different rates, a grouping algorithm can be developed to minimize the amount of null events. This will resemble a combination between the n-level search with the extra addition of the null events. In this case the search would be $O(N_n)$, with N_n the number of groups. This null event (or padding) method could also be combined with a binary search to keep the tree balanced. A review of this method can be found in Slepoy et al. (2008).

Figure 2 shows a summary of the searching algorithms discussed above.
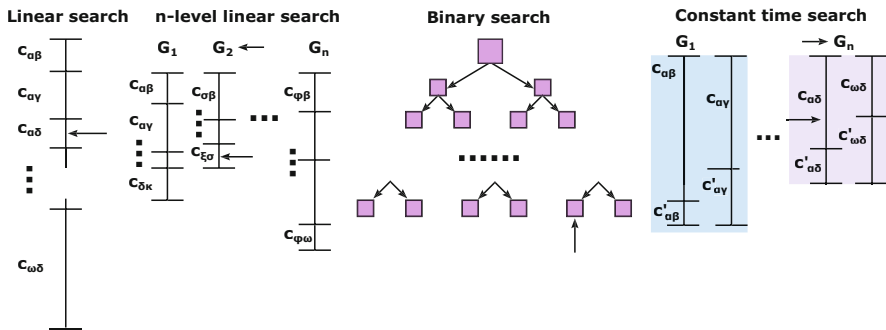
Fig. 2 Searching algorithms

5   Object Kinetic Monte Carlo Methods

The object kinetic Monte Carlo (OKMC) method uses a coarse-grained representation of the system to accelerate the wall time required to analyze its time evolution. In the context of materials under irradiation, the method averages out the underlying atomic structure to consider as possible event processes involving defects. Therefore, the algorithm follows the irradiation-created defects in time, solving for their migration and the interaction with one another and other kinds of features in the microstructure, such as grain boundaries, dislocations, alloying elements, or such. As it was mentioned in Sect. 1, TST is usually assumed to calculate the rates for each possible event, although, in general, experimental results or MD simulations can be also used to obtain such propensities, including dynamical corrections. The method results in a defect distribution evolving in time for a given initial microstructure (may be containing defect sinks such as interfaces or dislocations) under a given set of irradiation conditions, such as dose rate, temperature, pressure, and information about the nature of the defects created in the collision processes.

6   Density Functional Theory to Calculate Propensities

Density functional theory (DFT) is a mean field approach to solve the many-body Schrodinger equation (Hohenberg and Kohn 1964). It is based on the fact that the ground-state properties of a many-electron system are uniquely determined by an electron density that depends on only three spatial coordinates. As such, the many-body problem of M electrons with $3M$ spatial degrees of freedom is reduced to three spatial coordinates through the use of functionals of the electron density. DFT assumes that the electron dynamics are much faster than the nuclei and solves the ground-state electronic structure for a fixed positions of the nuclei, i.e., the

Born-Oppenheimer approximation is assumed to hold. Hohenberg and Kohn (1964) developed an energy functional for the system and demonstrated that the true ground state of such energy functional minimizes the functional. Subsequently, Kohn and Sham (1965) were able to reduce the many-body problem of interacting electrons in a static external potential to a solvable problem of noninteracting electrons evolving in an effective potential. The correctness of these methods mainly relies on the accuracy of this effective potential. This effective potential includes the external potential and the Coulomb interactions between electrons, with contributions from the exchange and correlation interactions. The simplest approach is the local-density approximation (LDA) that relies on the exact exchange energy for a uniform electron gas as given by the Thomas-Fermi model (Ceperley et al. 1980). More complex functionals for exchange and correlation have been developed and still are a very active field of research. One widely used example is the generalized gradient approximation (GGA) that incorporates information about the gradient of the electron density in the functionals to account for the non-homogeneity of the real electron density (Langreth and Mehl 1983; Becke 1988; Perdew et al. 1992). Hybrid functionals incorporating a portion of the exact exchange energy calculated from Hartree-Fock theory with the rest of exchange correlation from other sources are also extensively used. The problem of finding the ground-state electron density is variational and has to be solved iteratively in a self-consistent manner such that the effective potential and the wave functions are also obtained for such electron density.

DFT is widely used in chemistry and materials science to understand and predict the properties of complex systems under specific conditions. Although computationally more efficient than all-electron approaches, the solution of the DFT equations presents a computational bottleneck, which translates into limited system sizes and simulated times. In general, the computational scaling of this method is $O(n^3)$ with n the number of electrons since usually the diagonalization of the system Hamiltonian scales as n^3. The development of $O(n)$ methods is also a remarkably active research area (Niklasson et al. 2006; Witt et al. 2018).

In the context of materials under irradiation, this approach has been broadly used to estimate the properties of defects. Although usually calculated at 0 K, there have been some studies using ab initio MD to compute the required properties (Grabowski et al. 2007; Körmann et al. 2008; Grabowski et al. 2011; Glensk et al. 2015) at higher temperature. One of such properties is the defect formation energy. This is usually calculated as

$$E_{\text{for}}(D) = E[N_{\text{at}} + D] - (N_{\text{at}} + D)\mu \qquad (36)$$

where D represents the defect, N_{at} is the number of atoms in the bulk system, and μ is the cohesive energy of the material. This formation energy might be used to estimate the equilibrium defect populations.

Another important quantity is the migration energy ($E_m(D)$), which is defined as the difference of the energy of the system at the saddle point and the initial state

for a defect hopping. To compute this migration energy, the system is forced to converge to the minimum energy path (MEP) between the initial and final states. If the initial and final states are known, methods like the drag method (Trushin 2005) or the nudged elastic band (NEB) (Henkelman and Jonsson 2000; Henkelman et al. 2000) are commonly used. If the final state is unknown a priori, the dimer method (Henkelman and Jonsson 1999) or the Lanczos method (El-Mellouhi et al. 2008) are usually preferred. Recently, advanced methods have been developed to improve the predictions obtained with the traditional methods, such as free-end NEB (Zhu et al. 2007) or free-end adaptive NEB (Zhang et al. 2016). This $E_m(D)$ will be used in the HTST expression (Eq. 2) to compute the migration rate. The pre-factor is usually assumed constant and close to the Debye frequency of the material. However, under certain conditions, it has been shown that the explicit calculation of the pre-factor might significantly change the event rate. One of such conditions arises when the system is under nonuniform strain, in which case the pre-factor can change by orders of magnitude (Uberuaga et al. 2007). Vineyard pre-exponential factors can also be calculated within DFT, although the computational expense is high.

Binding energies are also crucial for the method to be predictive. The binding energy characterizes the interaction between two or more species in the system. It is defined as the energy gain in a given reactions that can be exothermic or endothermic:

$$E^b_{X-Y}(A) = E[S + X + Y] + E[S] - (E[S + X] + E[S + Y]) \tag{37}$$

where $E^b_{X-Y}(A)$ denotes the binding energy of species X and Y in system S, $E[S + X + Y]$ is the energy of the system with the interacting species, $E[S]$ is the reference system, and $E[S + X]$ and $E[S + Y]$ are the reference systems with specie X and Y, respectively. Important binding energies are, for example, from single defects to clusters, between defects and impurities or alloying elements, and between defects and sinks (such as grain boundaries or dislocations). This binding energy enters in the activation energy for a dissociation process, altering the propensity for such reaction to occur. Expressions derived from elasticity theory are commonly used to fit the values obtained from DFT and extrapolate for larger cluster sizes (Perini et al. 1984).

The interaction radius is also an important parameter needed in the OKMC method. In general, it depends on the species and its size. Usually, empirical relations based on DFT data are used. For defect clusters, a common expression is $r_n = Z \cdot \left(\frac{3n\Omega}{4\pi}\right)^{1/3} + r_0$, where Z is a bias factor depending on the nature of the defect, n is the number of single defects in the cluster, Ω is the atomic volume, and r_0 is a constant usually obtained with DFT.

DFT has also been used to obtain threshold displacement energies in materials of interest, i.e., the minimum energy that needs to be given to an atom such that a stable Frenkel pair is formed. This approach has been employed to study the energetic threshold depending on the orientation in materials such as α-Fe (Olsson et al. 2016) and W (De Backer et al. 2016).

7   Examples

In this section we will show a couple of examples in which this method has been applied to analyze the evolution of the microstructure under irradiation. The first one is for ferritic/martensitic steels, which is of great importance in nuclear applications. The second example studies the defect evolution in tungsten (W) and the interaction of the irradiation-created defects with He created from (n, α) nuclear reactions and the incoming flux from the plasma ashes.

7.1   Microstructural Evolution in α-Fe Under Irradiation

Irradiation of a metal and subsequent annealing while measuring electrical resistivity can provide important information about basic processes related to defect migration and defect stability, data that can be calculated with DFT or other atomic-scale models as mentioned in the previous section. C. C. Fu et al. (2005) using event kinetic Monte Carlo with parameters from DFT simulated the resistivity recovery curve of Fe after electron irradiation and correlated these results to the different peaks observed in experiments performed by Takaki et al. (1983). Figure 3 shows the derivative of the number of defects (N) as a function of temperature obtained in this case with OKMC and the same parameters as in Fu et al. (2005) for a dose of $2 \times 10^{-6} dpa$ (dpa = displacements per atom). The locations of the peaks found experimentally in Takaki et al. (1983) are marked in the same figure with arrows. These simulations show that the first peak (named stage ID2) appears because of the recombination of correlated Frenkel pairs. In this case the vacancy and self-interstitial belong to the same Frenkel pair in such a way that with one single self-interstitial jump, these defects recombine. Recombination of defects gives rise to a decrease in the resistivity of the material and therefore a peak in the derivative of the curve. The second stage, named IE, also corresponds to the recombination between vacancies and self-interstitials but for pairs of defects that are far away from each other, so more than one jump of the self-interstitial must occur for the defects to recombine. This peak appears at a temperature that corresponds to that of self-interstitial migration. The next stage, stage II, is due to the migration of self-interstitial clusters, while stage III is due to vacancy migration.

In the case of an alloy, if the concentration is low, the alloying element can be treated explicitly in the OKMC approach by treating every alloy atom as a defect. As mentioned above, in the OKMC approach, there is no underlying lattice, so, in principle, these atoms will be distributed randomly within the simulation cell. All the possible events that these atoms can perform, interaction with vacancies or self-interstitials, formation of clusters, or migration must be specified with their appropriate rates. For the case of FeCr alloys, there is information obtained from DFT on small clusters (Olsson et al. 2007; Olsson 2009). One important defect formed in FeCr alloys is the mixed interstitial dumbbell, which has a lower migration energy than the self-interstitial cluster, but also a low binding energy
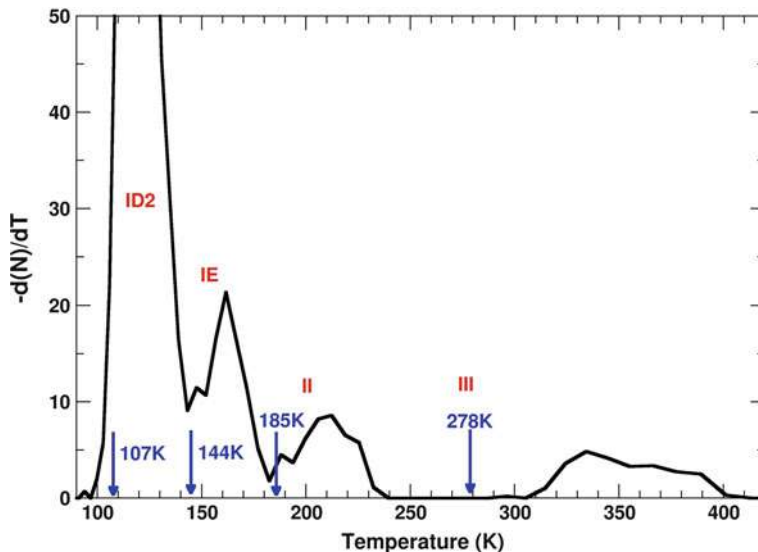
Fig. 3 Derivative of the number of defects as a function of temperature for electron irradiation of Fe at a dose of $2 \times 10^{-6} dpa$ from OKMC calculations and parameters from Fu et al. (2005)

(Domain et al. 2004). Experimentally, it is observed that in low-concentration alloys (between 0.01 and 0.1% Cr), the IE peak in the resistivity recovery measurements shifts toward lower temperatures and the shift increases with Cr concentration (Abe and Kuramoto 1999; Maury et al. 1987). Using an OKMC model with parameters from DFT as described in Gámez et al. (2011) that include the formation of mixed dumbbells (ICr) as well as other Cr clusters (I2Cr, ICr2, and I2Cr2), the shift was reproduced as it can be seen in Fig. 4. This figure shows the results for pure Fe and for FeCr at the three concentrations measured experimentally in Abe and Kuramoto (1999), 0.019 at%, 0.047 at%, and 0.095 at%), and how the IE peak moves toward lower temperatures with increasing Cr concentration in agreement with the experiments. The reason behind this shift is in fact the formation of the ICr mixed dumbbell.

One of the advantages of the OKMC approach is that system sizes larger than those in lattice KMC (LKMC) models can be simulated, typically cubes of hundreds of nanometers on each direction. This allows for calculations of damage accumulation as a function of irradiation to doses that can be achieved experimentally and to model the formation of clusters large enough to be visible under transmission electron microscopy (TEM). An example of these types of simulations that have been performed by several groups for different conditions, starting with the work of Heinisch (1990) (see, e.g., Domain et al. 2004; Malerba et al. 2008; Becquart and Domain 2010), is presented here. In this case, the accumulation of damage during continuous irradiation of Fe and Cu is simulated. Details of these calculations are
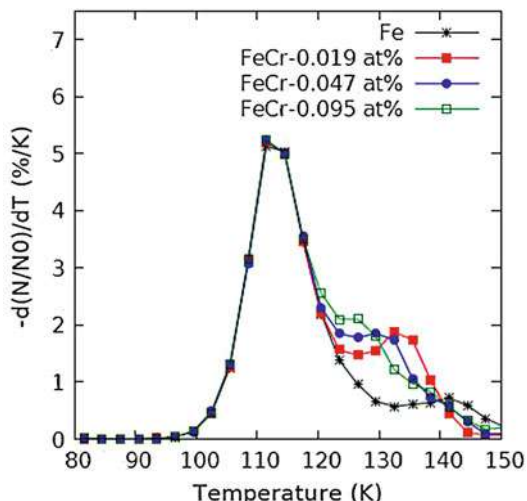
in Caturla et al. (2000). In brief, a database of collision cascades in Cu and Fe for 30 keV recoils obtained from molecular dynamics simulations is used as initial damage distribution. This picosecond time damage is very different in Fe and in Cu. Cu presents mostly clustered vacancies with surrounding self-interstitial clusters, while in Fe most of the vacancies are isolated or forming very small clusters. The values of migration energies and binding energies of defects used in the calculations were obtained from empirical potentials. Figure 5 shows the total concentration of visible clusters as a function of dose for the case of the two materials studied, Cu and Fe. Here, the visibility conditions were considered as follows: 20 defects for vacancies in Cu, equivalent to a stacking fault tetrahedra of 1.5 nm, 350 vacancies for a 1 nm void in Fe, and 50 self-interstitials for a 1 nm loop in Fe. This figure shows that there is an important difference in the damage accumulated in Fe vs. Cu, phenomenon also observed experimentally (Dai et al. 1997). The reason for this difference is related both to the initial damage distribution and the effect of impurities in these two materials. In Cu, where no impurities are considered in the simulation, most self-interstitials and self-interstitial clusters disappear through recombination with vacancies or at grain boundaries, and most of the damage is vacancy clusters, with an average size of about 28 vacancies which corresponds with a 2 nm stacking fault, in agreement with experimental observations (Zinkle and Farrell 1989). In Fe, self-interstitials are trapped by impurities, which result in visible clusters, while vacancies are too small to be observed under TEM (with less than 15 defects). In Fe most of the evolution of the clusters and growth to sizes that can be observed experimentally is going to occur through defect diffusion and defect coalescence. And the description of these interactions depends on the set of reactions that the user of the OKMC model has decided to consider as the important ones.
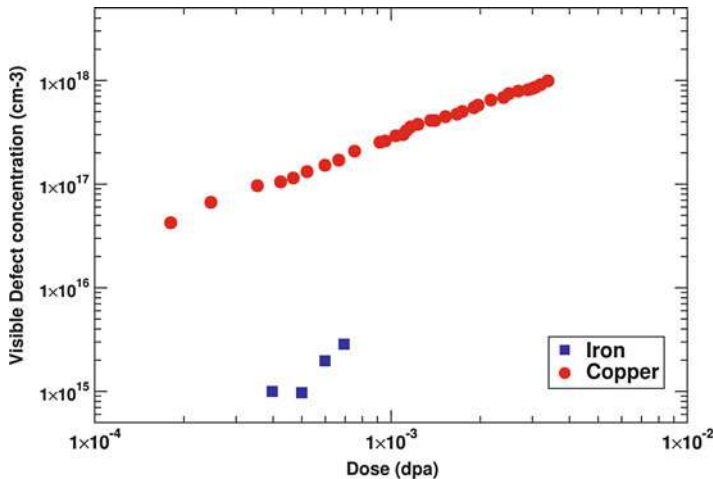
Fig. 5 Visible defect clusters as a function of irradiation dose for the case of Fe and Cu (see Caturla et al. 2000 for details)

7.2   Microstructural Evolution in W Under Fusion Conditions

Critical components in magnetic fusion reactors are required to have stringent properties including low neutron activation, high melting point, good thermomechanical properties, low sputter erosion, and low tritium retention/co-deposition. Tungsten (W) is the leading plasma-facing material (PFM) candidate due to its high melting temperature, low erosion rates, and small tritium retention. These advantages are unfortunately coupled with very low fracture toughness characterized by brittle transgranular and intergranular failure regimes, which severely restrict the useful operating temperature window and also create a range of fabrication difficulties. Exposure to He and H plasma leads to the formation of diverse microstructures (pits, holes, bubbles, nanostructured surface morphology termed as fuzz) (Baldwin and Doerner 2008) that severely limit its utilization. Therefore, understanding the mechanisms leading to the W failure is crucial to develop materials to be able to withstand the extreme environments posed by fusion energy generation.

DFT has been broadly used to obtain the data mentioned above and needed by the OKMC to evolve the microstructure. Using different DFT approaches, the formation energies of intrinsic defects have been calculated. The single vacancy (V) values in the literature range from ~3.1 to ~3.7 eV (Söderlind et al. 2000; Becquart and Domain 2007; Suzudo et al. 2014), while for self-interstitial atoms (SIA), the values depend on the configuration. Some discrepancy arises in the literature about the most stable configuration, with similar values for the $\langle 111 \rangle$ dumbbell, $\langle 111 \rangle$ crowdion, and a slightly rotated configuration, with values around ~9.6–9.7 eV (Nguyen-Manh and Dudarev 2006; Suzudo et al. 2014) and ~10.41 (Becquart and Domain 2007) depending on the approach. Binding energies of individual defects

Fig. 6 Binding energies of individual defects to their clusters for (**a**) vacancies, (**b**) self-interstitial atoms, and (**c**) He atoms (Becquart and Domain 2010)
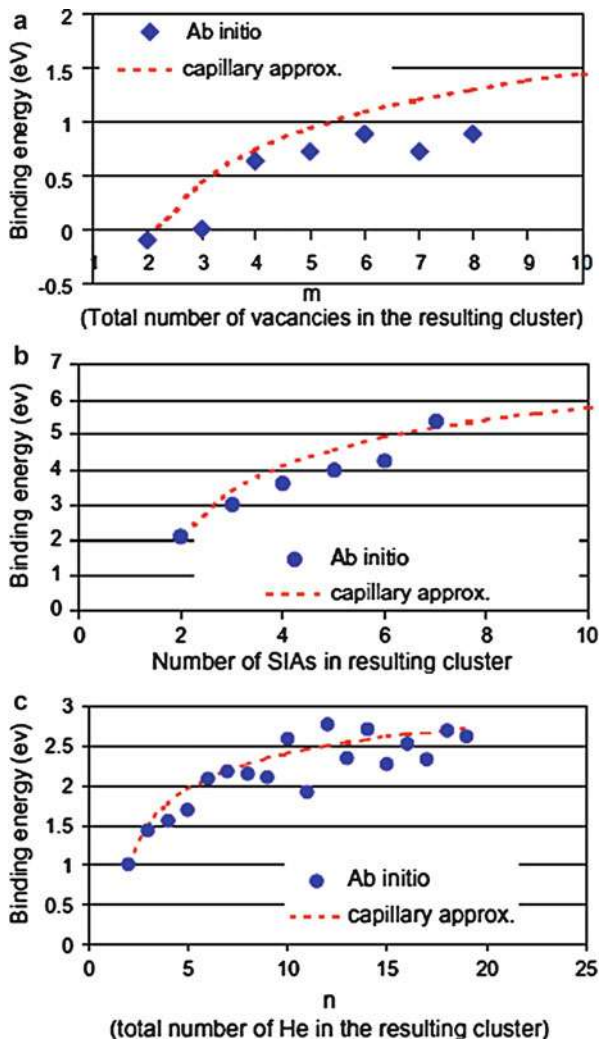
to defect clusters are also required. Becquart et al. calculated the ones for vacancies and interstitials using their DFT approach and use a capillary law for large defect clusters (>8 V for vacancy clusters and >7 SIA for interstitial clusters). Figure 6a and 6b shows the results for vacancies and self-interstitials, respectively. We observe that the capillary approximation fits well the ab initio data. One of the puzzling results in W is that the binding energy of two individual vacancies is repulsive, i.e., vacancies do not want to aggregate.

The migration energies for the intrinsic defects and their clusters are also needed. For individual defects, several sources are available in the literature (Becquart and Domain 2010; Suzudo et al. 2014; Huang et al. 2016), all describing single

vacancies with a large migration energy ($\sim$1.7 eV), while SIAs show a remarkably small migration barrier ($\sim$0.005–0.056) eV. Becquart et al. (2010) proposed an expression for the diffusion constant of defect clusters. In the case of vacancy clusters, the attempt frequency is given by $\nu = \nu_0(q^{-1})^{n-1}$, with $\nu_0 = 6 \times 10^{12}$ s^{-1} constant and equal for V and SIA, $q = 1000$, and n the number of individual defects in the cluster. Based on this model, Castin et al. (2017) have recently proposed an extension to account for the effect of grain boundaries and carbon impurities. The authors presented a sensitivity analysis of the parameters that mostly influence the onset of void swelling using their in-house OKMC code called MATEO concluding that the presence of carbon impurities in the W matrix strongly influences the kinetics of void formation due to the strong binding interaction between carbon and vacancies. They also showed the importance of correlated recombination after cascade events in the microstructure evolution. Mason et al. (2014) used a different parametrization of formation and migration of clusters to study loop accumulation and their interactions in a W thin film. In this case, they explicitly added elastic interactions between defects to analyze the evolution of collision cascades and compare the effect of impurity trapping and elastic trapping concluding that elastic loop trapping within a cascade can stabilize large loops at high temperatures, while they shrink at low temperatures. On the other hand, small loops glide to the surface at high temperature, hence accounting for the experimentally observed reduction in total loop density and the change in the loop size distribution (Fig. 7).

As mentioned above, the study of the effect of He and H is paramount to understand the response of W to plasma exposures and neutron irradiations. OKMC has also been used to study the evolution of W in the presence of these two light elements. For this end, parameterizations of the behavior of these elements and their interactions among them and with V and SIA have been developed. Becquart et al. (2010) used their parametrization to study the isochronal annealing of He desorption in W and compared the results to experimental observations, employing the LAKIMOKA OKMC code (Domain et al. 2004). Subsequently, they studied the microstructural evolution of W under irradiation accounting for He production. They observe that the OKMC model reproduces satisfactorily the microstructure formed in the track region by 800 ^{3}He implantation for different fluence conditions (see Fig. 8).

The influence of grain boundaries on the microstructure evolution accounting for He has also been studied within an OKMC model. Extending the parametrization of Becquart et al. (Becquart and Domain 2009; Becquart et al. 2010), Valles et al. (2015a, b) studied the retention of He in the sample depending on the density of grain boundaries under pulsed He irradiation. They observe that around 50% of the incoming ions were retained at the interfaces in nanocrystalline W, while in the case of larger grains, that number dropped to around 30%, which indicates a strong dependence of He retention with grain boundary configurations. They also conclude that in the nanocrystalline W, mixed He$_n$V$_m$ clusters contain more vacancies with a lower He/V ratio.

Valles et al. (2017a) have also studied the influence of H atoms in the defect evolution under irradiation and the effect of grain boundaries in the sample. They
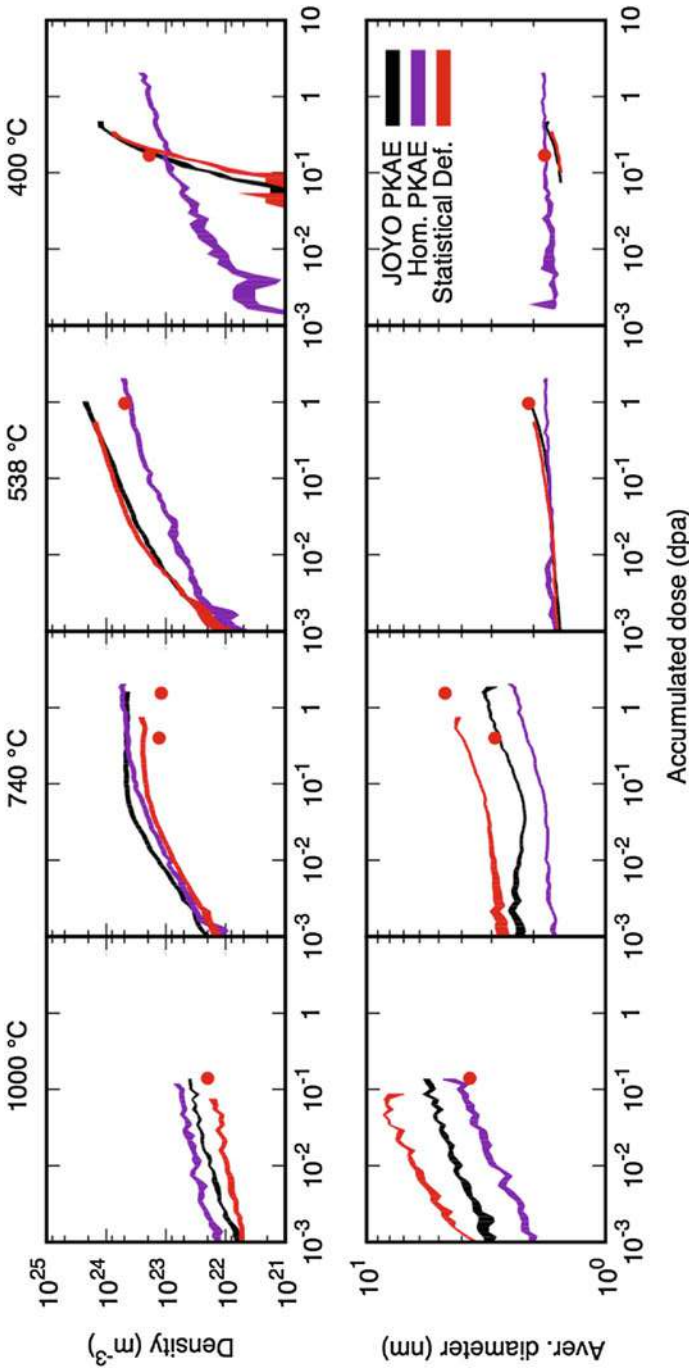
Fig. 7 Evolution with the accumulated dose of the void density and average void diameter found in OKMC simulations, as compared to experimental evidence (red dots) (Castin et al. 2017)
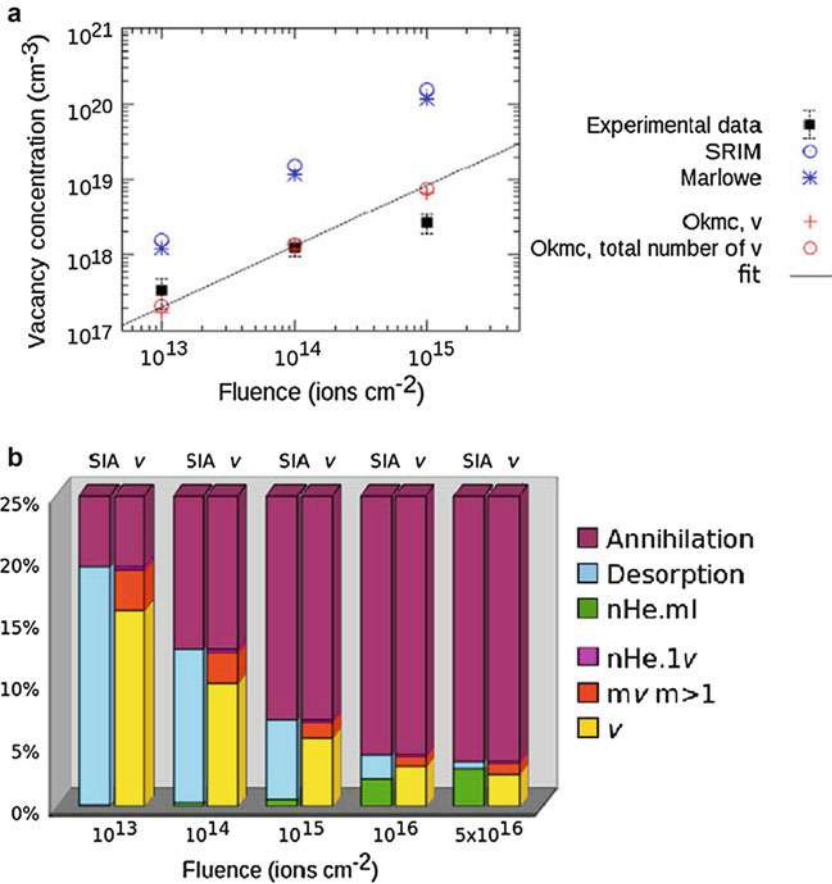
Fig. 8 (**a**) Vacancy concentration at the end of the implantation sequence for different fluences: a comparison of the experimental data with the SRIM, MARLOWE, or OKMC models. The black line is a linear fit (in the logarithmic scale) of the OKMC results. (**b**) The amount of desorption, annihilation, and clustering versus fluence at the end of the implantation stage as calculated by OKMC (Becquart et al. 2011)

extended the energetics for small clusters available in the literature (Henriksson et al. 2005; Fernandez et al. 2015; Lu et al. 2014; Ohsawa et al. 2010) to parametrize the H-V interactions using DFT calculations and considered H-SIA clusters as immobile. It is important to note that the H-H interaction is repulsive and, therefore, pure H clusters will not occur. However, in the presence of defects, the situation is different, with a strong binding between vacancies and H atoms. The authors conducted OKMC simulations using the MMonCa code (Martin-Bragado et al. 2013) and concluded that grain boundaries have a clear influence on the number density and distribution of vacancies, observing a larger concentration

in the nanocrystalline sample compared to a single-crystal sample. Also, they observed that H retention is highly influenced by both grain boundaries and vacancy concentration, the size of H_nV_m clusters varies slightly with the presence of grain boundaries, and grain boundaries act as fast diffusion pathways for H. Other KMC models on the same system reached similar conclusions (Oda et al. 2015).

Impurities, such as carbon and oxygen, or alloying elements or transmutation products will also influence the long-term damage evolution in W. DFT calculations have shown that both vacancies and He atoms strongly interact with solute atoms. Carbon and oxygen have also been found to be able to trap He atoms, increasing its retention in the material and the probability of forming bubbles. Self-interstitial atoms also interact with solute atoms and impurities (Suzudo et al. 2014; Gharaee et al. 2016; Huang et al. 2017, 2018). Some of these extra atoms modify the relative stability of SIA configurations, which in turn modifies the transport properties and the flux coupling between the defects and the solute elements. Very recently, Castin et al. (2018) have developed an OKMC model based on DFT calculations to study the role of carbon impurities in defect evolutions and also the presence of rhenium upon neutron irradiation, which also impacts the diffusivity of defects and therefore the microstructure evolution. Their model compares well with experiments in terms of both density of loops and voids.

Another material problem where the OKMC has been extensively applied was the analysis of the formation of a nanostructured surface morphology termed as fuzz appearing in W exposed to He plasma under conditions relevant for energy generation. The increased surface area and fragility of these nanostructured surfaces raise new concerns for the use of W as a fusion reactor PFM, particularly as a source of high-Z dust that will contaminate the plasma. Lasa et al. (2014) developed a model considering the He atoms and the bubbles as the objects in the system. He atoms are mobile but bubbles are not. Bubbles are considered to blast at a threshold distance from the surface that depends on the bubble radius. They concluded that the square root of time dependence of the growth rate is due to the He bubble growth and not to He diffusivity. The surface above the bubbles grows as interstitial loops are emitted from the bubbles with a preferential direction toward the surface. Bubble rupture near the surface will create roughness as seen in the experiments. Even though the model does not consider elastic interactions, nor surface diffusion or W knockout adatom formation, it reproduces the experimentally found growth dependence on time, which suggests that those mechanisms might not be crucial to explain fuzz formation. Building on this model, refining the parametrization, Valles et al. (2017b) studied the effect of temperature on fuzz growth. The authors identified the reason for fuzz formation as the formation of large He_nV_m clusters, which are only stable at intermediate temperatures (900–1900 K). At low temperatures (700 K), fuzz growth is prevented by the stability of small He_9V_1 clusters, hindering He emission. In contrast, at high temperatures (2500 K), fuzz does not grow due to the high He emission from He_nV_m clusters, leading to their dissolution and, thus, preventing He retention.

8   Limitations of the Method

The main limitation of the method is the fact that all possible mechanisms characterizing viable events have to be identified a priori and taken as input for the OKMC simulations. However, it is oftentimes the case that unforeseen mechanisms occur as the system evolves in time. One simple example often given in the literature is the adatom diffusion on a {100} free surface in fcc metals, which was long misunderstood until Feibelman showed that the exchange mechanism could be energetically favorable compared to direct migration (Feibelman 1990). Recent examples include the computational observation using MD that stacking fault tetrahedra (characteristic irradiation-created defects in fcc metals) might indeed migrate with diffusivities depending on their atomic structure, contrary to the commonly accepted assumption that described these defects as immobile (Martinez and Uberuaga 2015). It was shown that their mobility can significantly impact the long-term behavior of the system, precisely relying on an OKMC model as implemented in the code MMonCa (Martin-Bragado et al. 2013). Another interesting example is the understanding that small He clusters might indeed diffuse in a W matrix as Frenkel pairs are created and the self-interstitial annihilates with a different vacancy to the originally created, leading to net translation of the cluster. Using a cluster dynamics model as implemented in the code Xolotl (Xu et al. 2012), it was shown that the incorporation of this mechanism in the viable events significantly modifies the amount of He retained in the W wall and the bubble distribution. As the microstructure becomes more complex, the probability of finding new mechanisms increases, which might significantly modify the long-term evolution of the system (Uberuaga et al. 2018). Therefore, much attention needs to be given to understanding the atomistic processes occurring in the systems of interest under irradiation and their characterization with accurate rates for the OKMC calculations to be predictive.

It is also commonly assumed that the rates for the possible events are independent of the current configuration, i.e., there is no long-range interaction between defects that might modify their diffusivities. There are studies in the literature mitigating this limitation (Hudson et al. 2005; Wen et al. 2009; Subramanian et al. 2013; Mason et al. 2014). A common approach is to use a linear approximation in which the dipolar tensor of the defect is used to compute a variation of the energy at the saddle point and at the initial state due to elastic interactions (among defects and also between irradiation-created defects and pre-existing dislocations/grain boundaries (Vattré et al. 2016)). These modifications in the rates due to elastic interactions have been shown to modify the microstructural evolution of the system. A similar situation is encountered in the case of charged defects, in which long-range Coulomb interactions might be considered to bias defect diffusivities (Martin-Bragado et al. 2006). Although these long-range effects might be significant, they are not commonly considered based on computational arguments. Usually this problem scales as $O(N^2)$, with N the number of species, which reduces the efficiency of the algorithm and the attainable simulated time. This bottleneck could

be in part mitigated relying on fast multipole methods, which reduce the algorithmic complexity to $O(N)$ (Arsenlis et al. 2007), albeit with a large pre-factor, or Fourier transform methods ($n \log(n)$), with n the number of grid points (Hunter et al. 2011; Bertin et al. 2015), to accelerate the computation of the interactions.

In principle the method should be able to handle alloying elements and impurities provided that the whole parametrization is known. In practice this represents another challenge in terms of computation since the number of events can increase considerably. Also, the definition of local concentration and its relation with a certain cluster is unclear for non-dilute alloys. In this scenario, it is also important to consider the possibility of defect trapping, i.e., the presence of superbasins in which defects might flicker between nearby positions without much addition to the simulated time, reducing considerably the efficiency of the algorithm. Several methods have been proposed to alleviate this drawback that can indeed significantly increase the efficiency in such situations (Novotny 1995; Fichthorn and Lin 2013; Athènes and Bulatov 2014).

In common conditions in nuclear environments, the amount of irradiation-created defects per unit time and volume is considerable and increases with time until it reaches steady state (if attainable). Tracking all these defects in an efficient manner is a challenge. Parallel algorithms to take advantage of parallel architectures have been recently developed with the goal of improving both simulated time and system size (Shim and Amar 2005a, b; Martínez et al. 2008, 2011). Implementation on GPU architectures has also been considered with the same goal and showed promising results (Jiménez and Ortiz 2016).

9   Conclusions

We have reviewed in this chapter the kinetic Monte Carlo (KMC) methodology describing in detail the theory behind this powerful and flexible approach and how it can be used to study the microstructure evolution of systems under irradiation. The KMC approach described here is called object KMC (OKMC), which coarsegrains out the atomic structure to focus on elements (or objects) of interest, such as irradiation-created defects, impurities, transmutation or alloying elements, etc. These objects evolve in time according to prescribed rules characterized by probabilities per unit time (or rates), which enter in the OKMC algorithm to follow the system evolution. It is commonly assumed that the rates follow transition state theory, which relates activation free energies and attempt frequencies with such rates. We have seen how these parameters can be accurately calculated using density functional theory, which explicitly accounts for the electronic degrees of freedom. Although computationally expensive, which limits system size and total simulated time if dynamics are required, its accuracy, superior to that of empirical interatomic potentials makes it the method of choice to obtain the required values for single defects and small defect clusters, along with their interactions with impurities and alloying elements. We have shown two examples in which the evolution of the

system under irradiation has been studied using this methodology: ferritic steels and tungsten. In both systems we show that the results in terms of defect distributions can be satisfactorily compared to experimental observations. Finally, we have reviewed the limitations of the methodology and how current research approaches are tackling those limitations.

References

Abe H, Kuramoto E (1999) Interaction of solutes with irradiation-induced defects of electron-irradiated dilute iron alloys. J Nucl Mater 271–272:209–213

Arsenlis A, Cai W, Tang M, Rhee M, Oppelstrup T, Hommes G, Pierce TG, Bulatov VV (2007) Enabling strain hardening simulations with dislocation dynamics. Model Simul Mater Sci Eng 15:553–595

Athènes M, Bulatov VV (2014) Path factorization approach to stochastic simulations. Phys Rev Lett 113(23):230601

Baldwin M, Doerner R (2008) Helium induced nanoscopic morphology on tungsten under fusion relevant plasma conditions. Nucl Fusion 48(3):035001, 00193

Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic-behavior. Phys Rev A 38(6):3098–3100

Becquart CS, Domain C (2007) Ab initio calculations about intrinsic point defects and He in W. Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 255(1):23–26

Becquart CS, Domain C (2009) A density functional theory assessment of the clustering behaviour of He and H in tungsten. J Nucl Mater 386–388:109–111

Becquart CS, Domain C (2010) Modeling microstructure and irradiation effects. Metall Mater Trans A 42A:852

Becquart CS, Domain C, Sarkar U, DeBacker A, Hou M (2010) Microstructural evolution of irradiated tungsten: ab initio parameterisation of an OKMC model. J Nucl Mater 403(1–3): 75–88

Becquart CS, Barthe MF, De Backer A (2011) Modelling radiation damage and He production in tungsten. Phys Scripta T145:014048

Bertin N, Upadhyay MV, Pradalier C, Capolungo L (2015) A FFT-based formulation for efficient mechanical fields computation in isotropic and anisotropic periodic discrete dislocation dynamics. Model Simul Mater Sci Eng 23(6):065009

Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. J Comput Phys 17:10–18

Castin N, Bakaev A, Bonny G, Sand A, Malerba L, Terentyev D (2017) On the onset of void swelling in pure tungsten under neutron irradiation: an object kinetic Monte Carlo approach. J Nucl Mater 493:280–293

Castin N, Bonny G, Bakaev A, Ortiz C, Sand A, Terentyev D (2018) Object kinetic Monte Carlo model for neutron and ion irradiation in tungsten: impact of transmutation and carbon impurities. J Nucl Mater 500:15–25

Caturla MJ, Soneda N, Alonso E, Wirth BD, Diaz de la Rubia T, Perlado JM (2000) Comparative study of radiation damage accumulation in Cu and Fe. J Nucl Mater 276:13

Ceperley DM, Alder BJ (1980) Ground state of the electron gas by a stochastic method. Phys Rev Lett 45(7):4

Chatterjee A, Vlachos DG (2007) An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. J Comput-Aided Mater Des 14(2):253–308

Correa AA, Kohanoff J, Artacho E, Sánchez-Portal D, Caro A (2012) Nonadiabatic forces in ion-solid interactions: the initial stages of radiation damage. Phys Rev Lett 108(21):213201

Dai Y, Victoria M (1997) Defect cluster structure and tensile properties of copper single crystals irradiated with 600 MeV protons. MRS Proc 439:319–324

De Backer A, Sand A, Ortiz CJ, Domain C, Olsson P, Berthod E, Becquart CS (2016) Primary damage in tungsten using the binary collision approximation, molecular dynamic simulations and the density functional theory. Phys Scripta T167:014018

Domain C, Becquart CS, Malerba L (2004) Simulation of radiation damage in Fe alloys: an object kinetic Monte Carlo approach. J Nucl Mater 335:121

Draeger EW, Andrade X, Gunnels JA, Bhatele A, Schleife A, Correa AA (2017) Massively parallel first-principles simulation of electron dynamics in materials. J Parallel Distrib Comput 106: 205–214

Elcock EW (1959) Vacancy diffusion in ordered alloys. Proc Phys Soc 73:250

Elcock EW, McCombie CW (1958) Vacancy diffusion in binary ordered alloys. Phys Rev 109:605

El-Mellouhi F, Mousseau N, Lewis LJ (2008) Kinetic activation-relaxation technique: an off-lattice self-learning kinetic Monte Carlo algorithm. Phys Rev B 78:153202

Eyring H, Walter J, Kimball GE (1944) Quantum chemistry, 1st edn. Wiley, New York

Feibelman PJ (1990) Diffusion path for an Al adatom on Al(001). Phys Rev Lett 65(6):729–732

Fernandez N, Ferro Y, Kato D (2015) Hydrogen diffusion and vacancies formation in tungsten: density functional theory calculations and statistical models. Acta Mater 94:307–318

Fichthorn KA, Lin Y (2013) A local superbasin kinetic Monte Carlo method. J Chem Phys 138(16):164104

Fu CC, Dalla Torre J, Willaime F, Bocquet J-L, Barbu A (2005) Multiscale modelling of defect kinetics in irradiated iron. Nature Mater 4:68

Gámez L, Gámez B, Caturla MJ, Terentyev D, Perlado JM (2011) Object Kinetic Monte Carlo calculations of irradiated Fe-Cr dilute alloys: the effect of the interaction radius between substitutional Cr and self-interstitial Fe. Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 269(14):1684–1688

Gharaee L, Marian J, Erhart P (2016) The role of interstitial binding in radiation induced segregation in W-Re alloys. J Appl Phys 120(2):025901

Gibson MA, Bruck J (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. J Phys Chem A 104(9):1876–1889

Gillespie D (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22:403–434

Gillespie D (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81:2340–2361

Gilmer GH, Bennema P (1972) Simulation of crystal growth with surface diffusion. J Appl Phys 43(4):1347–1360

Glensk A, Grabowski B, Hickel T, Neugebauer J (2015) Understanding anharmonicity in fcc materials: from its origin to ab initio strategies beyond the quasiharmonic approximation. Phys Rev Lett 114(19):195901

Grabowski B, Hickel T, Neugebauer J (2007) Ab initio study of the thermodynamic properties of nonmagnetic elementary fcc metals: exchange-correlation-related error bars and chemical trends. Phys Rev B 76(2):024309

Grabowski B, Hickel T, Neugebauer J (2011) Formation energies of point defects at finite temperatures. Phys Status Solidi B 248(6):1295–1308. https://doi.org/10.1002/pssb.201046302

Hanusse P, Blanche A (1981) A Monte Carlo method for large reaction diffusion systems. J Chem Phys 74:6148

Heinisch HL (1990) Computer simulation of high energy displacement cascades. Radiat Eff Defects Solids 113:53

Henkelman G, Jonsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111(15):7010–7022

Henkelman G, Jonsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. J Chem Phys 113(22) 9978–9985

Henkelman G, Jonsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115(21):9657

Henkelman G, Uberuaga BP, Jonsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113(22):9901–9904

Henriksson KOE, Nordlund K, Krasheninnikov A, Keinonen J (2005) Difference in formation of hydrogen and helium clusters in tungsten. Appl Phys Lett 87(16):163113

Hohenberg P, Kohn W (1964) Inhomogenous electron gas. Phys Rev B 136:864–871

Huang G-Y, Juslin N, Wirth BD (2016) First-principles study of vacancy, interstitial, noble gas atom interstitial and vacancy clusters in bcc-W. Comput Mater Sci 123:121–130

Huang CH, Gharaee L, Zhao Y, Erhart P, Marian J (2017) Mechanism of nucleation and incipient growth of Re clusters in irradiated W-Re alloys from kinetic Monte Carlo simulations. Phys Rev B 96:094108

Huang CH, Gilbert MR, Marian J (2018) Simulating irradiation hardening in tungsten under fast neutron irradiation including Re production by transmutation. J Nucl Mat 499:204–215

Hudson TS, Dudarev SL, Caturla MJ, Sutton AP (2005) Effects of elastic interactions on post-cascade radiation damage evolution in kinetic monte carlo simulations. Philos Mag 85:661–675

Hunter A, Saied F, Le C, Koslowski M (2011) Large-scale 3D phase field dislocation dynamics simulations on high-performance architectures. Int J High Perform Comput Appl 25(2): 223–235

Jiménez F, Ortiz C (2016) A GPU-based parallel object kinetic monte carlo algorithm for the evolution of defects in irradiated materials. Comput Mater Sci 113:178–186

Kampen NGV (1992) Stochastic processes in physics and chemistry, 2nd edn. Elsevier, Amsterdam

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140(4A):1133

Körmann F, Dick A, Grabowski B, Hallstedt B, Hickel T, Neugebauer J (2008) Free energy of bcc iron: integrated ab initio derivation of vibrational, electronic, and magnetic contributions. Phys Rev B 78(3):033102

Langreth DC, Mehl MJ (1983) Beyond the local-density approximation in calculations of ground-state electronic properties. Phys Rev B 28(4):1809–1834

Lasa A, Tahtinen SK, Nordlund K (2014) Loop punching and bubble rupture causing surface roughening A model for W fuzz growth. EPL (Europhys Lett) 105(2):25002, 00010

Lu G-H, Zhou H-B, Becquart CS (2014) A review of modelling and simulation of hydrogen behaviour in tungsten at different scales. Nucl Fusion 54(8):086001

Malerba L, Caro A, Wallenius J (2008) Multiscale modelling of radiation damage and phase transformations: The challenge of FeCr alloys. J Nucl Mater 382:112–125

Martin-Bragado I, Tian S, Johnson M, Castrillo P, Pinacho R, Rubio J, Jaraiz M (2006) Modeling charged defects, dopant diffusion and activation mechanisms for TCAD simulations using kinetic Monte Carlo. Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 253(1–2):63–67

Martin-Bragado I, Rivera A, Valles G, Gomez-Selles JL, Caturla MJ (2013) MMonCa: an object kinetic Monte Carlo simulator for damage irradiation evolution and defect diffusion. Comput Phys Commun 184(12):2703–2710

Martinez E, Uberuaga BP (2015) Mobility and coalescence of stacking fault tetrahedra in Cu. Sci Rep 5:9084

Martínez E, Marian J, Kalos M, Perlado J (2008) Synchronous parallel kinetic Monte Carlo for continuum diffusion-reaction systems. J Comput Phys 227(8):3804–3823

Martínez E, Monasterio PR, Marian J (2011) Billion-atom synchronous parallel kinetic Monte Carlo simulations of critical 3D Ising systems. J Comp Phys 230:1359–1369

Mason DR, Yi X, Kirk MA, Dudarev SL (2014) Elastic trapping of dislocation loops in cascades in ion-irradiated tungsten foils. J Phys Condens Matter 26(37):375701

Maury F, Lucasson P, Lucasson A, Faudot F, Bigot J (1987) A study of irradiated FeCr alloys: deviations from Matthiessen's rule and interstitial migration. J Phys F Metal Phys 17(5): 1143–1165

Nguyen-Manh D, Dudarev S (2006) Multi-scale modelling of defect behavior in BCC transition metals and iron alloys for future fusion power plants. Mater Sci Eng A 423(1–2): 74–78

Niklasson A, Tymczak C, Challacombe M (2006) Time-reversible born-oppenheimer molecular dynamics. Phys Rev Lett 97(12):123001

Novotny MA (1995) Monte Carlo algorithms with absorbing Markov chains: fast local algorithms for slow dynamics. Phys Rev Lett 74(1):1

Oda T, Zhu D, Watanabe Y (2015) Kinetic Monte Carlo simulation on influence of vacancy on hydrogen diffusivity in tungsten. J Nucl Mater 467:439–447

Ohsawa K, Goto J, Yamakami M, Yamaguchi M, Yagi M (2010) Trapping of multiple hydrogen atoms in a tungsten monovacancy from first principles. Phys Rev B 82(18):184117

Olsson P (2009) Ab initio study of interstitial migration in Fe-Cr alloys. J Nucl Mater 386–388: 86–89

Olsson P, Domain C, Wallenius J (2007) Ab initio study of Cr interactions with point defects in bcc Fe. Phys Rev B 75:014110

Olsson P, Becquart CS, Domain C (2016) Ab initio threshold displacement energies in iron. Mater Res Lett 4(4):219–225

Opplestrup T, Bulatov V, Gilmer G, Kalos M, Sadigh B (2006) First-passage Monte Carlo algorithm: diffusion without all the Hops. Phys Rev Lett 97(23):230602

Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. Phys Rev B 46(11):6671–6687

Perini A, Jacucci G, Martin G (1984) Cluster free energy in the simple-cubic Ising model. Phys Rev B 29(5):2689–2697

Schulze TP (2002) Kinetic Monte Carlo simulations with minimal searching. Phys Rev E 65(3):036704

Shim Y, Amar J (2005a) Rigorous synchronous relaxation algorithm for parallel kinetic Monte Carlo simulations of thin film growth. Phys Rev B 71(11):115436

Shim Y, Amar JG (2005b) Semirigorous synchronous sublattice algorithm for parallel kinetic Monte Carlo simulations of thin film growth. Phys Rev B 71:125432

Sholl DS, Steckel JA (2009) Density functional theory - A practical introduction. Wiley, Hoboken

Slepoy A, Thompson AP, Plimpton SJ (2008) A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. J Chem Phys 128(20):205101

Söderlind P, Yang LH, Moriarty JA, Wills JM (2000) First-principles formation energies of monovacancies in bcc transition metals. Phys Rev B 61(4):2579–2586

Sørensen MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 112(21):9599–9606

Subramanian G, Perez D, Uberuaga BP, Tomé CN, Voter AF (2013) Method to account for arbitrary strains in kinetic Monte Carlo simulations. Phys Rev B 87(14):144107

Suzudo T, Yamaguchi M, Hasegawa A (2014) Stability and mobility of rhenium and osmium in tungsten: first principles study. Model Simul Mater Sci Eng 22(7):075006

Takaki S, Fuss J (1983) Dedek HKU, Schultz H. The resistivity recovery of high purity and carbon doped iron following low electron irradiation. Rad Eff 79:87–122

Trushin O, Karim A, Kara A, Rahman TS (2005) Self-learning kinetic Monte Carlo method: application to Cu(111). Phys Rev B 72:115401

Uberuaga BP, Hoagland RG, Voter AF, Valone SM (2007) Direct Transformation of Vacancy Voids to Stacking Fault Tetrahedra. Phys Rev Lett 99:135501

Uberuaga BP, Martínez E, Perez D, Voter AF (2018) Discovering mechanisms relevant for radiation damage evolution. Comput Mater Sci 147:282–292

Valles G, González C, Martin-Bragado I, Iglesias R, Perlado J, Rivera A (2015a) The influence of high grain boundary density on helium retention in tungsten. J Nucl Mater 457:80–87

Valles G, Cazalilla AL, Gonzalez C, Martin-Bragado I, Prada A, Iglesias R, Perlado J, Rivera A (2015b) A multiscale approach to defect evolution in tungsten under helium irradiation. Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms 352:100–103

Valles G, Panizo-Laiz M, González C, Martin-Bragado I, González-Arrabal R, Gordillo N, Iglesias R, Guerrero C, Perlado J, Rivera A (2017a) Influence of grain boundaries on the radiation-induced defects and hydrogen in nanostructured and coarse-grained tungsten. Acta Mater 122:277–286

Valles G, Martin-Bragado I, Nordlund K, Lasa A, Björkas C, Safi E, Perlado J, Rivera A (2017b) Temperature dependence of underdense nanostructure formation in tungsten under helium irradiation. J Nucl Mater 490:108–114

Vattré A, Jourdan T, Ding H, Marinica M-C, Demkowicz MJ (2016) Non-random walk diffusion enhances the sink strength of semicoherent interfaces. Nat Commun 7:10424

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3(1–2):121–127

Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett 78(20):3908

Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57(22):R13985

Voter AF (2007) Introduction to the kinetic Monte Carlo method, Ch. 1. Springer, Dordrecht

Voter AF, Doll JD (1984) Transition state theory description of surface self-diffusion: comparison with classical trajectory results. J Chem Phys 80:5832

Voter AF, Doll JD (1985) Dynamical corrections to transition state theory for multistate systems: surface self-diffusion in the rare-event regime. J Chem Phys 82:80–92

Was GS (2007) Fundamentals of radiation materials science. Springer, Berlin/Heidelberg/New York

Wen M, Takahashi A, Ghoniem NM (2009) Kinetics of self-interstitial cluster aggregation near dislocations and their influence on hardening. J Nucl Mater 392:386–395

Witt WC, del Rio BG, Dieterich JM, Carter EA (2018) Orbital-free density functional theory for materials research. J Mater Res 33:1–19

Xu L, Henkelman G (2008) Adaptive kinetic Monte Carlo for first-principles accelerated dynamics. J Chem Phys 129:114104

Xu D, Wirth BD, Li M, Kirk MA (2012) Defect microstructural evolution in ion irradiated metallic nanofoils: kinetic Monte Carlo simulation versus cluster dynamics modeling and in situ transmission electron microscopy experiments. Appl Phys Lett 101(10):101905

Young WM, Elcock EW (1966) Monte Carlo studies of vacancy migration in binary ordered alloys: I. Proc Phys Soc 89:735

Zhang J, Zhang H, Ye H, Zheng Y (2016) Free-end adaptive nudged elastic band method for locating transition states in minimum energy path calculation. J Chem Phys 145(9):094104

Zhu T, Li J, Samanta A, Kim HG, Suresh S (2007) Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. Proc Natl Acad Sci 104(9):3031–3036

Zinkle SJ, Farrell K (1989) Void swelling and defect cluster formation in reactor-irradiated copper. J Nucl Mater 168:262–267

Rate Theory: Cluster Dynamics, Grouping Methods, and Best Practices

101

Thomas Jourdan

Contents

Abstract

Cluster dynamics is a mean field method which is used to simulate the evolution of defect clusters in materials under aging and irradiation conditions. Its popularity is due to the simplicity of the underlying "cluster gas" model, which is able to simulate accurately and efficiently dilute cases, provided it is carefully parametrized. In this chapter cluster dynamics formalism is reviewed. Parametrization of cluster dynamics models and numerical schemes to efficiently solve cluster dynamics equations are discussed.

1   Introduction

Microstructure evolution under irradiation is driven by phenomena occurring over very different time and length scales. For example, displacement cascades are produced in a few picoseconds and span over a few tens of nanometers, whereas subsequent defect diffusion over micrometers requires a few seconds or more.

T. Jourdan (✉)

DEN-Service de Recherches de Métallurgie Physique, CEA, Université Paris-Saclay, Gif-sur-Yvette, France

e-mail: thomas.jourdan@cea.fr

Quantitative simulation of the impact of radiation damage over experimental time scales requires the modeling of all these processes. Although no single tool is capable of performing such a task, cluster dynamics has emerged as a versatile tool to explore various consequences of irradiation. In particular, when it is informed with atomistic calculations, it has proven capable of simulating precipitation in dilute alloys (Clouet et al. 2005; Jourdan et al. 2010), void and loop nucleation, growth, and coarsening (Ortiz et al. 2007; Xu et al. 2012b; Moll et al. 2013), without suffering the limitations of fully atomistic calculations.

A distinctive feature of cluster dynamics is its mean field character, which solves the length scale problem. Microstructure evolution is obtained by solving a set of ordinary differential equations (ODEs) on cluster concentrations. However, for some phenomena which are highly localized in space, such as displacement cascades, mean field formalism can be a drawback, and special care must be taken. Concerning the time scale problem, high-frequency events, which limit the efficiency of stochastic methods such as object kinetic Monte Carlo (OKMC), may also hinder the solving of cluster dynamics equations. More precisely, the large scatter in reaction rates leads to very stiff equations, for which dedicated methods must be used. In addition, to simulate large clusters, it can be essential to reduce the number of equations by approximating the initial ODE set. Grouping methods (Golubov et al. 2001a) and Fokker-Planck-based methods (Ghoniem and Sharafat 1980) are useful ways to achieve such reduction.

In this article we first present the cluster dynamics formalism. Then, parametrization of cluster dynamics models is discussed. In the last section, technical aspects for the efficient solving of cluster dynamics equations are presented.

2   Cluster Dynamics Formalism

Cluster dynamics is a mean field approach which considers the microstructure as a collection of clusters. These clusters can diffuse in the matrix and absorb and emit other clusters (Martin 2006; Clouet 2009). For this reason, this model is sometimes called a "cluster gas model." As such, cluster dynamics is essentially appropriate for dilute cases, where clusters can be distinguished unambiguously. Such cases include the precipitation of secondary phases in dilute alloys and the formation of point defect clusters. It bears much resemblance to OKMC and mostly requires the same input parameters.

Let us consider, for example, the formation of vacancy and self-interstitial clusters under irradiation in a pure metal. Assuming that only one type of cluster exists for each self-defect, point defect clusters can be conveniently identified by a single index n. We use $n < 0$ for a vacancy cluster containing $|n|$ vacancies and $n > 0$ for an interstitial cluster containing n self-interstitials. The evolution equation, for the concentration C_n of mobile clusters in class (n), reads (Hardouin-Duparc et al. 2002; Xu et al. 2012a; Jourdan et al. 2014):

$$\frac{dC_n}{dt} = G_n + \sum_{\substack{m \in \mathcal{M} \\ n-m \neq 0}} J_{n-m,n} - \sum_{m \in \mathcal{M}} J_{n,n+m} - \sum_{m \in \mathbb{Z} \setminus \{0\}} J_{m,m+n}$$

$$- \sum_j k_{j,n}^2 D_n \left(C_n - C_{n,j}^{\mathrm{e}} \right). \tag{1}$$

In this equation, G_n and D_n are the production rate and the diffusion coefficient of cluster (n), respectively, and $\mathcal{M}$ is the set of mobile species. The factor $k_{j,n}^2$ represents the sink strength of sink type j which is not explicitly treated with cluster dynamics variables, in analogy with classical rate theory approach (Nichols 1978). In the present case, network dislocations and grain boundaries, for example, are taken into account with sink strengths, whereas vacancy and interstitial clusters are explicitly treated with the dedicated equations on C_n. Equilibrium concentration of (n) near sink j is noted $C_{n,j}^{\mathrm{e}}$.

The flux $J_{n,n+m}$ is a net flux between classes (n) and ($n+m$), due to the mobility of (m) only. It is defined by:

$$J_{n,n+m} = \beta_{n,m} C_n C_m - \alpha_{n+m,m} C_{n+m}, \tag{2}$$

where $\beta_{n,m}$ and $\alpha_{n+m,m}$ are the absorption and dissociation rates, respectively. The effect of mobility of (n) is taken into account with the term $J_{m,m+n}$ in Eq. (1). Absorption rates are closely related to the sink strengths used in rate theory (Nichols 1978). They take different forms, depending on the cluster type, the presence of long-range interactions (elastic, but also electrostatic interactions for materials with charged defects Kotomin and Kuzovkov 1992), and the presence of an energy barrier for the association of clusters (Waite 1958). For example, if (n) is a cavity in an infinitely dilute system and if elastic interactions between (n) and (m) are neglected, we obtain, for a three-dimensional diffusion of (m) (Waite 1957):

$$\beta_{n,m} = 4\pi (r_n + r_m) D_m, \tag{3}$$

which corresponds to the classical sink strength $k_{n,m}^2 = 4\pi (r_n + r_m) C_n$. In Eq. (3) r_n is the radius of cluster (n). The emission rate of (m) by ($n + m$) is obtained by imposing that $J_{n,n+m}$ is zero at equilibrium, so

$$\alpha_{n+m,m} = \frac{\beta_{n,m}}{V_{\mathrm{at}}} \exp\left(-\frac{F_{n+m,m}^{\mathrm{b}}}{k_{\mathrm{B}} T} \right), \tag{4}$$

where $F_{n+m,m}^{\mathrm{b}} = F_n^{\mathrm{f}} + F_m^{\mathrm{f}} - F_{n+m}^{\mathrm{f}}$ is the binding free energy of cluster (m) to cluster (n) and F_n^{f} is the formation free energy of cluster (n). This free energy notably contains a configurational entropy term. It is important to note that the thermodynamic part of a cluster dynamics model is only given by the

emission coefficients. In particular, contrary to the classical nucleation theory (Feder et al. 1966), the chemical potential of the monomer in the matrix is not present. This is due to the fact that in cluster dynamics there is no distinction between the matrix, where heterophase fluctuations occur, and the secondary phase. All clusters, whatever their size, evolve due to the absorption and emission processes. This seamless description of the nucleation, growth, and coarsening processes is one of the strengths of the method.

If the hypothesis of infinite dilution is not well justified, corrections can be performed on the absorption (Nichols 1978; Brailsford and Bullough 1981) and on the emission (Lépinoux 2009; Berthier et al. 2010) rates.

Under irradiation, dislocation loops quickly reach sizes which are comparable to the mean distance between clusters. Loops interact with each other or with preexisting dislocations. Loop size and density saturate, while the network dislocation density reaches a steady state, as a result of incorporation of loops and dislocation recovery (Brager et al. 1977). Cluster dynamics, as described above, is unable to simulate such phenomena. Although the validity of cluster dynamics at high doses is questionable, it may be desirable to include loop interactions and dislocation recovery into cluster dynamics formalism to perform qualitative modeling in support of experiments. This has been done recently (Zouari et al. 2011; Michaut et al. 2017), based on evolution equations proposed for rate theory (Stoller and Odette 1987) and then adapted for cluster dynamics (Jourdan 2015).

Up to now, cluster dynamics has been described for a pure metal containing self-defects or a dilute binary alloy under thermal aging. For more complex materials, a natural extension of Eq. (1) is to describe clusters through tuples $(n_1, n_2, \ldots, n_p)$, where p is the number of components in clusters. This has been done mostly in the case of self-defect clusters in the presence of gas atoms, such as helium and hydrogen (Ghoniem et al. 1983; Golubov et al. 2007; Ortiz et al. 2007; Marian and Bulatov 2011; Moll et al. 2013). For multicomponent alloys, one of the difficulties is to parametrize the model. In addition, negative flux couplings can arise, which cannot be taken into account in cluster dynamics due to its mean field character. Even if a reliable parametrization is available, the number of equations to solve becomes too large as soon as three components or more are present. For all these reasons, approximations to the initial problem can be necessary. Some recent attempts have aimed at simulating either precipitation or self-defect cluster evolution in multicomponent alloys (Ke et al. 2017; Mamivand et al. 2017), using equations similar to Eq. (1) coupled with CALPHAD approach.

Although surfaces can be taken into account in an effective way through sink strengths, it is sometimes necessary to describe them explicitly. For this purpose, a spatial dependency of rate equations on depth is introduced. The system is however assumed to be locally homogeneous, so that the mean field hypothesis of cluster dynamics remains valid. In practice, a diffusion term is added to Eq. (1) (Ortiz et al. 2004; Xu et al. 2012a; Jourdan et al. 2014), and the system is discretized along the depth. This approach is well adapted to simulate, for example, nonuniform damage creation by ion irradiation or irradiation of thin foils. Spatially resolved cluster dynamics in three dimensions has also been developed (Dunn and Capolungo 2015).

3 Parametrization of Cluster Dynamics

Among the parameters of cluster dynamics models, diffusion coefficients and binding energies at low temperature for small clusters can be estimated by atomistic calculations (Domain and Becquart 2001; Fu et al. 2004; Fu and Willaime 2008). Analytical laws for binding energies can be used to extrapolate values for larger clusters (Soneda and Diaz de la Rubia 2001; Varvenne et al. 2014; Alexander et al. 2016). A more difficult task is to calculate the binding *free* energies, so very often entropic contributions are neglected. However, it should be noted that to lead to quantitative results, such contributions should be included. In particular, it has been shown that cluster dynamics can reproduce atomistic kinetic Monte Carlo (AKMC) results on precipitation under thermal aging, provided that cluster free energies, taking into account configurational entropies, are calculated with the same energy model (Clouet et al. 2005; Jourdan et al. 2010) (Fig. 1). Another example is the annealing of loops in iron after helium implantation (Moll et al. 2013). Using an atomistically-informed model for the thermodynamics of helium bubbles (Jourdan and Crocombette 2011) appeared crucial to properly model the annealing kinetics of interstitial dislocation loops, which evolve due to Ostwald ripening by vacancy emission. Modeling the kinetics permitted, in turn, to *deduce* the vacancy free
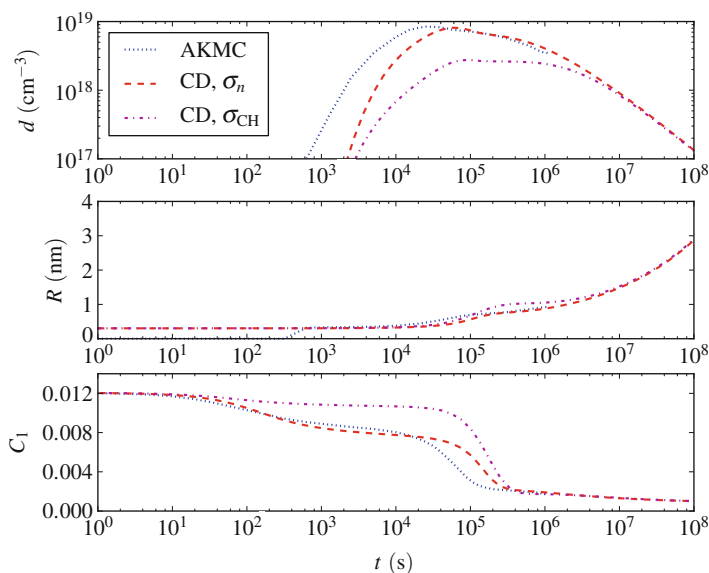
Fig. 1 Density of copper precipitates, mean radius, and monomer concentration in AKMC and cluster dynamics (CD) in a Fe-1.34at.%Cu alloy during thermal aging at 500 °C. Two different models for interface free energies in cluster dynamics are used, a Cahn-Hilliard approach (σ_{CH}) or a size-dependent interface energy (σ_n), obtained by the method of overlapping distributions, using the same energy model as for AKMC simulations (Jourdan et al. 2010)

activation energy from experiments. In general, under thermal aging, results are more sensitive to the precise values of emission coefficients than to the absorption coefficients.

For complex alloys, setting up a physically based thermodynamic parametrization is much more difficult than for model materials, if not impossible. A common approach is to use the so-called gray alloy approach. Only self-defect clusters are considered, and their properties reflect the chemical composition of the materials. Such an approach must somehow be calibrated on experiments and has a limited predictive character. However, it can be useful to perform sensitivity studies or to investigate the relevance of some mechanisms (Zouari et al. 2011; Michaut et al. 2017).

Under irradiation, microstructures of self-defect clusters are also very sensitive to coefficients $\beta_{n,m}$. Indeed, the growth rate of these clusters depends on the net flux of interstitials and vacancy to them. Since in general interstitials and vacancies are created in equal numbers, the growth of self-defect clusters results from tiny differences between interstitial and vacancy fluxes. These differences, which are crucial to explain phenomena such as void swelling and irradiation creep, are essentially explained by elastic effects which favor the absorption of interstitials by dislocations. In this "dislocation bias" model (Greenwood et al. 1959; Bullough and Perrin 1970), the preferential absorption of interstitials by dislocations leads, by conservation of matter, to a net flux of vacancies to voids. Therefore simulating quantitatively the growth of voids and dislocation loops requires precise knowledge of absorption coefficients. Obtaining sink strengths from experiments requires some assumptions (Kiritani et al. 1975; Golubov et al. 2001b), such as the fraction of freely migrating species. Alternatively, sink strengths can be evaluated numerically (Dubinko et al. 2005; Rouchette et al. 2014; Heinisch et al. 2000). It has been shown, however, that sink strengths greatly depend on the precise properties of point defects, such as point defect anisotropy at saddle point (Dederichs and Schroeder 1978; Skinner and Woo 1984; Carpentier et al. 2017), so up to now general laws to be implemented in cluster dynamics are still lacking.

Besides bias values, cluster dynamics results under irradiation are known to be very sensitive to production rates G_n. To define them, it is tempting to perform molecular dynamics (MD) simulations of displacement cascades, count the remaining clusters at the end of the simulations, and average the cluster populations over several runs. However, this method overestimates cluster densities in general. Indeed, at the end of MD simulations, clusters are still spatially correlated, whereas cluster dynamics can only describe a homogenous distribution of clusters. To circumvent this problem, an idea is to let self-defects and self-defect clusters diffuse over longer times, with OKMC, for example (Adjanor et al. 2010; Becquart and Domain 2011). It can be expected that if the temperature is high enough, the damage will be less correlated at the end of the simulation, so a transfer to cluster dynamics is more valid. However, it is not always easy to define a proper annealing time, since fast diffusers (typically interstitials) will homogenize more rapidly than slow diffusers (typically vacancies). For this reason it seems more legitimate to work on homogenization distance instead of homogenization time. This homogenization distance is given by cluster dynamics itself: for example, two clusters a few nanometers apart can be considered either as spatially correlated if the
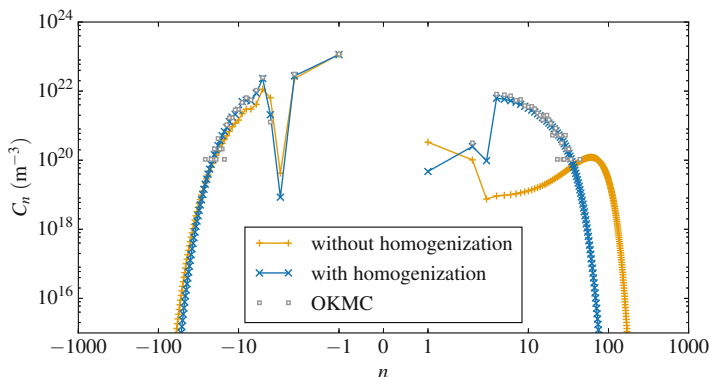
Fig. 2 Cluster distributions obtained with full OKMC simulations or cluster dynamics with two different source terms. Cascades are produced with the primary damage code MAR-LOWE (Robinson 1989), so only mono-interstitials and mono-vacancies are generated. For the non-homogenization case, a zero-time homogenization process was actually imposed to trigger close pair recombinations and clustering, but no thermally activated diffusion. Such small times are typical of MD simulations. Homogenization was performed using the procedure published in Jourdan and Crocombette (2012). Calculations were performed at 300 K up to 10^{-3} dpa, for an irradiation of iron with 20 keV iron ions and a damage rate of 10^{-4} dpa.s^{-1} (see Jourdan and Crocombette 2012 for more details)

density of clusters in cluster dynamics is low (e.g., at the beginning of irradiation) or uncorrelated if the average distance between defects in cluster dynamics is also a few nanometers. On this basis, a procedure has been proposed to provide a homogenized source term which is time dependent, even under constant irradiation (Jourdan and Crocombette 2012). A comparison between cluster dynamics with homogenized source term and full OKMC simulations, for an irradiation with 20 keV iron ions, is shown in Fig. 2. It should be noted that if only isolated Frenkel pairs are created by irradiation, this method is equivalent to reducing the production rate of self-interstitials and vacancies, due to recombinations which occur during homogenization. For ion and neutron irradiations, the source terms also contain self-defect clusters, either because they are created during the MD simulations or because they result from the diffusion and agglomeration of defects during the homogenization process. Therefore, this procedure naturally includes some aspects of "production bias" rate theory model (Golubov et al. 2012). One-dimensional migration of interstitial clusters, which is another cornerstone of this model, can be implemented by appropriate absorption coefficients (Singh et al. 1997).

4   Solving Cluster Dynamics Equations

Cluster concentrations at a given time are obtained by solving an ODE set containing equations similar to Eq. (1). This ODE set is known to be very stiff, which means that the use of explicit schemes requires very small time steps, so it may not be possible to run the problem until the desired physical time. It was soon recognized

that implicit methods based on multistep methods, such as the GEAR package (Hindmarsh and Gelinas 1971) or the more recent CVODES solver (Hindmarsh et al. 2005), are more appropriate for such problems (Ghoniem and Sharafat 1980). These methods require the solving of linear systems based on the Jacobian matrix of the ODE set. Without optimization of this step, the calculation may be intractable for memory and computation time problems. For systems containing more than a few thousands of equations, it is clearly advantageous to provide the analytical form of the Jacobian matrix, instead of letting the solver evaluate it by finite differences. In addition, if only a small proportion of cluster classes are mobile, the Jacobian matrix is sparse, and dedicated solvers can be used. Examples of the use of sparse direct solvers are given in Xu et al. (2012a) and Jourdan et al. (2014). A final improvement consists in resizing the cluster size space on the fly, in order to avoid useless consideration of zero concentration clusters (Jourdan et al. 2014).

Even when the sparsity of the Jacobian matrix is taken into account, the computation time and memory use can become too large as the maximum cluster size increases. For example, to simulate the formation of a 10 nm cavity in iron filled with helium atoms over a spatial discretization of 100 slices, it is necessary to consider more than 10^{11} equations, which is well beyond the current capabilities of computation resources. In this case, approximating the master equation (1) becomes necessary. Two different kinds of methods can be distinguished:

- Grouping methods, first developed by Kiritani (1973), consist in solving equations for the moments of the cluster distribution over groups of equations. It was shown by Golubov et al. (2001a) and Ovcharenko et al. (2003) that accurate results can be obtained by taking into account the zeroth and first moments of the distribution. One of the advantages of this method is to be able to handle the mobility of large clusters (Golubov et al. 2007). However, it does not guarantee that cluster concentrations remain positive.
- Fokker-Planck-based methods. Although they can also be viewed as grouping methods since the final number of equations is reduced, the mathematical justification is completely different. For large values of n, the master equation for immobile species can be written as a Fokker-Planck equation (Goodrich 1964; Wolfer et al. 1977):

$$\frac{\partial}{\partial t} C(x, t) = -\frac{\partial}{\partial x} [F(x, t) C(x, t)] + \frac{\partial^2}{\partial x^2} [D(x, t) C(x, t)], \quad (5)$$

with $x_n = n$, $C(x_n, t) \approx C_n(t)$ and

$$F(x, t) = \sum_{m \in \mathcal{M}} m [\beta_m(x) C_m(t) - \alpha_m(x)], \quad (6)$$

$$D(x, t) = \frac{1}{2} \sum_{m \in \mathcal{M}} m^2 [\beta_m(x) C_m(t) + \alpha_m(x)]. \quad (7)$$

This equation can then be discretized on a coarse mesh to reduce the number of equations to solve. Special care must be taken to connect the master equation used for small clusters to the discretized Fokker-Planck equation used for large clusters (Jourdan et al. 2014). Formulating the problem with a conservative form and using a finite volume approach is particularly convenient to ensure that matter is exactly conserved (Jourdan et al. 2014). The last important point is to choose an appropriate discretization scheme for the advection term. Recently, it has been shown that high-resolution, monotonicity-preserving schemes permit to accurately reproduce reference cluster distributions obtained by solving all equations, without introducing negative concentrations nor spurious oscillations (Jourdan et al. 2016) (Fig. 3). Although they are more computationally expensive than simple upwind schemes, high-resolution schemes do not introduce numerical diffusion and should therefore be preferred when high precision is needed on cluster
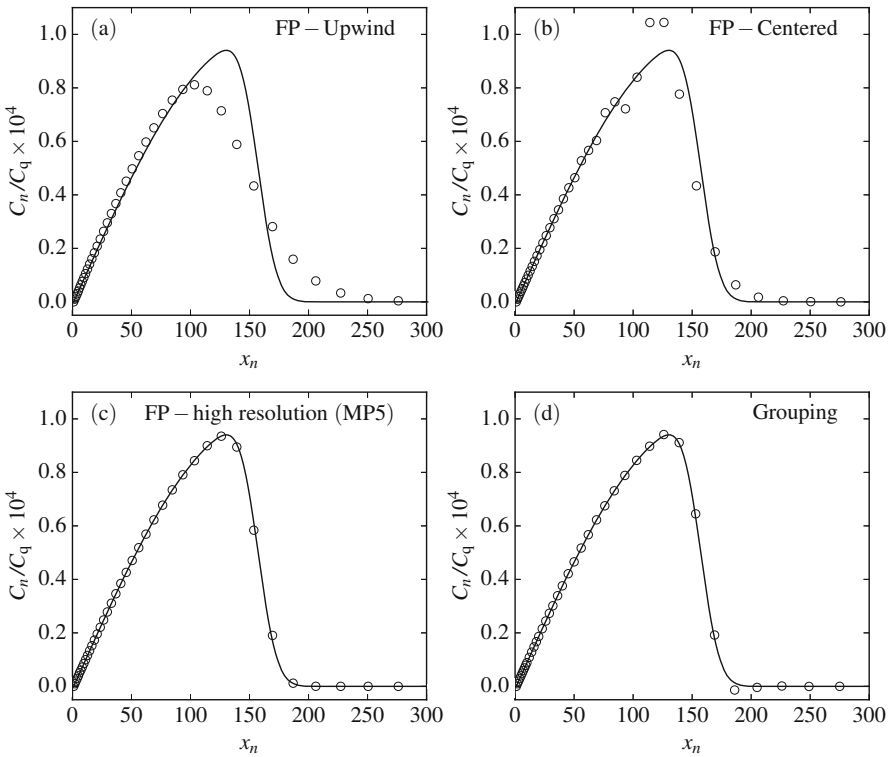
Fig. 3 Long-time limit cluster distributions for the clustering of quenched-in vacancies (Koiwa 1974) using different approximations (circles): (**a, b, c**) Fokker-Planck (FP) with upwind, centered, and high-resolution schemes for the advection term and (**d**) grouping method. Grouping and Fokker-Planck with high-resolution discretization scheme give very similar results, except in the distribution tail where negative concentrations occur with the grouping method (Jourdan et al. 2016)

distributions. The Fokker-Planck method can be generalized to multidimensional cases and permits to drastically reduce the number of equations. An example is given in Moll et al. (2013), where a spatialized simulation of helium implantation in iron could be performed. One drawback of this method, however, is that it is not really appropriate to treat the mobility of large clusters.

Alternatively, stochastic methods can be used to solve cluster dynamics equations. Among them, the simplest one is the so-called stochastic simulation algorithm (SSA) proposed by Gillespie (1976) and recently used for cluster dynamics (Marian and Bulatov 2011; Hoang et al. 2015). A finite volume V is envisaged, and equations on the number of clusters $N_n = C_n V$, derived from Eq. (1), are solved by a kinetic Monte Carlo method, using the appropriate rates for all possible events. An advantage of this method is that it does not require the solving of linear systems, so the memory requirements are lower than for deterministic solving, especially if large cluster spaces are considered. However, time steps are in general very low due to high-frequency events, so it is less efficient than deterministic solving for systems with one- or two-dimensional cluster spaces. Coupling stochastic approaches for large cluster sizes to deterministic solving for small clusters has been proposed, in order to avoid the stochastic treatment of high-frequency events (Surh et al. 2004; Gherardi et al. 2012; Terrier et al. 2017).

5   Conclusions and Perspectives

Over the last two decades, cluster dynamics has proven an interesting method to simulate irradiation processes over long time scales. Two aspects must be considered with care to obtain reliable results within the domain of validity of cluster dynamics: the numerical schemes to solve cluster dynamics equations and the parametrization of models. Efficient and accurate numerical schemes exist for time integration and for grouping methods. Additional efficiency improvements could be obtained by using hybrid methods, which couple deterministic to stochastic solving. Concerning the parametrization of cluster dynamics models, atomistic calculations now provide a large amount of data that can be used either directly in cluster dynamics or to parametrize scaling laws. Although no complete parametrization, from cluster free energies to sink strengths, exists even in pure metals, all tools are available to fill this gap. A remaining issue is to be able to treat heterogeneities in materials, which is not possible with current cluster dynamics models.

References

Adjanor G, Bugat S, Domain C, Barbu A (2010) Overview of the RPV-2 and INTERN-1 packages: from primary damage to microplasticity. J Nucl Mater 406:175

Alexander R, Marinica MC, Proville L, Willaime F, Arakawa K, Gilbert MR, Dudarev SL (2016) Ab initio scaling laws for the formation energy of nanosized interstitial defect clusters in iron, tungsten, and vanadium. Phys Rev B 94:024103

Becquart CS, Domain C (2011) Modeling microstructure and irradiation effects. Metall Mater Trans A 42A:852

Berthier F, Braems I, Maras E, Creuze J, Legrand B (2010) Extending cluster dynamics to concentrated and disordered alloys: the linear-chain case. Acta Mater 58:2387

Brailsford AD, Bullough R (1981) The theory of sink strengths. Philos Trans R Soc Lond A 302:87

Brager HR, Garner FA, Gilbert ER, Flinn JE, Wolfer WG (1977) Stress-affected microstructural development and the creep-swelling interrelationship. In: Bleiberg ML, Bennett JW (eds) Radiation effects in breeder reactor structural materials. The Metallurgical Society of AIME, New York, p 727

Bullough R, Perrin RC (1970) The mechanism and kinetics of void growth during neutron irradiation. In: Bement A (ed) Irradiation effects on structural alloys for nuclear reactor applications. ASTM International, West Conshohocken, pp 317–329

Carpentier D, Jourdan T, Le Bouar Y, Marinica MC (2017) Effect of saddle point anisotropy of point defects on their absorption by dislocations and cavities. Acta Mater 136:323

Clouet E (2009) Modeling of nucleation processes. In: Furrer DU, Semiatin SL (eds) Fundamentals of modeling for metals processing. ASM Handbook, vol 22A. ASM International, Materials Park, Ohio, p 203

Clouet E, Barbu A, Lae L, Martin G (2005) Precipitation kinetics of $A1_3Zr$ and Al_3Sc in aluminum alloys modeled with cluster dynamics. Acta Mater 53:2313. https://doi.org/10.1016/j.actamat.2005.01.038

Dederichs PH, Schroeder K (1978) Anisotropic diffusion in stress fields. Phys Rev B 17:2524

Domain C, Becquart CS (2001) Ab initio calculations of defects in Fe and dilute Fe-Cu alloys. Phys Rev B 65:024103

Dubinko VI, Abyzov AS, Turkin AA (2005) Numerical evaluation of the dislocation loop bias. J Nucl Mater 336:11

Dunn AY, Capolungo L (2015) Simulating radiation damage accumulation in α-Fe: a spatially resolved stochastic cluster dynamics approach. Comput Mater Sci 102:314

Feder J, Russel KC, Lothe J, Pound GM (1966) Homogeneous nucleation and growth of droplets in vapours. Adv Phys 15:111

Fu CC, Willaime F (2008) First principles calculations in iron: structure and mobility of defect clusters and defect complexes for kinetic modelling. C R Phys 9:335

Fu CC, Willaime F, Ordejón P (2004) Stability and mobility of mono- and di-interstitials in α-Fe. Phys Rev Lett 92:175503

Gherardi M, Jourdan T, Le Bourdiec S, Bencteux G (2012) Hybrid deterministic/stochastic algorithm for large sets of rate equations. Comput Phys Commun 183:1966

Ghoniem NM, Sharafat S (1980) A numerical solution to the Fokker-Planck equation describing the evolution of the interstitial loop microstructure during irradiation. J Nucl Mater 92:121

Ghoniem NM, Sharafat S, Williams JM, Mansur LK (1983) Theory of helium transport and clustering in materials under irradiation. J Nucl Mater 117:96

Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22:403

Golubov SI, Ovcharenko AM, Barashev AV, Singh BN (2001a) Grouping method for the approximate solution of a kinetic equation describing the evolution of point-defect clusters. Philos Mag A 81:643

Golubov SI, Singh BN, Trinkus H (2001b) On recoil-energy-dependent defect accumulation in pure copper Part II. Theoretical treatment. Philos Mag A 81:2533

Golubov SI, Stoller RE, Zinkle SJ, Ovcharenko AM (2007) Kinetics of coarsening of helium bubbles during implantation and post-implantation annealing. J Nucl Mater 361:149

Golubov SI, Barashev AV, Stoller RE (2012) Radiation damage theory. In: Konings RJM (ed) Comprehensive nuclear materials, vol 1, Elsevier, Oxford, pp 357–391

Goodrich FC (1964) Nucleation rates and the kinetics of particle growth II. The birth and death process. Proc R Soc Lond A 277:167

Greenwood GW, Foreman AJE, Rimmer DE (1959) The role of vacancies and dislocations in the nucleation and growth of gas bubbles in irradiated fissile material. J Nucl Mater 1:305

Hardouin-Duparc A, Moingeon C, Smetniansky-de-Grande N, Barbu A (2002) Microstructure modelling of ferritic alloys under high flux 1 MeV electron irradiations. J Nucl Mater 302:143

Heinisch HL, Singh BN, Golubov SI (2000) The effects of one-dimensional glide on the reaction kinetics of interstitial clusters. J Nucl Mater 283–287:737

Hindmarsh AC, Gelinas RJ (1971) GEAR: ordinary differential equation system solver. Technical report UCID-30001, Lawrence Radiation Laboratory

Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, Woodward CS (2005) SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. ACM Trans Math Softw 31:363. https://doi.org/10.1145/1089014.1089020

Hoang TL, Marian J, Bulatov VV, Hosemann P (2015) Computationally-efficient stochastic cluster dynamics method for modeling damage accumulation in irradiated materials. J Comput Phys 300:254

Jourdan T (2015) Influence of dislocation and dislocation loop biases on microstructures simulated by rate equation cluster dynamics. J Nucl Mater 467:286

Jourdan T, Crocombette JP (2011) A variable-gap model for calculating free energies of helium bubbles in metals. J Nucl Mater 418:98

Jourdan T, Crocombette JP (2012) Rate theory cluster dynamics simulations including spatial correlations within displacement cascades. Phys Rev B 86:054113

Jourdan T, Soisson F, Clouet E, Barbu A (2010) Influence of cluster mobility on Cu-precipitation in α-Fe: a cluster dynamics modeling. Acta Mater 58:3400

Jourdan T, Bencteux G, Adjanor G (2014) Efficient simulation of kinetics of radiation induced defects: a cluster dynamics approach. J Nucl Mater 444:298

Jourdan T, Stoltz G, Legoll F, Monasse L (2016) An accurate scheme to solve cluster dynamics equations using a Fokker-Planck approach. Comput Phys Commun 207:170

Ke H, Wells P, Edmondson PD, Almirall N, Barnard L, Odette GR, Morgan D (2017) Thermodynamic and kinetic modeling of Mn-Ni-Si precipitates in low-Cu reactor pressure vessel steels. Acta Mater 138:10

Kiritani M (1973) Analysis of the clustering process of supersaturated lattice vacancies. J Phys Soc Jpn 35:95

Kiritani M, Yoshida N, Takata H, Maehara Y (1975) Growth of interstitial type dislocation loops and vacancy mobility in electron irradiated materials. J Phys Soc Jpn 38(6):1677

Koiwa M (1974) On the validity of the grouping method – comments on "analysis of the clustering process of supersaturated lattice vacancies". J Phys Soc Jpn 37:1532

Kotomin E, Kuzovkov V (1992) Phenomenological kinetics of Frenkel defect recombination and accumulation in ionic solids. Rep Prog Phys 55:2079

Lépinoux J (2009) Modelling precipitation in binary alloys by cluster dynamics. Acta Mater 57:1086. https://doi.org/10.1016/j.actamat.2008.10.048

Mamivand M, Yang Y, Busby J, Morgan D (2017) Integrated modeling of second phase precipitation in cold-worked 316 stainless steels under irradiation. Acta Mater 130:94

Marian J, Bulatov VV (2011) Stochastic cluster dynamics method for simulations of multispecies irradiation damage accumulation. J Nucl Mater 415:84

Martin G (2006) Reconciling the classical nucleation theory and atomic scale observations and modeling. Adv Eng Mater 8:1231

Michaut B, Jourdan T, Malaplate J, Renault-Laborne A, Sefta F, Décamps B (2017) Cluster dynamics modeling and experimental investigation of the effect of injected interstitials. J Nucl Mater 496:166. https://doi.org/10.1016/j.jnucmat.2017.09.018

Moll S, Jourdan T, Lefaix-Jeuland H (2013) Direct observation of interstitial dislocation loop coarsening in α-iron. Phys Rev Lett 111:015503

Nichols FA (1978) On the estimation of sink-absorption terms in reaction-rate-theory analysis of radiation damage. J Nucl Mater 75:32

Ortiz CJ, Pichler P, Fühner T, Cristiano F, Colombeau B, Cowern NEB, Claverie A (2004) A physically based model for the spatial and temporal evolution of self-interstitial agglomerates in ion-implanted silicon. J Appl Phys 96:4866

Ortiz CJ, Caturla MJ, Fu CC, Willaime F (2007) He diffusion in irradiated α-Fe: an ab-initio-based rate theory model. Phys Rev B 75:100102(R)

Ovcharenko AM, Golubov SI, Woo CH, Huang H (2003) GMIC++: grouping method in C++: an efficient method to solve large number of Master equations. Comput Phys Commun 152:208

Robinson MT (1989) Slowing-down time of energetic atoms in solids. Phys Rev B 40:10717

Rouchette H, Thuinet L, Legris A, Ambard A, Domain C (2014) Quantitative phase field model for dislocation sink strength calculations. Comput Mater Sci 88:50

Singh BN, Golubov SI, Trinkaus H, Serra A, Osetsky YN, Barashev AV (1997) Aspects of microstructure evolution under cascade damage conditions. J Nucl Mater 251:107

Skinner BC, Woo CH (1984) Shape effect in the drift diffusion of point defects into straight dislocations. Phys Rev B 30:3084

Soneda N, Diaz de la Rubia T (2001) Migration kinetics of the self-interstitial atom and its clusters in bcc Fe. Philos Mag A 81:331

Stoller RE, Odette GR (1987) A composite model of microstructural evolution in austenitic stainless steel under fast neutron irradiation. In: Garner FA, Packan NH, Kumar AS (eds) Radiation-induced changes in microstructure: 13th international symposium (part I). ASTM STP 955. ASTM, Philadelphia, p 371

Surh MP, Sturgeon JB, Wolfer WG (2004) Master equation and Fokker-Planck methods for void nucleation and growth in irradiation swelling. J Nucl Mater 325:44

Terrier P, Athènes M, Jourdan T, Adjanor G, Stoltz G (2017) Cluster dynamics modelling of materials: a new hybrid deterministic/stochastic coupling approach. J Comput Phys 350:280. https://doi.org/10.1016/j.jcp.2017.08.015

Varvenne C, Mackain O, Clouet E (2014) Vacancy clustering in zirconium: an atomic-scale study. Acta Mater 65:65

Waite TR (1957) Theoretical treatment of the kinetics of diffusion-limited reactions. Phys Rev 107:463

Waite TR (1958) General theory of bimolecular reaction rates in solids and liquids. J Chem Phys 28:103

Wolfer WG, Mansur LK, Sprague JA (1977) Theory of swelling and irradiation creep. In: Bleiberg ML, Bennett JW (eds) Radiation effects in breeder reactor structural materials, The Metallurgical Society of AIME, New York, p 479

Xu D, Wirth BD, Li M, Kirk MA (2012a) Combining *in situ* transmission electron microscopy irradiation experiments with cluster dynamics modeling to study nanoscale defect agglomeration in structural materials. Acta Mater 60:4286

Xu D, Wirth BD, Li M, Kirk MA (2012b) Defect microstructural evolution in ion irradiated metallic nanofoils: kinetic Monte Carlo simulation versus cluster dynamics modeling and in situ transmission electron microscopy experiments. Appl Phys Lett 101:101905

Zouari M, Fournier L, Barbu A, Bréchet Y (2011) Cluster dynamics prediction of the microstructure evolution of 300-series austenitic stainless steel under irradiation: influence of helium. In: Busby JT, Ilevbare G, Andresen PL (eds) Proceedings of the 15th international conference on environmental degradation of materials in nuclear power systems – water reactors. Springer, Cham, p 1371

Experimental Validation of Models: *In Situ* TEM for Radiation Damage

102

Kazuto Arakawa and Michael P. Short

Contents

Abstract

In situ transmission electron microscopy (TEM) is a powerful technique for validating dynamic models of radiation-produced defects in metals. In this chapter, we examine key advances from the last ten years in *in situ* TEM

K. Arakawa (✉)
Department of Materials Science, Faculty of Science and Engineering, Shimane University, Matsue, Japan
e-mail: arakawa@riko.shimane-u.ac.jp

M. P. Short
Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: hereiam@mit.edu

which have enabled the validation of models and theories of radiation-induced defects, including TEM at the atomic scale and techniques for extracting defect information from nanoscale TEM. We conclude with an outlook and perspective on advances to come in the next ten years, expected to further aid in radiation damage model validation.

1 Introduction

Nuclear fission and fusion materials are primarily degraded by the accumulation of lattice defects produced upon knock-on displacement of host atoms under irradiation – point defects (vacancies and self-interstitial atoms (SIAs)) and their clusters (dislocation loops and cavities (voids and bubbles), etc.) (Was 2007). In understanding radiation damage, precise knowledge of its component processes – the production of defects and their movement and interaction with others – is crucial. Therefore, dynamics of radiation-produced defects in comparatively simple pure metals and model alloys have been extensively studied by modeling methods such as density functional theory (DFT), molecular dynamics (MD), kinetic Monte Carlo (kMC), and rate equations (RE) (Dudarev 2013; Marian et al. 2017). Needless to say, these modeling studies require experimental validation. Nowadays, *in situ* transmission electron microscopy (TEM) is recognized as a powerful technique for model validation. This chapter focuses on the use of *in situ* TEM for radiation-produced defects in "pure" metals and model alloys as a model validation technique. Following the description of the capabilities and limitations of the *in situ* TEM in Sect. 2, the key advances in the last ten years are provided in Sect. 3. Finally, a perspective of the outlook for the next ten years is given in Sect. 4.

2 Capabilities and Limitations of the *In Situ* TEM for Visualizing Radiation-Produced Defects

TEM is the most powerful technique for the observation of local structures such as defects and precipitates within solids with atomic to nanoscale spatial resolution. In *in situ* TEM (Fig. 1), the dynamic behavior of such local structures is probed in response to applied stimuli, such as heating, cooling, deformation, and irradiation with high-energy electrons, ions, and photons. Here, the typical imaging mechanisms of TEM and its capabilities and limitations are briefly provided to give context.

TEM enables both reciprocal-space (diffraction) and real-space imaging for the same region. It functions by passing an incident electron wave through a specimen, whereby the wave function of the transmitted electron wave at the exit surface of the specimen (exit wave) is Fourier transformed at the back focal plane of a magnetic objective lens. It is again Fourier transformed at the image plane of the objective lens. The distributions of the absolute square of the wave functions (intensities) at these planes correspond to the diffraction and real-space magnified image, respectively.

Fig. 1 Example of a transmission electron microscope (TEM) combined with a low-energy ion accelerator at Shimane University in Japan (Arakawa et al. 1999). Using this instrument, the dynamic behavior of local structures such as defects within solids can be monitored with nanoscale spatial resolution and recorded with 30 frames per second, in response to ion irradiation under heating, cooling, or deformation

Among various imaging methods, the most typical ones adopted for defect imaging are atomic high-resolution TEM (HRTEM) (De Graef 2003; Kirkland 2010) and nanoscale "low-resolution" bright-field and dark-field TEM (De Graef 2003; Hirsch et al. 1965; Jenkins and Kirk 2001). Normally, the specimens are illuminated by an almost ideal plane wave for conventional TEM. Instead, they are illuminated by a convergent electron beam for scanning TEM (STEM) (Pennycook and Nellist 2011). Even "low-resolution" bright-field and dark-field TEM and STEM are very useful and often utilized, because they are superior in terms of applicability to thicker specimens, wider areas, and temporal resolution. Principally, conventional

TEM is superior to STEM in terms of temporal resolution, although STEM also has remarkable advantages. Therefore, conventional TEM is generally utilized for *in situ* TEM probing dynamical behavior of defects. Hereafter, only conventional TEM is described.

2.1 Atomic High-Resolution TEM (HRTEM)

HRTEM can be adopted only for comparatively thin specimens less than a few 10's of nanometers in thickness. Multiple reflections in the diffraction signal are selected for imaging using a comparatively large objective aperture at the back focal plane of the objective lens. The transmitted plane wave at the exit surface of a specimen is modulated by the electrostatic potentials around individual atoms within the specimen. For very thin specimens with only a few to several atomic layers in thickness, only the phase of the electron wave is shifted by $\sigma V_p (x, y)$, where σ is the interaction constant being a function of the electron acceleration voltage and $V_p (x, y)$ is the projected potential onto the observation plane (x–y plane). This phase modulation is converted to an amplitude modulation by the objective lens. HRTEM is therefore called phase-contrast imaging. For a specific condition with a specific objective lens defocus (Scherzer defocus) and objective aperture radius (Scherzer frequency), the final image corresponds to the distribution of $1-2\sigma V_p (x, y)$, i.e., the projection of the atomic arrangement onto the observation plane. This image is called the structure image. However, for thicker specimens, both the phase and amplitude of the electron wave are modulated in a complex manner. Then, the image reflects only the spatial frequencies corresponding to the atomic arrangement. This image is called the lattice image. Therefore, interpretation of the HRTEM images often requires image simulations for model structures (Kirkland 2010). To the author's knowledge, the best spatial resolution recorded to date in HRTEM is 43 pm, which corresponded to the (633) lattice fringe of tungsten (Akashi et al. 2015). The HRTEM, therefore, can be applied to observation of the core structure of defects. In contrast, the direct detection of moving atomic-scale defects such as point defects has been performed only for very limited systems with only a few atomic layers of thickness (e.g., carbon nanotube) (Suenaga et al. 2007), and it has not been achieved for metals, to the author's knowledge.

2.2 Nanoscale "Low-Resolution" Bright-Field and Dark-Field TEM

Bright-field and dark-field TEM can be adopted for comparatively thick specimens from tens to hundreds of nanometers in thickness. Only a transmitted or excited diffracted reflection in the diffraction signal is selected for bright-field or dark-field TEM, respectively, using a comparatively small objective aperture at its back-focal plane. In normal bright-field and dark-field TEM, the electron incident direction deviates from precise low-index directions, and only systematic reflections (reciprocal lattice vectors, 0**g**, 1**g**, 2**g**, 3**g**, etc.) are highly excited. In the process

of electron waves transmitted through a specimen, only those corresponding to these reflections (ϕ_0, $\phi_{\mathbf{g}}$, $\phi_{2\mathbf{g}}$, $\phi_{3\mathbf{g}}$, etc.) interfere with one another, which is called dynamical diffraction. In the two-beam approximation, where only the transmitted wave ϕ_0 and a diffracted wave $\phi_{\mathbf{g}}$ are considered, the following Howie–Whelan equations on dynamical diffraction are established (inelastic scattering is neglected here, for simplicity):

$$\begin{cases} \frac{d\phi_0(x,y,z)}{dz} = \frac{i\pi}{\xi_0}\phi_0\left(x, y, z\right) + \frac{i\pi}{\xi_{\mathbf{g}}} \exp\left(-2\pi i\left(-sz - \mathbf{g}\cdot \mathbf{R}\left(x, y, z\right)\right)\right) \phi_{\mathbf{g}}\left(x, y, z\right) \\ \frac{d\phi_{\mathbf{g}}(x,y,z)}{dz} = \frac{i\pi}{\xi_{\mathbf{g}}} \exp\left(-2\pi i\left(sz + \mathbf{g}\cdot \mathbf{R}\left(x, y, z\right)\right)\right) \phi_0\left(x, y, z\right) + \frac{i\pi}{\xi_0}\phi_{\mathbf{g}}\left(x, y, z\right) \end{cases},$$

(1)

where $\mathbf{R}(x, y, z)$ is the displacement around a defect (z-coordinate along the depth in a specimen), ξ_0 and $\xi_{\mathbf{g}}$ are parameters called extinction distances being functions of the corresponding structure factor, and s corresponds to the deviation of the diffraction vector for $\phi_{\mathbf{g}}$ from $\mathbf{g}$ and means the deviation from the exact Bragg condition (Hirsch et al. 1965). These equations clearly show that defect contrast arises from the factor $\mathbf{g} \cdot \mathbf{R}(x, y, z)$ and only the component of $\mathbf{R}(x, y, z)$ being parallel to $\mathbf{g}$ contributes to defect contrast. The contrast in the image due to this mechanism is called diffraction contrast. The bright-field or dark-field images correspond to the distribution of $|\phi_0(x, y)|^2$ or $|\phi_{\mathbf{g}}(x, y)|^2$ at the exit surface of the specimen, respectively. For example, in imaging dislocations, what is imaged is spatially extended $\mathbf{R}(x, y, z)$ around a dislocation (Hirth and Lothe 1982). Even so, in weak-beam dark-field TEM, where intensity of the adopted reflection is weak under considerable deviation of the s value from zero, the spatial resolution can reach approximately 2 nm (Jenkins and Kirk 2001). In ordinary bright-field TEM, the spatial resolution is worse. The displacement field $\mathbf{R}(x, y, z)$ around a dislocation is primarily parallel to its Burgers vector, $\mathbf{b}$. The dislocation contrast, therefore, almost vanishes when $\mathbf{g} \cdot \mathbf{b} = 0$ is satisfied. Using this relationship, the direction of $\mathbf{b}$ can be determined. In imaging stacking faults, defect contrast also arises from $\mathbf{R}(x, y, z)$ of the stacking faults. In contrast, in imaging cavities, the local structure-factor reduction due to the presence of a cavity is imaged. It is therefore called structure-factor contrast (Jenkins and Kirk 2001).

2.3   Capabilities and Limitations of *In Situ* TEM

The dynamic behavior of defects can be monitored and recorded through a camera, in response to various stimuli. Temporal resolution normally ranges from 10 to 30 frames per second (fps) for ordinary cameras. Nowadays, even 400 fps can be achieved for high-speed and high-sensitivity cameras. As basic stimuli, heating and cooling are applicable. For commercially available specimen holders, the maximum temperature is approximately 1573 K, and the minimum temperature is approximately 10 K. For irradiation sources, electrons and ions are available (Hinks 2015). For *in situ* high-energy electron irradiation, high-voltage electron microscopes with

acceleration voltages higher than 1000 kV are applicable. For *in situ* ion irradiation, microscopes combined with ion accelerators are applicable (Fig. 1).

The most remarkable limitation for *in situ* TEM is the necessity of the use of thin-foil specimens, ranging from a few nanometers to a few hundred nanometers in thickness, as described above. In considerably thin specimens, the specimen surface can significantly influence the behavior of defects. For example, mobile defects can escape to the free surface. This is one of the major reasons why "low-resolution" bright-field and dark-field TEM suitable to comparatively thick specimens are very useful. The second limitation is that directly tracking considerably fast-moving defects is impossible, because of the limitation in temporal resolution. The third limitation is that directly probing moving atomic-scale defects is still practically impossible, as described above. Another limitation for the *in situ* TEM is the energy of electrons for observation, with 200 keV in kinetic energy for ordinary 200-kV microscopes. For example, even 200 keV electron irradiation of aluminum specimens causes knock-on displacements to produce additional point defects, thereby changing the image by measuring it. In *in situ* TEM experiments, these limitations should be considered.

3   Key Advances in the Last Ten Years in the *In Situ* TEM for Radiation-Produced Defects

Here, the key advances in the last ten years in the *in situ* TEM for radiation-produced defects in metals are briefly provided. For this purpose, typical successful examples are shown, for direct evaluation of the dynamic properties of slow-moving nanoscale defects and indirect evaluation of the dynamic properties of invisible atomic-scale defects.

3.1   Direct Evaluation of Dynamic Properties of Nanoscale Defects

As described above, in *in situ* TEM, temporal resolution is limited to a few tens of milliseconds at best. Also, practically, only nanoscale defects can be directly tracked. However, in some cases, motion of nanoscale defects is considerably slow, which can be directly tracked by normal *in situ* TEM. There, the most remarkable key advance in the last ten years is that dynamic behaviors of *individual* nanoscale defects have been observed to evaluate their dynamic properties, which can be compared with modeling results.

3.1.1   Motion of Nanoscale Defects and Their Interaction with Other Defects

One of the most successful examples is an experiment for 1D glide diffusion of SIA dislocation loops (Arakawa et al. 2007). This phenomenon is critical in the accumulation process of defects in metals upon irradiation. There had previously
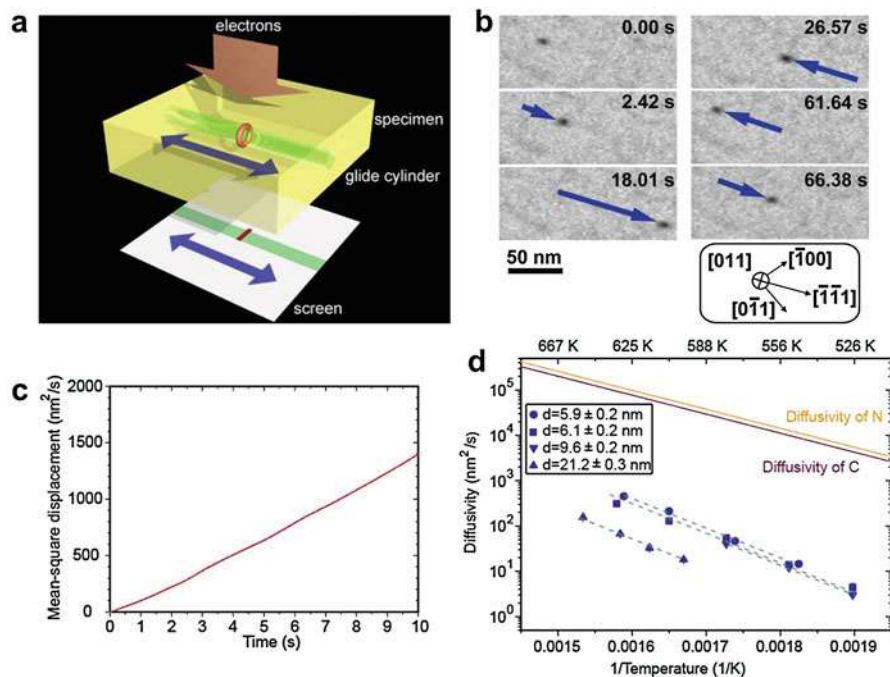
Fig. 2 TEM visualization of 1D glide diffusion of 1/2<111> SIA dislocation loops in high-purity iron (Arakawa et al. 2007). (**a**) Experimental setup of the TEM observation of a 1D-diffusing loop. (**b**) Bright-field TEM observation of an almost isolated 1D-diffusing 1/2[11$\bar{1}$] loop 5.9 nm in diameter at 575 K. (**c**) Mean squared displacement as a function of time for the loop shown in (**b**). (**d**) Arrhenius plot of the diffusivity of almost isolated 1/2<111> loops. Matching symbols correspond to the same loop with diameter *d*. The diffusivities of carbon and nitrogen in the matrix (Lord and Beshers 1966) are also drawn

been extensive MD (Marian et al. 2002; Osetsky et al. 2003; Soneda and Diaz de La Rubia 2001; Terentyev et al. 2007; Wirth et al. 2000) and theoretical studies (Barashev et al. 2000; Dudarev 2002) of this effect. Here, the diffusivity or activation energy for diffusion was directly compared between the TEM experiment and the MD calculations.

Figure 2 shows the TEM experiment on 1D diffusion of loops with **b** = 1/2<111> in high-purity iron (purity, 99.998 wt.%) (Arakawa et al. 2007). Figure 2a shows the experimental setup. In this experiment, the thin-foil effect on 1D diffusion was eliminated even for thin-foil specimens by setting the specimen surface to be parallel to the loop glide direction (**b** direction). Here, the effect of the electron irradiation for the observation was negligible, because its energy was only 200 keV, and therefore its associated PKAs were below the displacement threshold energy. Figure 2b shows an example of a bright-field TEM observation of an almost isolated 1/2[11$\bar{1}$] loop 5.9 nm in diameter at 575 K, which undergoes 1D diffusion along its **b** direction. By directly tracking the loop position with a time step of the frame

interval (1/30 s), the loop motion was analyzed. Figure 2c shows the mean squared displacement as a function of time for the loop, as shown in Fig. 2b. From the slope of the line in Fig. 2c, the diffusivity for this loop at this temperature can be directly evaluated. Figure 2d shows the Arrhenius plot of the evaluated diffusivity for various loops with diameter d. It clearly shows that the temperature dependence of the diffusivity satisfies Arrhenius' law. From the slope of the lines, the activation energy, E, for the diffusion can be evaluated. As a result, the E value was 1.3 eV, found to be independent of the loop size. These experimental results can be compared with corresponding MD results (Marian et al. 2002; Osetsky et al. 2003; Soneda and Diaz de La Rubia 2001; Terentyev et al. 2007; Wirth et al. 2000). The size independence of the diffusivity matches the MD results. The experimental E value (1.3 eV) is, however, considerably greater than the E value obtained by MD for perfectly pure metals (0.1 eV). The considerably high E value in the experiment was attributed to the effect of impurity atoms with high diffusivity such as carbon and nitrogen forming a Cottrell atmosphere around loops and following the loop motion (Fig. 2d), although it is still controversial.

This experimental technique has been also successfully applied to other cases, such as 2D climb diffusion of loops (Swinburne et al. 2016), elastic interaction among loops (Dudarev et al. 2010), collision processes between two loops (Arakawa et al. 2011; Arakawa et al. 2017), and the dynamical interaction of bubbles with grain boundaries (Ono et al. 2007) and collision cascades (Ono et al. 2009). Thus, it has become possible to make direct experimental validation of modeling results for dynamic properties of nanoscale defects under certain conditions.

3.1.2 Growth or Shrinkage of Nanoscale Defects

Ordinarily, bright-field and dark-field *in situ* TEM with a wide field of view is applied to visualize dynamic defects. In contrast, even *in situ* HRTEM with a narrow field of view can be applied to the evaluation of the growth or shrinkage processes of static defects.

Figure 3 is an example, which shows the atomistic growth process of a void in magnesium by absorption of vacancies produced under electron irradiation (Xu et al. 2013). This process was interpreted in terms of the vacancy diffusion process as calculated by MD (Xu et al. 2013).

This experimental technique is principally applicable to other defects. Actually, even using weak-beam dark-field TEM, the nucleation and propagation of unit jog at the periphery of SIA dislocation loops have been detected in the past (Arakawa et al. 2000). This could be extended to HRTEM.

3.2 Indirect Evaluation of Dynamic Properties of Invisible Atomic-Scale Defects

Directly tracking moving atomic-scale volumetric defects in metals is practically impossible, unless the specimen is so thin that defects can too easily reach free surfaces. Understanding dynamic behaviors of these invisible defects behind visible-
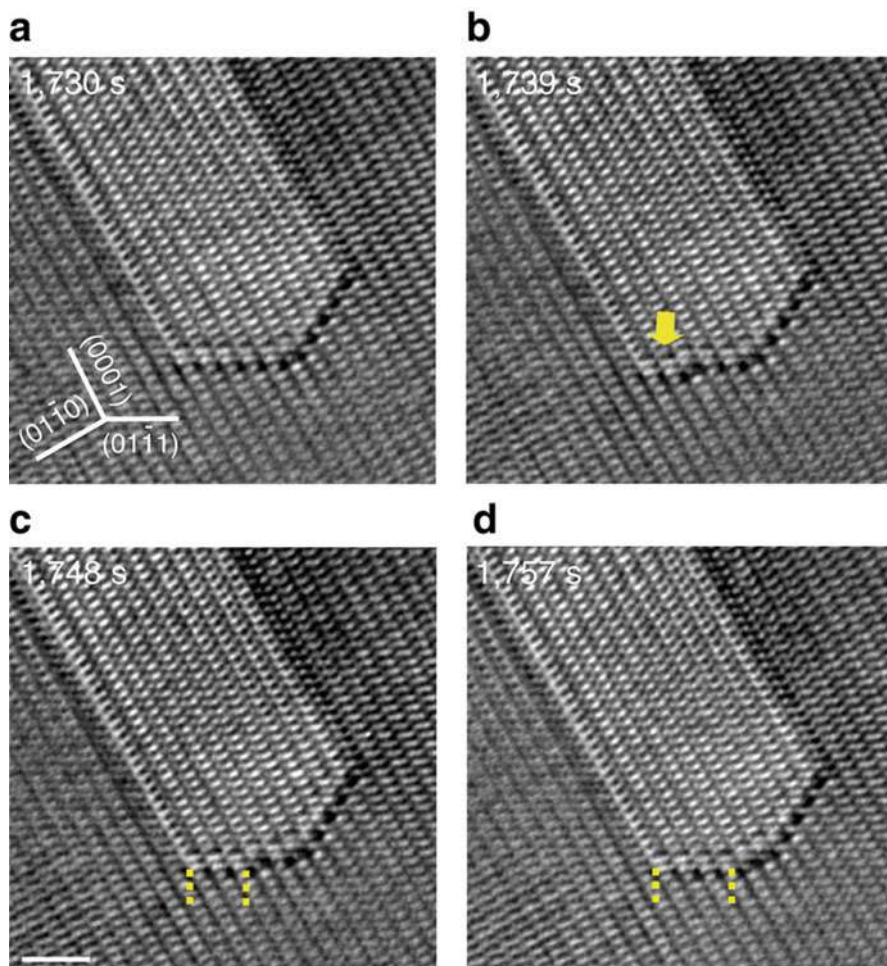
Fig. 3 HRTEM observation of the void growth process in magnesium under 200 keV electron irradiation at room temperature (Xu et al. 2013). (**a**) Facet plane indexes. (**b**) A vacancy layer nucleated as indicated by the *yellow arrow*. (**c**) The vacancy layer extended on the facet. (**d**) The growth of the facet. Scale bar, 2 nm

defect accumulation is, however, considerably important and feasible. To extract the dynamic properties of these invisible defects, the formation and growth processes of visible nanoscale point defect clusters under irradiation have been compared between TEM experimental results and corresponding modeling results. In the past, the primary modeling technique for such comparison has been rate theory based on mean-field theory (Kiritani 1994). However, in the last ten years, even kMC has been often adopted to such comparisons. kMC takes into account even spatial correlations among defects (Amino et al. 2011), and it is more suitable to quantitative comparisons. Here, examplary studies in which *in situ* TEM is used to monitor or infer defect evolution or migration are shown.

3.2.1 Fast-Moving Point Defects

In high-energy electron irradiation, only point defects (vacancies and SIAs) are principally produced as primary damage, and the production of point defects is almost spatially homogeneous (Was 2007). Therefore, high-energy electron irradiation is suitable to extract dynamic properties of point defects.

Figure 4 shows the comparison between a TEM experiment and corresponding kMC for the formation and growth process of SIA dislocation loops in high-purity (purity, 99.9999 wt.%) tungsten at 16 K under 2000 keV electron irradiation (Amino et al. 2016). At this temperature, vacancy thermal migration is frozen, and the moving defects are only SIAs. Figure 4a shows weak-beam dark-field TEM images of the loops for different irradiation times for a beam flux of 3.0×10^{22} m^{-2}s^{-1}. From these TEM photographs, both average size and number density of loops were measured as a function of irradiation time. These experimental data were compared with kMC. In the kMC simulation, almost all the parameters for the dynamic properties of SIAs were regarded as fitting parameters, such as the migration dimension (3D or 1D), migration activation energy (E_m^I), and reaction radius values
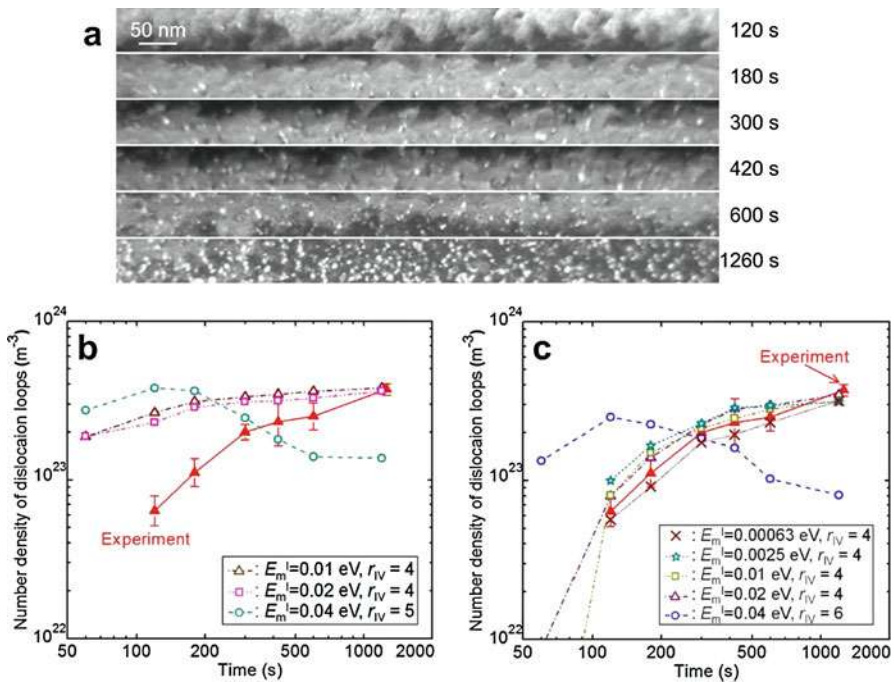
Fig. 4 Comparison between TEM and kMC for the formation and growth of SIA dislocation loops in high-purity tungsten under 2000-keV electron irradiation (Amino et al. 2016). (**a**) Weak-beam dark-field TEM images of loops for different irradiation times. (**b**) Comparison between the experiment and kMC for temporal variation in loop number density. In the kMC, the SIA migration dimension is 3D. (**c**) Comparison between the experiment and kMC for temporal variation in loop number density. In the kMC, the SIA migration dimension is 1D

with an impurity atom (r_{IX}), a vacancy (r_{IV}), and another SIA (r_{II}). Figures 4b, and 4c show examples for the comparison. The 3D migrating SIA never reproduces the experimental result (Fig. 4b). In, contrast, simulating 1D migrating SIAs reproduces the experimental result for the other parameters within a limited range (Fig. 4c). This result matches a DFT result showing that the most stable SIA structure is a crowdion undergoing 1D migration (Nguyen-Manh et al. 2006). The range within the parameter space reproducing the experimental results was considerably narrow. Thus, the parameters for the SIA dynamics can be fixed, through the comparison between TEM and kMC.

3.2.2   Collision Cascade Defects

In the primary damage process of ion and neutron irradiation, point defect clusters are also generated as well as point defects, which is called the collision cascade (Was 2007). The collision cascade has never been directly observed by any experiment. Its structure calculated by MD depends on the interatomic potentials chosen (Terentyev et al. 2006), the correct choice and quality of which are often unresolved matters of great debate. Therefore, modeling the defect accumulation processes under collision cascade is less straightforward. Even so, some efforts for the comparison between TEM and kMC (and cluster dynamics) for ion irradiation have been performed. For example, validation of the dynamic properties of invisible small defects has been attempted (Xu et al. 2012). Also, the trapping centers of intrinsically fast-moving defects were examined (Mason et al. 2014). These types of studies are growing in number, and represent an area of considerable growth for using *in situ* TEM to validate primary radiation damage. Ultimately a carefully planned suite of *in situ* TEM experiments could even help inform the choice of the best MD interatomic potentials to use, further boosting the efficacy of more complex simulations not as easily validated by experiments.

4   Outlook of the Next Ten Years of the *In Situ* TEM for Radiation-Produced Defects

Efforts for reducing the limitations of the TEM are expected, as increasing especially the time resolution of *in situ* TEM represents a way to both acquire more data for existing experiments and temporally resolve new regimes of radiation defect accumulation. For example, temporal resolution will be significantly improved using microscopes with pulsed electron sources. The ideal spatial resolution for a 15-ns pulse is approximately 1 nm (LaGrange et al. 2012). This type of high-speed technique will be applied to fast-moving nanoscale defects (Arakawa et al. 2014). Efforts for directly observing dynamic behaviors of atomic-scale defects are expected, by improving the signal-to-noise ratio in the image. For example, this new technique will be applied to the nucleation process of point defect clusters.

In addition to simply improving the resolution in both time and space of the *in situ* TEM, carefully coordinated sets of experiments can be used to rapidly resolve current scientific questions and matters of debate. For example, a set of

TEM observations on the self-ion irradiation of pure tungsten after one second of irradiation (Yi 2015) has garnered two competing explanations for the observed cluster size distribution (Sand 2017, Jin 2018). The former explains the defect size distribution using primary radiation damage and sub-cascade splitting, while the latter uses cluster dynamics and rate theory to explain the deviation from the power law size distribution. Neither study can be confirmed right or wrong without more data. An *in situ* study, acquiring sufficiently resolved TEM images with tens of millisecond resolution over a period of minutes, could help to determine whether dynamical effects are responsible for the cluster size distribution or not.

The example above may have other ramifications for the choice of MD inter-atomic potentials, constantly a matter of debate. The sheer data throughput rate of *in situ* TEM can generate enough information to apply techniques of machine learning, first to decide which available potential is most suitable to reproduce an experiment, then as a training database to iteratively design better potentials for even better experimental reproduction.

Finally, the combination of direct visualization of radiation-induced defects in the *in situ* TEM with complimentary *in situ* measurements of material property evolution can yield ultra-high throughput structure-property relationships in irradiated materials. For example, studies such as those of Xu et al (Xu 2013) which directly visualize void growth could be combined with new techniques such as *in situ* transient grating spectroscopy (TGS) of the thermo-elastic properties of irradiated metals (Dennett 2019). Theories relating changes in elasticity due to defect structures can be combined with the directly visualized structures themselves, helping to reconcile theory, simulation, and experiment at once. As these capabilities are further developed and enhanced, it will always fall to the *in situ* TEM to provide the direct structural visualization to map new structure-property relationships at unprecedented resolution.

References

Akashi T, Takahashi Y, Tanigaki T, Shimakura T, Kawasaki T, Furutsu T, Shinada H, Müller H, Haider M, Osakabe N, Tonomura A (2015) Aberration corrected 1.2-MV cold field-emission transmission electron microscope with a sub-50-pm resolution. Appl Phys Lett 106:074101

Amino T, Arakawa K, Mori H (2011) Reaction rate between 1D migrating self-interstitial atoms: an examination by kinetic Monte Carlo simulation. Philos Mag 91:3276–3289

Amino T, Arakawa K, Mori H (2016) Detection of one-dimensional migration of single self-interstitial atoms in tungsten using high-voltage electron microscopy. Sci Rep 6:26099

Arakawa K, Tsukamoto T, Tadakuma K, Yasuda K, Ono K (1999) In-situ observation of the microstructural evolution in germanium under the low-energy helium ion irradiation. J Electron Microsc 48:399–405

Arakawa K, Satoh Y, Arai S, Kiritani M (2000) Fluctuation of point defect reactions observed during the growth of dislocation loops under electron irradiation. Philos Mag A 80:2041–2055

Arakawa K, Ono K, Isshiki M, Mimura K, Uchikoshi M, Mori H (2007) Observation of the one-dimensional diffusion of nanometer-sized dislocation loops. Science 318:956–959

Arakawa K, Amino T, Mori H (2011) Direct observation of the coalescence process between nanoscale dislocation loops with different Burgers vectors. Acta Mater 59:141–145

Arakawa K, Amino T, Mori H (2014) One-dimensional glide motion of "naked" 1/2<111> prismatic dislocation loops in iron. ISIJ Int 54:2421–2424

Arakawa K, Amino T, Yasuda H, Mori H (2017) Dislocation junction formation via the coalescence between nanoscale 1/2<111> prismatic dislocation loops in iron. ISIJ Int 57:2065–2069

Barashev AV, Osetsky YN, Bacon DJ (2000) Mechanism of one-dimensional glide of self-interstitial atom clusters in α-iron. Philos Mag A 80:2709–2720

De Graef M (2003) Introduction to conventional transmission electron microscopy. Cambridge University Press, Cambridge

Dennett CA, Buller D, Hattar K, Short MP (2019) Real-time thermomechanical property monitoring during ion beam irradiation using *in situ* transient grating spectroscopy. Nucl Instrum Meth B 440:126–138

Dudarev SL (2002) Thermal mobility of interstitial defects in irradiated materials. Phys Rev B 65:224105

Dudarev SL (2013) Density functional theory models for radiation damage. Annu Rev Mater Res 43:35–61

Dudarev SL, Gilbert MR, Arakawa K, Mori H, Yao Z, Jenkins ML, Derlet PM (2010) Langevin model for real-time Brownian dynamics of interacting nanodefects in irradiated metals. Phys Rev B 81:15

Hinks JA (2015) Transmission electron microscopy with in situ ion irradiation. J Mater Res 30:1214–1221

Hirsch PB, Howie A, Nicholson RB, Pashley DW, Whelan MJ (1965) Electron microscopy of thin crystals. Butterworths, London

Hirth JP, Lothe J (1982) Theory of dislocations. Wiley, New York

Jaime M, Charlotte SB, Christophe D, Sergei LD, Mark RG, Richard JK, Daniel RM, Kai N, Andrea ES, Lance LS, Tomoaki S, Brian DW (2017) Recent advances in modeling and simulation of the exposure and response of tungsten to fusion energy conditions. Nucl Fusion 57:092008

Jenkins ML, Kirk MA (2001) Characterization of radiation damage by transmission electron microscopy. Institute of Physics, Bristol/Philadelphia

Jin M, Permann C, Short MP (2018) Breaking the Power Law: Multiscale Simulations of Self-Ion Irradiated Tungsten. J Nucl Mater 504:33–40 (2018)

Kiritani M (1994) Microstructure evolution during irradiation. J Nucl Mater 216:220–264

Kirkland EJ (2010) Advanced computing in electron microscopy, 2nd edn. Springer, New York

LaGrange T, Reed BW, Santala MK, McKeown JT, Kulovits A, Wiezorek JMK, Nikolova L, Rosei F, Siwick BJ, Campbell GH (2012) Approaches for ultrafast imaging of transient materials processes in the transmission electron microscope. Micron 43:1108–1120

Lord AE, Beshers DN (1966) The mechanical damping of iron from room temperature to 400°C at 7 megacycles/sec. Acta Metall 14:1659–1672

Marian J, Wirth BD, Perlado JM (2002) Mechanism of formation and growth of <100> interstitial loops in ferritic materials. Phys Rev Lett 88:255507

Mason DR, Yi X, Kirk MA, Dudarev SL (2014) Elastic trapping of dislocation loops in cascades in ion-irradiated tungsten foils. J Phys Condens Matter 26:375701

Nguyen-Manh D, Horsfield AP, Dudarev SL (2006) Self-interstitial atom defects in bcc transition metals: group-specific trends. Phys Rev B 73:020101(R)

Ono K, Miyamoto M, Arakawa K (2007) Dynamical interaction of helium bubbles with grain boundaries in Fe and Fe–9Cr ferritic alloy. J Nucl Mater 367–370:522–526

Ono K, Miyamoto M, Arakawa K, Birtcher RC (2009) Dynamical interaction of helium bubbles with cascade damage in Fe–9Cr ferritic alloy. J Nucl Mater 386–388:177–180

Osetsky YN, Bacon DJ, Serra A, Singh BN, Golubov SI (2003) One-dimensional atomic transport by clusters of self-interstitial atoms in iron and copper. Philos Mag 83:61–91

Pennycook SJ, Nellist PD (eds) (2011) Scanning transmission electron microscopy imaging and analysis. Springer, Berlin

Sand AE, Mason DR, Backer AD, Yi X, Dudarev SL, Nordlund K (2017) Cascade fragmentation: deviation from power law in primary radiation damage. Mater Res Lett 5:357–363. https://doi.org/10.1080/21663831.2017.1294117

Soneda N, Diaz de La Rubia T (2001) Migration kinetics of the self-interstitial atom and its clusters in bcc Fe. Philos Mag A 81:331–343

Suenaga K, Wakabayashi H, Koshino M, Sato Y, Urita K, Iijima S (2007) Imaging active topological defects in carbon nanotubes. Nat Nanotechnol 2:358–360

Swinburne TD, Arakawa K, Mori H, Yasuda H, Isshiki M, Mimura K, Uchikoshi M, Dudarev SL (2016) Fast, vacancy-free climb of prismatic dislocation loops in bcc metals. Sci Rep 6:30596

Terentyev D, Lagerstedt C, Olsson P, Nordlund K, Wallenius J, Becquart CS, Malerba L (2006) Effect of the interatomic potential on the features of displacement cascades in α-Fe: a molecular dynamics study. J Nucl Mater 351:65–77

Terentyev DA, Malerba L, Hou M (2007) Dimensionality of interstitial cluster motion in bcc-Fe. Phys Rev B 75:13

Was GS (2007) Fundamentals of radiation materials science. Springer, Berlin

Wirth BD, Odette GR, Maroudas D, Lucas GE (2000) Dislocation loop structure, energy and mobility of self-interstitial atom clusters in bcc iron. J Nucl Mater 276:33–40

Xu D, Wirth BD, Li M, Kirk MA (2012) Defect microstructural evolution in ion irradiated metallic nanofoils: kinetic Monte Carlo simulation versus cluster dynamics modeling and in situ transmission electron microscopy experiments. Appl Phys Lett 101:101905

Xu W, Zhang Y, Cheng G, Jian W, Millett PC, Koch CC, Mathaudhu SN, Zhu Y (2013) In-situ atomic-scale observation of irradiation-induced void formation. Nat Commun 4:2288

Yi X, Sand A, Mason D, Kirk M, Roberts S, Nordlund K, Dudarev S (2015) Euro Phys Lett 110:36001

Modeling Radiation-Induced Segregation and Precipitation: Contributions and Future Perspectives from Artificial Neural Networks

103

Nicolas Castin and Lorenzo Malerba

Contents

Abstract

Radiation-induced segregation and precipitation is one of the main factors for the changes in macroscopic properties taking place in steels under irradiation. For instance, reactor pressure vessel steels and ferritic-martensitic steels are known to harden and embrittle within the lifetime of the reactors. The development

N. Castin (✉)
The Belgian Nuclear Research Centre (SCK·CEN), Nuclear Materials Science (NMS) Institute, Structural Materials (SMA) Expert Group, Mol, CP, Belgium
e-mail: nicolas.m.b.castin@gmail.com; nicolas.castin@sckcen.be

L. Malerba
The Belgian Nuclear Research Centre (SCK·CEN), Nuclear Materials Science (NMS) Institute, Structural Materials (SMA) Expert Group, Mol, CP, Belgium

Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avda., Spain
e-mail: lorenzo.malerba@sckcen.be

of quantitative predictive models for these changes is a very challenging task, because of the highly multiscale nature of the mechanisms taking place at the microstructural level and the chemical complexities involved. Complete physical-based approaches such as object kinetic Monte Carlo models are promising tools to take this challenge, but their parametrization must be accurately and adequately elaborated. In this chapter, we revise how highly powerful and flexible numerical tools offered by machine learning systems, specifically artificial neural networks, have contributed to these models. Perspectives for future developments are also discussed, in light of the state-of-the art modeling activities in the community.

1   Introduction

Irradiation with bombarding particles results in a continuous creation of point defects in the bulk of the studied material, i.e., vacancies and self-interstitial atoms (SIA). The increased concentration of vacancies thus enhances the kinetics of diffusion-driven process (such as precipitation of insoluble species) but might also alter the thermodynamic equilibrium. Moreover, SIA exhibit a profoundly different behavior from vacancies, potentially giving rise to processes not normally observed in materials. For instance, reactor pressure vessel steels and ferritic-martensitic steels are well known to harden and embrittle under neutron irradiation (Petrequin 1995), potentially limiting the lifetime of the power plants. Today, experimental evidence (Bergner et al. 2014) demonstrated that the main responsible for this degradation are nano-sized diffuse solute clusters, invisible to the transmission electronic microscope (TEM) but observable with the atom probe tomography (APT) (Kuksenko et al. 2013; Meslin et al. 2010), which are fully radiation-induced features.

Because it all starts at the atomic level, only a multiscale strategy can derive thorough and nonempirical physical models capable of describing the mechanisms leading to the formation of these nano-sized embrittling solute clusters. Figure 1 depicts an overview of a modeling approach based on kinetic Monte Carlo (KMC) methods, which is one of the possible choices to bridge the gap from the electronic to coarse-grained level. Interested readers are directed to general overviews for alternative approaches (Becquart et al. 2010; Becquart and Domain 2011). In the lowest left corner, two fundamental sources of input data are (a) calculations based on first-principle physics to address the electronic structure, most often using the density functional theory (DFT), and (b) any kind of experimental evidence, e.g., formation energies, diffusion coefficients, phase diagrams, etc. Unquestionably, these remain the only true final judge on any model prediction. The story ends at the coarse-grained level, which is in this work implemented by the object kinetic Monte Carlo (OKMC) method. There, the processes undergone by the defects injected during irradiation are described as thermally activated events that correspond, for the objects involved, generally clusters of point defects, to specifically designed laws
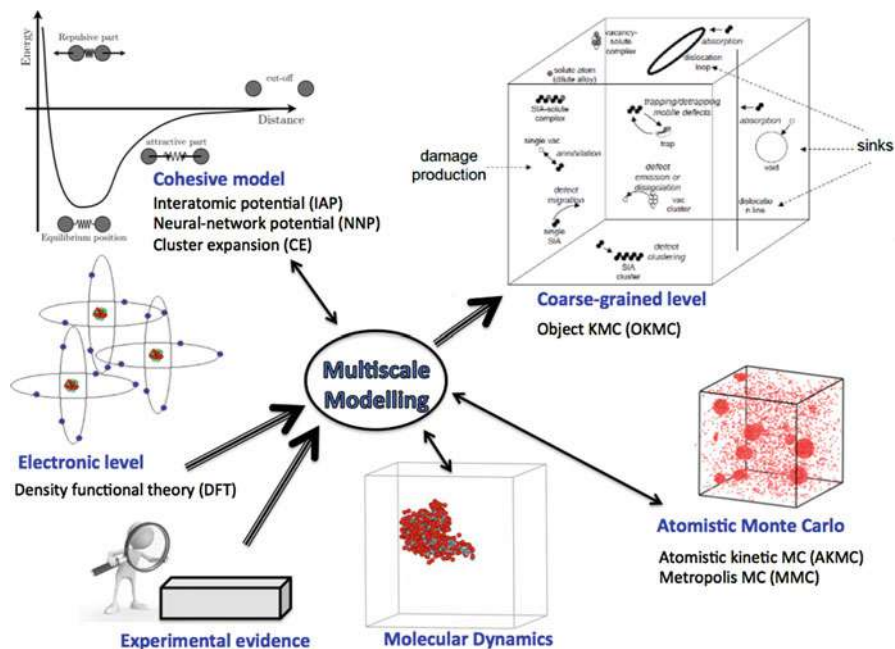
Fig. 1 Multiscale modeling approach for describing the effects of irradiation in structural materials

for mutual interactions, mobility, and stability. A substantial number of reactions must be carefully parametrized in OKMC models, which are achieved summoning a variety of methods and models at lower scales, ranging from DFT, molecular dynamics, and atomistic kinetic Monte Carlo (AKMC).

OKMC models have been widely used to describe the microstructural evolution in materials under irradiation, in terms of radiation defects, i.e., vacancies, SIA, and their clusters. For example, a model of this type was successfully developed (Jansson and Malerba 2013) to simulate irradiation processes in Fe-C systems, as reference system for models addressing steels, i.e., with a more complex composition. However, the only objects explicitly treated were point defects and their clusters. The parameters describing their migration and dissociation were specific for Fe, and the effect of C was effectively introduced in terms of traps for mobile clusters, without explicitly introducing C atoms. Likewise, models of this type have been extended to model chemically more complex systems, e.g., Fe-Mn-Ni as representative of RPV steels (Chiapetto et al. 2015a) or FeCr as representative of ferritic-martensitic steels (Chiapetto et al. 2015b). The effect of the presence of solutes was implicitly introduced in terms of changes of parameters due to the presence of solutes, without explicitly introducing them in the simulation box. This was achieved in a simplified way, assuming that the solute atoms are always uniformly and randomly distributed in the simulation volume, so that their effect

is inherently independent of local fluctuations of composition. Clearly, the solutes must be explicitly included in future models for describing the formation of their clusters.

Playing the role of reference metrics for the physical description of the studied material, vital inputs in multiscale modeling approaches are, undoubtedly, cohesive models. Given a set of atoms described by their coordinates in space and their chemical species, their task is to provide a numerical value for the total energy of the system in any configuration. They thus determine the driving forces, e.g., solute transports by point defects, and eventually the formation of solute-rich clusters. In a quantum mechanism framework, ab initio methods, e.g., DFT, are, for many good reasons, regarded as the most reliable choice to explore the energy landscape of the system. There are, however, limitations. Firstly, some degree of empiricism remains. Unfortunately, a blind use of DFT may, in some cases, provide imperfect predictions (Olsson et al. 2010). Secondly, even with state-of-the-art computing facilities, using DFT still implies a huge cost in CPU resources which is often unaffordable and, sometimes, unwarranted. For these reasons, alternative cohesive models are still extensively used, empirical interatomic potentials (IAP) being a very popular example. On the one hand, in addition to the largely reduced computing cost, the main advantage of these potentials is the possibility to achieve a tunable compromise between various target properties. It can thus mix DFT-originated with experimental-originated data, as conveniently as required. For example, an FeCu potential (Pasianot and Malerba 2007) was fitted to faithfully reproduce the experimentally observed Cu solubility limit in Fe, whereas DFT is known to underestimate it (Messina et al. 2017). The affordable computing time allows large and complex systems, containing up to several million atoms, to be studied. On the other hand, the simplicity proper to IAP also implies an intrinsic limitation in their capacities of making accurate prediction of various sets of properties at the same time.

Suitable low-computing-cost cohesive models allow the gap between the basic atomic level and the realm of longer timescale Monte Carlo (MC) models to be bridged. Many more gaps remain to bridge, however. Even using IAP, the precise calculation of the energy barriers the system has to overcome through thermal activation, which are at the core of MC methods, poses severe computing time limitations. Repetitive and heavy routine calculations are indeed required at key steps of a MC simulation, significantly limiting the range of applicability of the method. Similar limitations arise in connection with storage and integration of data/knowledge moving from one scale to a higher one. An illustrative example is the calculation of migration energies of single point defects in a chemically changing environment. These need to be routinely calculated using a dedicated method in order to let a Monte Carlo simulation that describes diffusion-driven processes to progress, but they entail a prohibitive computational cost if the calculation is performed in a precise way each time. Bridging the gap between models thus here means finding a numerical solution for either speeding up the calculation of the migration energies or, from another standpoint, extrapolating knowledge in the MC model from a limited amount of examples.

Machine learning techniques, specifically artificial neural networks (ANN), are very promising tools for taking this challenge. These tools have been progressively coupled to atomistic Monte Carlo models to improve their physical reliability, within affordable computing loads (Castin and Malerba 2010; Castin et al. 2014). ANN are flexible and powerful regression techniques, capable to assimilate complex and inexpressible laws of interactions from a limited number of explicit examples. Once the ANN are properly designed, they provide the values required for the model at a very limited computing cost, thanks to their mathematical simplicity. The remaining of this chapter is organized as follows. In Sect. 2, we start by providing a basic description of ANN techniques and the underlying mathematical framework. Next, in Sect. 3, we describe how ANN were used to create enhanced AKMC tools, providing novel and accurate models for solute transport and segregation. Both rigid-lattice and lattice-free concepts are presented. Similar mathematical concepts to the latter are then further exploited in Sect. 4 to define ANN-based cohesive models, directly fitted from DFT. Finally, we provide visions and recommendations for future perspectives in Sect. 5.

2   Artificial Neural Networks: A Practical Form of Artificial Intelligence

Artificial neural networks (ANN) are a concept of *weak artificial intelligence*, in a group of paradigms often denoted to as *machine learning*, or *computational intelligence*. They can be seen as self-learning systems aimed at extracting hidden knowledge from their environment, in order to take relevant actions when entirely new situations are encountered. For the particular application to the design of numerical predicting tools, for instance, they are *surrogate models* sparing the user the (otherwise fastidious) need to explicitly formulate knowledge about the problem at hand.

A complete description of the theoretical and mathematical frameworks of ANN goes beyond the scope of this paper. For a complete and detailed reference textbook, interested readers are directed to Bishop (1995) or to a more concise overview in Bishop (1994).

2.1   The Multilayer Perceptron: A Universal Approximation Machine

The so-called *multilayer perceptron* depicted in Fig. 2 provides appropriate solutions for the design of general numerical regressions. Inspired from biological neural networks such as the human brain, the fundamental idea is to create a network of simple processing units (either called neurons or nodes in the literature), thus constructing a sophisticated and complex response out of a set of simple individual rules. For the sake of simplicity, at least from the mathematical point of view, ANN are typically constituted of organized layers of nodes. On the left-hand side in Fig. 2,
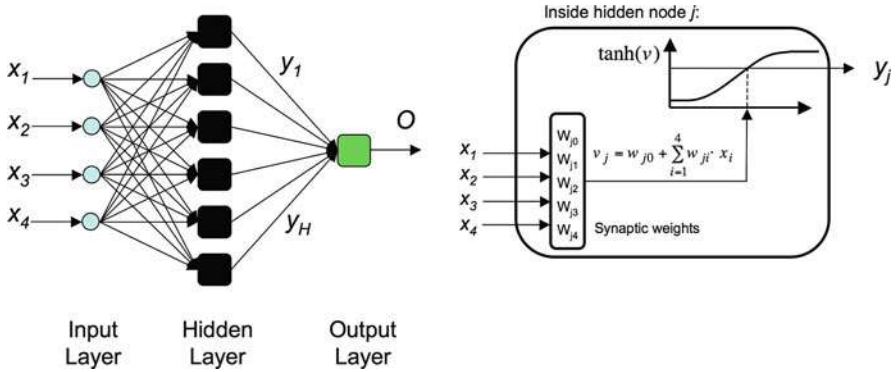
Fig. 2 Example of feed-forward multilayer perceptron with four input variables (x_1, x_2, x_3, and x_4), one hidden layer with $H = 6$ nodes, and one output (O). The right part of the figure shows the detail inside a node of the hidden layer. The signal is propagated from the left to the right of the network, in a layer-by-layer fashion

the input layer is the collection of the raw input signals for the whole network. On the right-hand side, the last layer is called the output layer, providing the answer of the network. In between, several intermediate layers may be introduced (there is only one in the figure). They are called *hidden layers*, for the reason that they are, in practice, invisible to the user. There is in theory no restriction about how nodes of different layers can be connected to each other. For simplicity, many networks do not allow backward connections: a given node never receives as input-output signals coming from nodes of either the same layer or from the next layers. Such a network is qualified as *feed-forward*. Lastly, the network shown in the figure is said to be *fully connected*, because all nodes in a given layer receive as input all output signals coming from the immediately preceding layer, and no *layer bypass* is allowed. The inputs for the first (and only) hidden layer are the raw input signals of the network.

The interest of the feed-forward multilayer perceptron, in its simplest form as depicted in Fig. 2, is that it fulfills the *universal approximation theorem* (Hornik et al. 1989): for any continuous function $F(x_1, \ldots, x_n)$, there exists an ANN with a finite number of hidden nodes that fulfill the following condition, for all sets of inputs x_i and all ε:

$$|F(x_1, \ldots, x_n) - O(x_1, \ldots, x_n)| < \varepsilon \tag{1}$$

Here, $O(x_1, \ldots, x_n)$ is the output of the ANN. The multilayer perceptron can therefore, in theory, be regarded as a universal approximation machine capable of perfectly assimilating any problem with a numerical character. It is safe and reasonable to assume that such degree of idealness cannot be dreamed of in practice, unless perhaps for some academical examples. Concretely, it is the authors' opinion that the flexibility and generality of the ANN internal structure, avoiding the user to explicitly formulate knowledge about the problem at hand, come at

an unavoidable cost. Indeed, ANN might be very successful in understanding a given problem from a limited amount of examples, but expecting it to perform a truly general understanding is hazardous. In other words, it is safe to assume that ANN is inherently devoted (not to say limited) to interpolation problems, whereas the extrapolation outside the domain of the input space covered during training is uncertain. Therefore, during any application of ANN, it is important to make sure that the domain of applicability can be determined. This can in fact be delicate in some cases, as discussed later.

2.2   Supervised Training

In this work, ANN are designed to implement numerical regression tools. This is therefore a *supervised training problem*, and training can be regarded as an optimization problem that consists in the minimization of the following objective function:

$$f(N_H, \bar{w}) = \sum_{i=1}^{N_T} (o_i(N_H, \bar{w}) - d_i)^2 \qquad (2)$$

Here N_T is the number of available examples of I/O for training, d_i is the desired output for the example i, and o_i is the corresponding prediction by the network. The latter is a function of the ANN architecture, i.e., the number N_H of nodes in the hidden layer and, finally, the vector $\bar{w}$ of synaptic weights. Assuming that N_H is fixed, the minimization of function f is thus the problem of determining the optimal numerical value of the synaptic weights $\bar{w}$: it can be undertaken by any classical method for nonlinear optimization. In this work, we used the method proposed by Levenberg (1994) and Marquardt (1963) (LM).

A major concern while training ANN is to guarantee that the predictions for new sets of inputs are equally accurate, compared to predictions on the available set of examples used for training. Indeed, without control, there is a risk that the ANN does not develop a general logic but in reality rather memorizes the complete set of available examples, as illustrated in Fig. 3. We see that reasonable interpolation is achieved by the ANN if the latter is not too complex, i.e., if N_H is small. Predictions are however not equally accurate for all training examples. If the number of hidden nodes is increased, the ANN manages more accurate predictions for all known points but clearly loses generality. This pathology cannot be easily identified only on the basis of a limited amount of examples of I/O, especially if the dimensionality of the problem at hand is large. ANN training must therefore be *regularized*. A common regularization approach is called *early stopping* and is based on the idea that memorization of the provided examples, or more generally can we say *network overspecialization*, develops only at a certain moment of training, i.e., after a certain number of training iterations, called *epochs* in ANN jargon. The most natural way to prevent it is therefore to divide the available table of examples of I/O into two
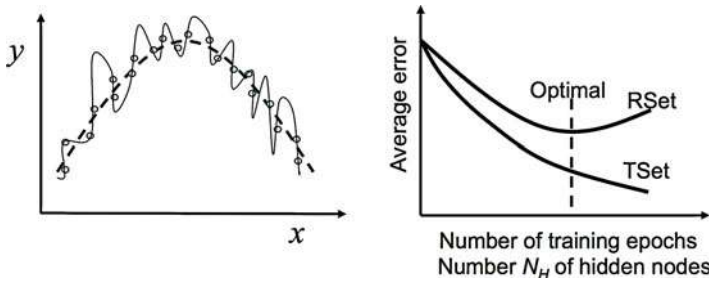
Fig. 3 (Left) Schematic illustration of ANN overspecialization for a simple problem with one variable x and one output y. Blue points represent data in the training set. The dashed line shows predictions by an ANN with a small number N_H of hidden nodes, whereas the plain line shows predictions of another ANN where too many hidden nodes were introduced. (Right) Typical evolution of the average error committed on the training set (TSet) and reference set (RSet) with increasing number of training epochs (N_H being fixed) or with increasing N_H (at the end of training)

different and nonoverlapping sets: (a) The training set is used to minimize function f in Eq. (2). Only these examples are thus used to calculate the gradients with the LM algorithm, and the synapses are updated during each epoch taking only them into account. (b) The reference set is used to measure, after every epoch, the average error of prediction on new cases. Figure 3 shows the typical evolution of the average error of prediction on both sets during the training epochs. The error committed on the training set always decreases. The error on the reference set, however, ceases to decrease from a certain epoch and then starts to increase, as a clear sign of the onset of overspecialization. Training is therefore interrupted at that moment.

At this stage, the determination of the optimal network architecture is still an open question. In our applications, we invariably found optimal architectures using no more than one hidden layer in the ANN. Determining the optimal architecture is thus a mono-parametric study: networks with increasing N_H are trained separately, and the one committing the lowest error on the reference set is finally retained, as depicted in Fig. 3. Too small a N_H understandably leads to higher errors of prediction, because not enough degrees of freedom are available in the network. Oppositely, too high a N_H increases the risk of overspecialization, and the error on the reference set increases as well.

3 Enhanced Atomistic Kinetic Monte Carlo Models

Atomistic kinetic Monte Carlo (KMC) methods are widespread simulation tools dedicated to describe diffusion-controlled phenomena at the atomic level. Generally based on a rigid-lattice approach, they feature an explicit spatial characterization of the diffusion of lattice defects and atoms, enabling a detailed investigation of the kinetics of formation of fine microstructural features. The evolution of the alloy proceeds through migration events of single defects (vacancies and/or interstitials)

(Young and Elcock 1966; Bortz et al. 1975; Fichthorn and Weinberg 1991), which are stochastically selected at each step based on their transition rates:

$$\Gamma = \Gamma_0 \cdot \exp\left(\frac{-E_m}{k_B T}\right) \tag{3}$$

Here, k_B is Boltzmann's constant, T the absolute temperature, Γ_0 the attempt frequency, and E_m the migration energy, generally evaluated in static atomic-level calculations. The accuracy of the latter parameter is thus crucial to ensure the physical reliability of the model, as it embodies both thermodynamic and kinetic properties of the system being studied.

Migration rates associated with single point defects are traditionally computed with several approaches; see, e.g., Chatterjee and Vlachos (2007) and Becquart and Domain (2010) for extensive reviews. Here, we use the nudged elastic band (NEB) method (Henkelman and Jónsson 2000), which provides the migration energy given both initial and final states of the desired transition. The direct on-the-fly use of any cohesive models (e.g., IAP or DFT) to calculate energy barriers is impractical, because the exact saddle-point configurations in each transition event are unknown and must be sought with time-consuming procedures. A concrete solution based on ANN to overcome this technical limitation is depicted in Fig. 4. The AKMC module is found on the left-hand part of the figure, where an example of migration event for a single vacancy is indicated by the black arrow. The local atomic configuration (LAC) is defined by the species located in the closest neighbor lattice nodes to both initial and final vacancy sites, denoted as A to M and encircled by a blue line in the figure. This LAC is described by a vector of numerical signals, aimed at communicating with other modules of the simulation code. The migration energies are obtained in the NEB module, using a suitable cohesive model to evaluate total energy and atomic forces. This may be directly returned to the AKMC module (on-the-fly mode), but most importantly it is also stored in a database aimed at designing an ANN in replacement of the NEB module. The ANN is obviously expected to accurately return the migration energies corresponding to the same configurations as previously added in the database, but most importantly it is aimed at making faithful predictions for new (never previously calculated) configurations, i.e., for any atomic configuration that may be encountered during the AKMC simulation. This prediction is very fast produced by any computer, thereby gaining orders of magnitude in terms of computing time.

3.1   Assuming a Rigid-Lattice Description of the Studied System

In rigid-lattice models, the atoms in the studied system are assumed to occupy always the nodes of a perfect lattice, e.g., bcc. This hypothesis is reasonable if single vacancies are the only defect present in the system, but it can also include more complex defects, such as vacancy-solute clusters (Castin et al. 2012), single SIA (Terentyev et al. 2012), or even small SIA clusters (Castin 2011). In these cases,
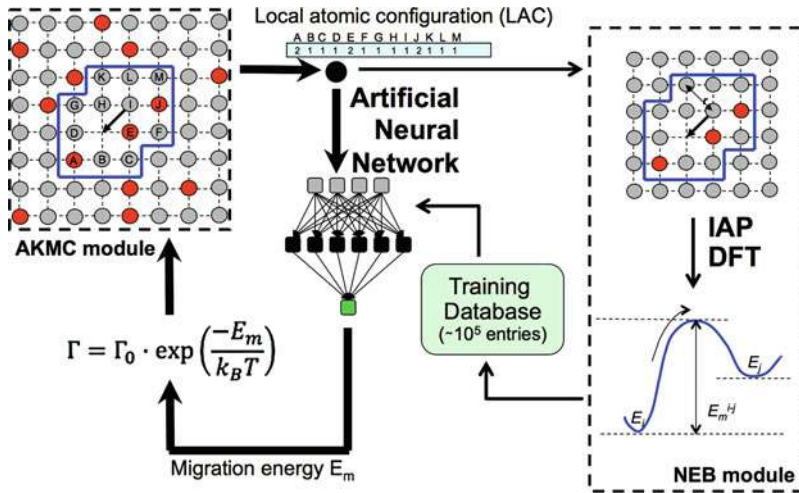
Fig. 4 AKMC model aided with ANN for fast and accurate on-the-fly estimations of the activation energies associated to the migration of single point defects
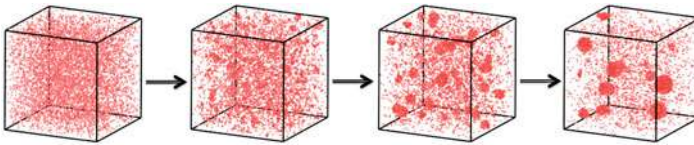
Fig. 5 Snapshots from an AKMC simulation of a thermal annealing experiment in a Fe-1.34% Cu at 500°C

the LAC vector in Fig. 4 may remain limited to a minimal amount of information: since the atomic coordinates are deducible from the crystallographic structure, only the chemical nature of each lattice site needs to be specified. The LAC is thus unequivocally defined with a vector of integers: each of them takes a predefined value that varies with the chemistry (e.g., value 1 stands for Fe, 2 for Cu, 3 for Ni, etc.) for the atoms sitting at sites A to M in the figure.

This methodology was applied to simulate thermal annealing experiments in Fe-based model alloys. Primarily using IAP as cohesive models, the FeCr system was first considered (Castin and Malerba 2010), and also the FeCu system (Castin et al. 2011), modeling the kinetics of precipitation. The physical accuracy of these models was later on significantly enhanced (Messina et al. 2017; Castin et al. 2017), using DFT directly as cohesive model. In a cubic simulation box with periodic boundary conditions, the prescribed content in solutes is initially introduced in random positions, as illustrated on the left-hand side in Fig. 5, together with a single vacancy. The simulation proceeds by computing its possible migration events toward each 1 nn position (eight in bcc structures). The individual migration energies are evaluated by the procedure illustrated in Fig. 4, which entirely pilots the evolution of the

system. Homogeneous precipitation of solutes eventually takes place, as depicted in Fig. 5, if so dictates the thermodynamics embedded in the underlying cohesive model, reflected by the ANN if the error of prediction is low enough.

In the absence of a formal feedback on the ANN predictions, their reliability for new atomic configurations must be maximized with a proper choice of the examples in both the training and the reference sets. Mathematically speaking, a new configuration is a new combination of integers in the LAC vector. Since each of these integers is seen at all positions many times in the training database (in other words, many cases are included with either an Fe or an X atom sitting in a given lattice site), new configurations are actually never in an extrapolative area in the input space. However, a new configuration can be considered as extrapolative from a physical point of view, if it describes a case governed by the kinds of interactions that were never, at least qualitatively, included in the training set. One can show (Castin and Malerba 2010) that the ANN predictive capabilities are maximized if the training data encompasses all expected stages of the thermal annealing process, as depicted in Fig. 5: Configurations where the solute atoms are randomly diluted in the matrix, or found under the form of small and big clusters, should be equally represented in the ANN training database.

The accuracy of prediction performed by the ANN trained to predict the vacancy migration energies, and the results of the AKMC simulations, is summarized in Fig. 6. In the FeCr case, the DFT-based model leads to predictions in better agreement with experimental evidence than the one based on IAP. Both densities and average sizes of the formed solute clusters coincide. In the FeCu case, the mechanism for the coherent stages of Cu precipitation in Fe was revealed (Castin et al. 2011), further stressing the importance of Cu cluster mobility (Soisson and Fu 2007): classical Ostwald ripening is not sufficient to explain the rapid kinetics of Cu precipitation. This was rendered possible, thanks to the model hybridization, and the accurate parametrization enhanced by the ANN. Even if the predictions obtained from an IAP resulted very satisfactory, the DFT-based parametrization maximized the agreement with experimental evidence. The solubility limit as predicted by DFT did not match the experimental one (Messina et al. 2017), with a consequent overestimation of the cluster density. Nevertheless, the model significantly improved from the point of view of time rescaling, necessary to convert the time in the MC simulation into a physical time comparable to the experiment, which resulted more consistent.

3.2 Applications Beyond the Rigid-Lattice Limit

The ANN-based AKMC algorithm as described in the previous section is not directly applicable to systems where the rigid-lattice assumption is no longer valid, e.g., near free surfaces, grain boundaries, in the presence of dislocations, or in the presence of nanostructural features such as dislocation loops. By removing the rigid-lattice assumption, the definition of the transitions and especially of the LAC becomes much more complex. Therefore, a different formulation is needed.
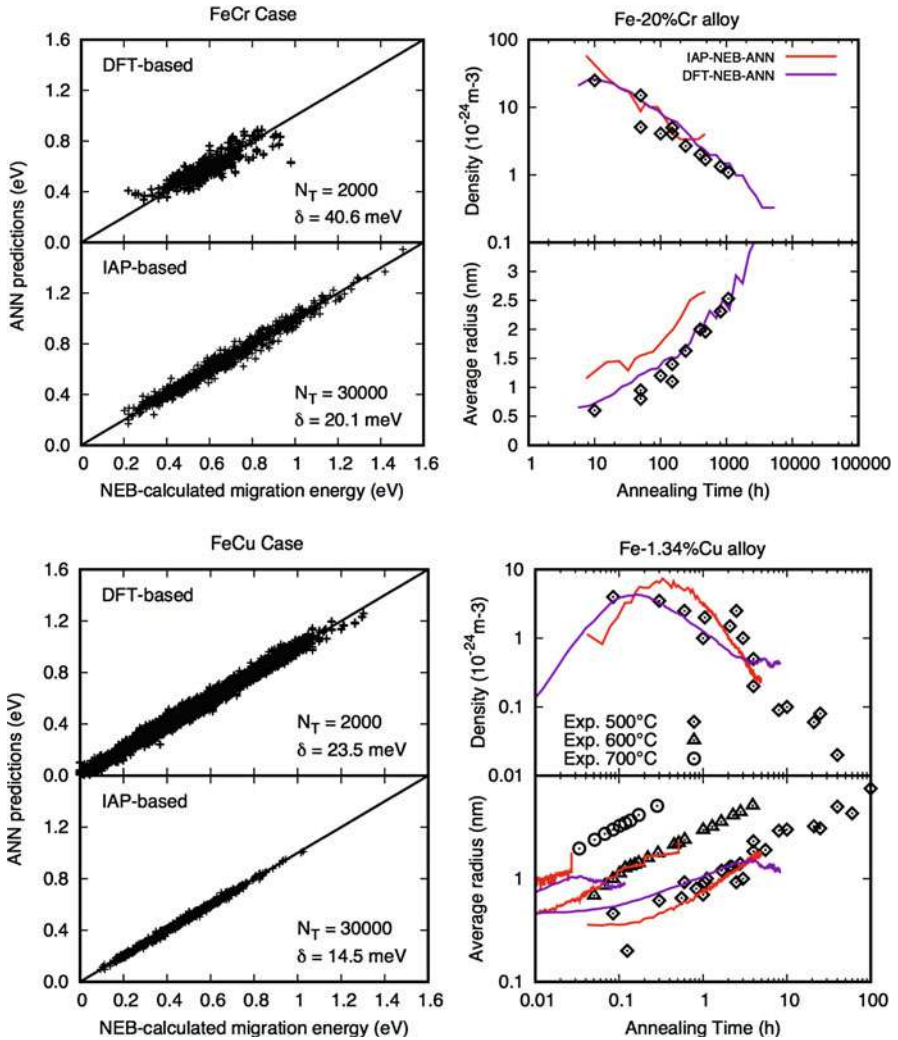
Fig. 6 Application of the ANN-based AKMC model to thermal annealing experiments. The ANN accuracy of prediction for the vacancy migration energies is shown on the left panels. N_T is the number of examples in the training database, and δ is the residual average error of prediction, after training. On the right panels, the series "IAP-NEB-ANN" refers to the ANN trained from NEB calculations using an IAP, whereas "DFT-NEB-ANN" is based on DFT, (top) in Fe-20%Cr at 500°C. Experimental data is taken from Novy et al. (2009), Novy (2009), and Bley (1992), (bottom) in Fe-1.34%Cu at 500, 600, and 700 °C. Experimental data is taken from Goodman et al. (1973), Kampmann and Wagner (1986), Mathon et al. (1997), and Perez et al. (2005)

Fully lattice-free AKMC models do exist. They constantly explore the local curvature of the potential energy surface and find transitions to nearby basins looking for saddle points, as specified by the transition state theory (Fichthorn and Weinberg 1991). For example, in 2001, Henkelman and Jonsson used the

dimer method (Henkelman and Jónsson 1999), which in theory searches for all possible transitions without the need to make assumptions. Other authors developed different schemes, though accomplishing the same finality, e.g., the ART method (Barkema and Mousseau 2001; Béland et al. 2011). In these, the definition of the migration events, and the calculation of their corresponding migration barriers, is made on the fly. Clearly, the advantage is the flexibility with respect to the simulated system. However, the required computing time is prohibitive for performing long simulations, inherently limiting the practical application of the method to no more than a few thousands of events. This is undoubtedly insufficient to study long-term and slow processes such as precipitation or depletion of solutes at interfaces.

Similar ideas to the ANN-based model described in the previous sections may be developed in a lattice-free framework (Castin et al. 2014), for proposing a compromise between the afore-discussed extremes. The search for any possible transition events with a fully general method is equivalent to test migration vectors in the $3N$ space, starting from the present state of the system. This allows the definition of a large number of migration events, which can become unmanageable. Additionally, many events found are likely to be minor (e.g., slight displacement of a single atom to a very nearby stable position), thus not letting the system advance significantly in time. Instead, we define a generic procedure for defining transitions, assuming the most likely events. Let's consider the example of a grain boundary in FeCr alloys. Making legitimate assumptions, migration events may be defined looking for sites with enough open volume (thus playing the role of a pseudo-vacancy) (Castin et al. 2014), using a geometrical criterion based on Voronoi vertexes, as shown in Fig. 7a. In the figure, seven atoms are found within a prescribed maximal distance from an eligible site, thus leading to the definition of seven transition events (the migration of any of them toward the site). A vector of migration is thus naturally defined for each transition, as illustrated in Fig. 7b. The direction of evolution for the system is thus reduced to a vector in three dimensions (indicated as $\bar{d}^{(e)}$ in Fig. 7b), expressing the relative translation for the migrating atom.

The activation energy associated with each event may be defined using a semirigid-lattice procedure: the migrating atom is first rigidly translated to the destination site, and later on full static relaxation using conjugate gradients is applied. The obtained configuration is defined as the final state assigned to the event, and NEB is applied to evaluate the migration energy. Similar to the idea illustrated in Fig. 4, many examples of transition events are gathered in a database, until enough data are available for designing an ANN, aimed at fully replacing the NEB calculations. Similar to the rigid-lattice AKMC, its purpose is strictly to provide a numerical estimate of the migration energy. The required inputs are, again, a description of the initial state (calculated from the relaxed atomic coordinates in the $3N$ space). The direction of migration is, also, implicitly provided, because the z-axis of the attached referential system of spherical coordinates is, by convention, aligned with the migration vector $\bar{d}^{(e)}$. It is worth noting that the final state of the system, i.e., after a given migration event is applied, is not known. Different from the rigid-lattice case, it cannot be fully deduced without the application of static relaxation.
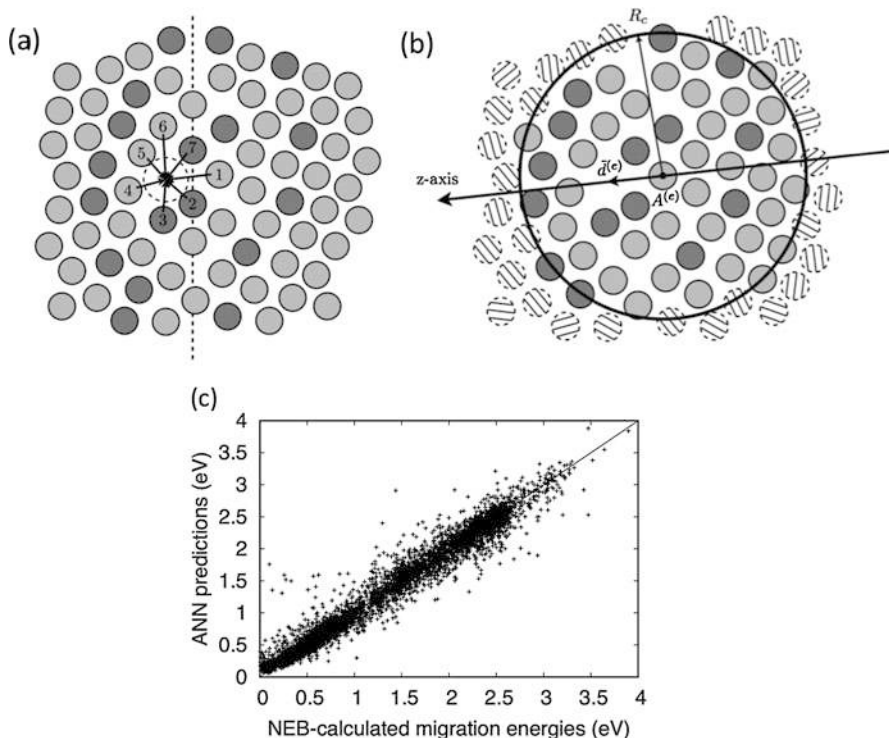
Fig. 7 Application of the ANN-based AKMC model to a lattice-free problem. (**a**) Definition of possible transition events with a generic procedure. (**b**) Definition of a migration vector (denoted as $\bar{d}^{(e)}$) and a LAC (atom within a given cutoff radius R_c) for a given event. (**c**) Accuracy of the prediction of the event migration energy, achieved by the ANN after training

Following a basic Monte Carlo algorithm, a list of events is established at every step of the simulation using the generic procedure, and the ANN is used to estimate the associated migration energies. One event is thus stochastically selected among the others. From this point, only static relaxation is necessary, after the migration atom is rigidly placed at its destination. The application of NEB is then optional and may be reserved for occasional feedback on the ANN predictions.

The major advantage of the here proposed lattice-free AKMC model is that the CPU cost is reduced to the minimum necessary. Instead of applying a general transition search method many times at each step of the simulation (for which it is in principle necessary to define a list of events), this operation is performed only once, i.e., for the selected event only. Once again, this algorithm entirely relies on the capacity of an ANN to make faithful predictions of the event migration energies, given as input a description of the initial state, and of the migration vector $\bar{d}^{(e)}$. We demonstrate the feasibility, successfully training an ANN for the case of a single vacancy migrating near a grain boundary in FeCr alloys, as shown in Fig. 7c. Taking advantage of the symmetry of the system, the ANN input variables were taken to be

the moduli of the following C_{nlm} coefficients. These are, by construction, invariant with respect to rigid rotations around the migration vector $\bar{d}^{(e)}$ in Fig. 7.

$$C_{nlm} = \frac{1}{N_{\text{RC}}} \sum_{i=1}^{N_{\text{RC}}} S_i R_n(r_i) Y_{lm}(\theta_i, \varphi_i) \tag{4}$$

Here, n, m, and l are integers defining the complexity for an expansion in series, with $0 \leq m \leq l$. The summation in Eq. (4) is performed over all neighboring atoms i found within the prescribed cutoff R_C from the migrating atom, located in space using their relative $(r_i, \theta_i, \varphi_i)$ spherical coordinates. Functions $R_n(r_i)$ are series of orthogonal radial functions, and Y_{lm} is the Laplace spherical harmonics. Finally, S_i is an integer describing the chemical species of the atom i, e.g., 1 for Fe, 2 for Cr, etc. The vector of ANN input variables is defined as:

$$\bar{C} \triangleq [|C_{nlm}|] \tag{5}$$

with $1 \leq n \leq N_{\text{Max}}$ and $0 \leq l \leq L_{\text{Max}}$. The number N_C of ANN input variables is thus given by:

$$N_C = N_{\text{Max}} \left(L_{\text{Max}} + \frac{L_{\text{Max}}(L_{\text{Max}} + 1)}{2} \right) \tag{6}$$

The ANN shown in Fig. 7c was defined taking $N_{\text{Max}} = 8$ and $L_{\text{Max}} = 36$, thus with $N_C = 5616$. It was trained on the basis of a database collecting a total of 31000 NEB calculation. The average residual error of prediction is 123 meV.

4   Neural-Network Potentials Fitted from DFT

As previously evoked, cohesive models are a central element in a multiscale modeling strategy. Density functional theory (DFT) is, in this respect, often an ideal and thus preferred choice. Cluster expansion models, e.g., the FeCrW by Bonny et al. (2016), were originally proposed to extrapolate DFT data for rigid-lattice configurations (van de Walle and Ceder 2002). In this case, the total energy is decomposed in contributions from clusters of atoms, according to predefined patterns: these are, e.g., pairs of atoms found in first nearest-neighbor position, triangles, quadruplets, etc. While this approach has been especially taken for addressing simulations in bulk materials driven by vacancy defects (der Ven and Ceder 2005), it has also been used to consider configurations with self-interstitials (Ngayam-Happy et al. 2012). It is worth noting that the deployment of a CE in the latter case is far more delicate, mainly because of the increasing complexity in the geometry of the explored configurations and the ensuing reduction of exploitable symmetries. As a consequence, the required numbers of clusters in the expansion for a proper general description rapidly explodes.

Given these limitations, neural-network potentials (NNP) are a very promising alternative method. They naturally benefit from the generality and portability proper of ANN, therefore not making any assumption regarding the kind of interactions between the chemical species in the target alloy. Such qualities are predicated to surpass the abilities of EAM-like potentials or any CE, provided that enough data are available for training. As in the case of the CE, NNP are meant to learn from DFT directly, and no experimental data can be directly incorporated. Different from CE, however, NNP are not limited to a rigid-lattice formalism nor by the geometrical complexity of the configurations described.

A method for designing NNP specifically aimed for KMC models in binary Fe-based alloys was proposed (Castin et al. 2017). Consistent with the prerequisites for high-dimensional potentials, the total energy of a given atomic configuration is decomposed as:

$$E_{\text{ANN}}^{(\text{Tot})} = \sum_{a=1}^{N} E_{\text{ANN}}^{(X(a))}(\rho^{(a)}) \tag{7}$$

Here, ρ is the local atomic density. Superscript (a) refers to a particular atom within the N constituting the studied configuration, and $X(a)$ denotes the chemical species for atom (a), i.e., $X(a)$ = Fe, Cu, Cr, …. Functions $E_{\text{ANN}}^{(X(a))}$ are atomic energy functions (AEF), providing an estimation of the energy assigned to every atom of the corresponding chemical species; subscript ANN refers to the fact that each AEF is implemented by an individual ANN. Their input variables are a description of the local atomic density, using symmetry functions defined as follows:

$$Q_{nl}^{(a)} \triangleq \sum_{m=-l}^{l} |C_{nlm}^{(a)}|^2 \tag{8}$$

with $C_{nlm}^{(a)}$ defined above in Eq. (4) but the S_i coefficients for determining the chemical species are not included. Instead, this information is included by expanding the vector of ANN input variables in distinct components. Considering the example of a binary FeX alloy, three distinct $\bar{Q}$ vectors are combined, each including the contribution from chemical sub-ensembles in the neighboring atoms, i.e.:

$$\bar{Q}^{(a)} \triangleq \left[Q_{\text{FeFe}}^{(a)} Q_{\text{FeX}}^{(a)} Q_{\text{XX}}^{(a)} \right] \tag{9}$$

Here, $Q_{\text{Fe}}^{(a)}$, $Q_{\text{X}}^{(a)}$, and $Q_{\text{FeX}}^{(a)}$ are identical as defined in Eq. (5) but only consider either Fe atoms, X atoms, or both, respectively. Given N_{Max} and L_{Max}, the number N_Q of symmetry functions (corresponding to the ANN input variables) is thus given by:

$$N_{\text{Q}} = 3N_{\text{Max}}(L_{\text{Max}} + 1) \tag{10}$$

NNP were fitted for both the bcc FeCu and the FeCr systems (Castin et al. 2017) and also for the FeNi case. Primarily aiming to incorporate DFT-based energies

in rigid-lattice Monte Carlo models, training configurations were extracted from NEB calculations of single point defects migration energies (both single vacancy and single self-interstitial). To fix the ideas, each NNP was fitted from 2000 to 5000 NEB calculation (requiring 10–30 million CPU hours), providing a total number of 20,000–50,000 atomic configurations. The vector of input variables was defined using $N_{Max} = 5$ and $L_{Max} = 10$, thus leading to $N_Q = 165$. The accuracy of prediction, after training, is summarized in Fig. 8. Similar results were later on obtained for the FeNi system, as also shown in the figure.

In addition to refine the models for thermal annealing experiments (previously shown in Fig. 6), the potentials were used to evaluate the phase diagram, using a Metropolis Monte Carlo method. As shown in Fig. 9, we see that DFT predicts a Cr solubility limit that is very consistent with an extensive review of experimental data (Bonny et al. 2008, 2010); alloying Cr atoms should remain fully solute in the ferritic matrix at all temperatures, up to a concentration near 9at%Cr from which α' phases form.
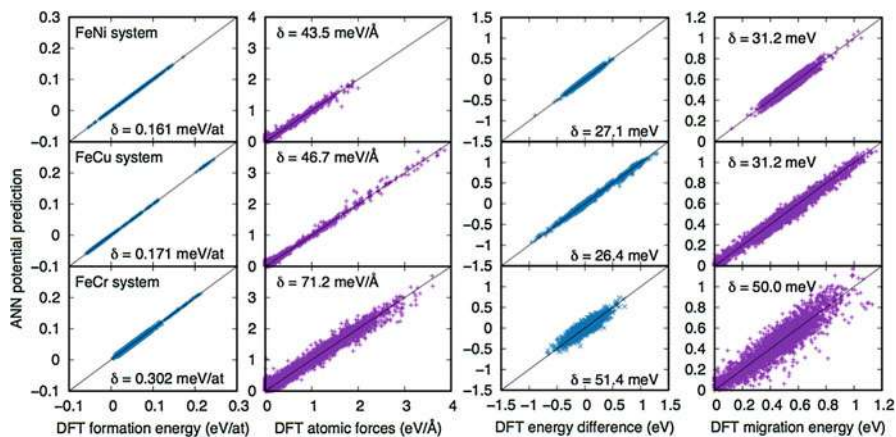
Fig. 8 Accuracy of prediction achieved by the ANN potential for the FeNi, FeCu, and FeCr systems. δ is the mean absolute error of prediction

Fig. 9 Phase diagram for the FeCr binary alloy, as predicted with Metropolis MC based on rigid-lattice NNP. The dashed line is a review from experimental data (Bonny et al. 2008, 2010)
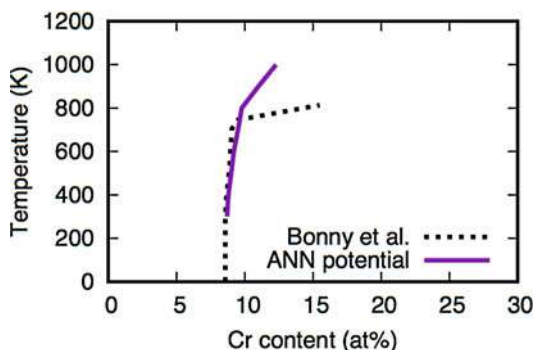
5   Future Perspectives: Proper Modeling of Radiation-Induced Hardening in Ferritic Steels

Artificial neural networks and machine learning schemes in general may find manifold applications in a multiscale modeling framework. The examples provided here are essentially all based on the idea of predicting how chemically complex local atomic configurations and/or specific strain fields influence the activation energy of selected thermally activated events, shortcutting computationally heavy calculation methods. Appropriately trained ANN are then used to inform more or less standard atomistic KMC models. Similar schemes can be used to calculate other quantities than activation energies and to transfer atomistic details to non-atomistic models, thereby helping in the effort of bridging between scales. Here a few examples are discussed.

As previously mentioned in Sect. 1, OKMC models were proposed for describing the microstructural evolution in RPV and F/M steels, assuming a gray matrix approximation as far as solute atoms are concerned. A future development of these models would consist in making the parameters of migration and dissociation of point-defect clusters sensitive to the local composition. As a matter of fact, phenomena of radiation-enhanced or radiation-induced heterogeneous precipitation are expected to create significant fluctuations in the local composition. Thus, vacancies will form complexes with solute atoms, mobile SIA clusters and dislocation loops will be repelled or attracted by regions rich in a given solute, while immobilized defects of the same type will become decorated by solutes, especially when these are dragged by point defects. In order to describe these processes, it becomes necessary to know, for example, how the formation, trapping, and migration energy of point-defect clusters change as a function of the local composition. Such a type of information cannot be calculated on the fly. First of all, such calculation would offset the advantage of non-atomistic modeling tools, i.e., the fact of simulating relatively long timescale. Secondly, concentration-dependent quantities are inherently averages or else randomly selected values for different configurations with the same local solute concentrations. ANN trained to calculate these quantities as functions of the local atomic configurations would allow the introduction of the chemical and atomistic detail in models that in fact do not include atoms but handle only, in the best case, local concentrations.

Therefore, a more realistic description of solute dragging by single point defects could be provided, beyond the assumption of infinite dilution limit, incorporating the effect of local chemical configurations. For instance, the dragging of a solute by a point defect may be reduced by the presence of other solutes, or the carrying defect may more favorably bind to a different solute encountered along the way. Given an adequate cohesive model as input, the ANN-based technique described in Sect. 3 could be integrated in the OKMC model to provide on-the-fly appropriate parameters at negligible computational cost.

Another example concerns the behavior of SIA loops decorated by solutes. By DFT some data on the relevant binding energy can be obtained in the presence of a

limited number of solutes at a time, and it is impossible to explore all possible ways in which solutes will aggregate around a loop. A valuable improvement in the model would be to train ANN to accurately predict the binding energy between SIA loops and decorating solutes, ideally on DFT data. Given as input a loop size in number of SIA and a content in solutes (either as an explicit set of spatial coordinates or as a composition in regions of space), ANN could assimilate how the effective binding energy changes.

Similarly, and without mutual exclusion, ANN could be used to predict how the migration energy of mobile defects is influenced by the presence, in non-atomistic models such as those of OKMC type, of strain fields, such as those created by dislocation loops and lines, grain boundaries, etc. It is of course possible to describe this effect by associating with each extended defect the strain field that it generates, as calculated in an elasticity theory framework, treating, for example, single point defect as elastic dipoles and parametrizing their migration energy on DFT as function of the dipole/strain field interaction. However, this is an approximation that slows down significantly the calculation and is limited to relatively simple strain fields, the description of which becomes complicated when, for example, the materials are anisotropic, or when strain-field superposition occurs. In contrast, ANN could be trained in this sense directly on atomistic models, to provide, say, the migration energy of a single point defect as function of the local strain field. Moving to different scales and processes, when applying dislocation dynamic models to describe plastic deformation in irradiated materials, it becomes necessary to define local rules of interactions between specific types of dislocations and obstacles. These obstacles can be solute clusters or precipitates of different types and compositions, loops orientated in different ways, decorated or not by solutes, etc. By molecular dynamics, it is possible to explore only a limited number of configurations, temperature being also a key variable to define the result of the dislocation/obstacle reaction. Once again, ANN trained to predict the key parameters that govern the interaction between dislocations and defect, depending on the specific features of both, would enable dislocation dynamic models to become more local and take into account variables, such as chemical composition of obstacles, which are currently very difficult or impossible to include in these models.

All the examples given concern quite strictly machine learning applied to varying atomistic configurations. In a more general and futuristic way, machine learning may one day shortcut and replace numerical simulation models that require long computing times to provide an answer. These models, e.g., suite of codes that, through different scales, provide the increase of yield strength as a consequence of irradiation for a given temperature, flux, fluence, materials composition, etc., may be used to provide examples of results for given initial and operating conditions, on which suitable machine learning schemes could be trained, possibly also completed by experimental examples. These machine learning schemes would then eventually be able to provide equivalent results, at negligible computational cost.

6   Conclusive Remarks

In this chapter, we have surveyed our work aimed at achieving a more physically accurate parametrization of atomic-scale modeling, more specifically kinetic Monte Carlo models devoted to the description of irradiation-induced microstructure changes in metals and alloys. However, the value of the concepts and examples here presented goes beyond the field of materials modeling, as there exist many more research fields where ANN, and machine learning techniques in general, can provide substantial contributions and help building more advanced and accurate models. It is in fact common in many modeling activities to face the need for more powerful and flexible regression techniques, which may appreciatively enhance the quality and applicability of the model.

References

Barkema G, Mousseau N (2001) The activation-relaxation technique: an efficient algorithm for sampling energy landscapes. Comput Mater Sci 20:285

Becquart CS, Domain C (2010) Introducing chemistry in atomistic kinetic Monte Carlo simulations of Fe alloys under irradiation. Phys Status Solidi B 247:9

Becquart CS, Domain C (2011) Modeling Microstructure and Irradiation Effects. Met Mat Trans A 42:852

Becquart CS, Barbu A, Bocquet J, Caturla M, Domain C, Fu CC, Golubov S, Hou M, Malerba L, Ortiz C, Souidi A, Stoller R (2010) Modeling the long-term evolution of the primary damage in ferritic alloys using coarse-grained methods. J Nucl Mater 406:39

Béland LK, Brommer P, El-Mellouhi F, Joly JF, Mousseau N (2011) Kinetic activation-relaxation technique. Phys Rev E 84:046704

Bergner F, Pareige C, Hernández-Mayoral M, Malerba L, Heintze C (2014) Application of a three-feature dispersed-barrier hardening model to neutron-irradiated Fe–Cr model alloys. J Nucl Mater 448:96

Bishop CM (1994) Neural networks and their applications. Rev Sci Instrum 65:1803

Bishop CM (1995) Neural networks for pattern recognition. Clarendon, Oxford

Bley F (1992) Neutron small-angle scattering study of unmixing in Fe-Cr alloys. Acta Metall Mater 40:1505

Bonny G, Terentyev D, Malerba L (2008) On the $\alpha - \alpha'$ miscibility gap of Fe–Cr alloys. Scr Mater 59:1193

Bonny G, Terentyev D, Malerba L (2010) New Contribution to the Thermodynamics of Fe-Cr Alloys as Base for Ferritic Steels. J Phase Equilibria Diffus 31:439

Bonny G, Castin N, Domain C, Terentyev D (2016) Density functional theory-based cluster expansion to simulate thermal annealing in FeCrW alloys. Philos Mag 97:1

Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. J Comput Phys 17:10

Castin N (2011) Atomistic kinetic Monte Carlo simulation of precipitation and segregation in metals for nuclear applications, using a novel methodology based on artificial neural networks. Ph.D thesis, Université Libre de Bruxelles (ULB), Belgium

Castin N, Malerba L (2010) Calculation of proper energy barriers for atomistic kinetic Monte Carlo simulations on rigid lattice with chemical and strain field long-range effects using artificial neural networks. J Chem Phys 132:074507

Castin N, Pascuet MI, Malerba L (2011) Modeling the first stages of Cu precipitation in α-Fe using a hybrid atomistic kinetic Monte Carlo approach. J Chem Phys 135:064502

Castin N, Pascuet M, Malerba L (2012) Mobility and stability of large vacancy and vacancy-copper clusters in iron: An atomistic kinetic Monte Carlo study. J Nucl Mater 429:315

Castin N, Fernandez J, Pasianot R (2014) Predicting vacancy migration energies in lattice-free environments using artificial neural networks. Comput Mater Sci 84:217

Castin N, Messina L, Domain C, Pasianot R, Olsson P (2017) Improved atomistic Monte Carlo models based on ab-initio-trained neural networks: Application to FeCu and FeCr alloys. Phys Rev B 95:214117

Chatterjee A, Vlachos D (2007) An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. J Comput-Aided Mater Des 14:253

Chiapetto M, Malerba L, Becquart C (2015a) Nanostructure evolution under irradiation in FeMnNi alloys: A "grey alloy" object kinetic Monte Carlo model. J Nucl Mater 462:91 and references therein

Chiapetto M, Malerba L, Becquart C (2015b) Effect of cr content on the nanostructural evolution of irradiated ferritic/martensitic alloys: an object kinetic monte carlo model. J Nucl Mater 465:326

der Ven AV, Ceder G (2005) Vacancies in ordered and disordered binary alloys treated with the cluster expansion. Phys Rev B 71:054102

Fichthorn KA, Weinberg WH (1991) Theoretical foundations of dynamical Monte Carlo simulations. J Chem Phys 95:1090

Goodman SR, Brenner SS, Low JR (1973) An FIM-atom probe study of the precipitation of copper from lron-1.4 at. pct copper. Part I: Field-ion microscopy. Metall Trans 4:2363

Henkelman G, Jónsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111:7010

Henkelman G, Jónsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. J Chem Phys 113:9978

Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115:9657

Hornik K, Stinchcombe M, White H (1989) Multilayer Feedforward Networks Are Universal Approximators. Neural Netw 2:359

Jansson V, Malerba L (2013) Simulation of the nanostructure evolution under irradiation in Fe–C alloys. J Nucl Mater 443:274

Kampmann R, Wagner R (1986) Phase Transformations in Fe-Cu-Alloys-SANS-Experiments and Theory. In: Janot C, Petry W, Richter D, Springer T (eds) Atomic transport and defects in metals by neutron scatterings. Springer, New York, pp 73–77

Kuksenko V, Pareige C, Pareige P (2013) Cr precipitation in neutron irradiated industrial purity Fe–Cr model alloys. J Nucl Mater 432:160

Levenberg K (1994) A method for the solution of certain non-linear problems in least squares. Q Appl Math 2:164

Marquardt D (1963) An Algorithm for Least-Squares Estimation of Nonlinear Parameters. SIAM J Appl Math 11:431

Mathon MH, Barbu A, Dunstetter F, Maury F, Lorenzelli N, de Novion CH (1997) Experimental study and modelling of copper precipitation under electron irradiation in dilute FeCu binary alloys. J Nucl Mater 245:224

Meslin E, Radiguet B, Pareige P, Barbu A (2010) Kinetic of solute clustering in neutron irradiated ferritic model alloys and a French pressure vessel steel investigated by atom probe tomography. J Nucl Mater 399:137

Messina L, Castin N, Domain C, Olsson P (2017) Improved atomistic Monte Carlo models based on ab-initio-trained neural networks: Application to FeCu and FeCr alloys. Phys Rev B 95:214117

Ngayam-Happy R, Becquart C, Domain C, Malerba L (2012) Formation and evolution of MnNi clusters in neutron irradiated dilute Fe alloys modelled by a first principle-based AKMC method. J Nucl Mater 426:198

Novy S (2009) Mécanismes De Vieillissement à Très Longue Échéance Des Aciers Inoxydables Austéno-ferritiques. Ph.D thesis, University of Rouen, France

Novy S, Pareige P, Pareige C (2009) Atomic scale analysis and phase separation understanding in a thermally aged Fe–20at.%Cr alloy. J Nucl Mater 384:96

Olsson P, Klaver T, Domain C (2010) Ab initio study of solute transition-metal interactions with point defects in bcc Fe. Phys Rev B 81:054102

Pasianot R, Malerba L (2007) Interatomic potentials consistent with thermodynamics: The Fe–Cu system. J Nucl Mater 360:118

Perez M, Perrard F, Massardier V, Kebler X, Deschamps A, de Monestrol H, Pareige P, Covarel G (2005) Low-temperature solubility of copper in iron: experimental study using thermoelectric power, small angle X-ray scattering and tomographic atom probe. Philos Mag 85:2197

Petrequin P (1995) A review of formulas for predicting irradiation embrittlement of reactors vessel materials. In: International working group on life management of nuclear power plants. International Atomic Energy Agency, Vienna, pp 95–147

Soisson F, Fu CC (2007) Cu-precipitation kinetics in α-Fe from atomistic simulations: vacancy-trapping effects and Cu-cluster mobility. Phys Rev B 76:214102

Terentyev D, Castin N, Ortiz C (2012) Correlated recombination and annealing of point defects in dilute and concentrated Fe–Cr alloys. J Phys: Condens Matter 24:475404

van de Walle A, Ceder G (2002) Automating First-Principles Phase Diagram Calculations. J Phase Equilib 23:348

Young WM, Elcock EW (1966) Monte Carlo studies of vacancy migration in binary ordered alloys: I. Proc Phys Soc Lond 89:735

Part XV
Multiscale Modeling of Diseases

Multiscale Modeling of Diseases: Overview

104

George E. Karniadakis and Ming Dao

Contents

Abstract

Some of the most recent advances in multiscale modeling of human diseases are briefly summarized. These progresses benefited from the rapid development of advanced multiscale modeling techniques as well as the development and availability of new mechanobiology experimental techniques and diverse imaging modalities. The close interdisciplinary collaborations between modeling scientists, experimentalists, and medical doctors enabled unprecedented quantitative understanding of human pathologies. Two specific examples are given on the recent advances in multiscale modeling of malaria – an infectious disease and sickle cell disease – a genetic disease. Future outlook is presented, followed by a brief introduction of the chapters included in this section.

G. E. Karniadakis (✉)
Division of Applied Mathematics, Brown University, Providence, RI, USA
e-mail: george_karniadakis@brown.edu

M. Dao
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: mingdao@mit.edu

1   Overview of Recent Advances

All biological systems in a human, e.g., from a big organ such as the heart to a simple cell such as the red blood cell (RBC), are multiscale and multi-rate in nature with characteristic spatiotemporal scales disparate by several orders of magnitude. Diseases often affect the cellular structure of diverse cells, the tissues, the signaling mechanisms, but also one or multiple organs. For example, parasitic infectious diseases or other hereditary hematologic disorders, such as sickle cell disease, are often associated with major morphological changes and affect the viscoelastic properties of RBCs, that is, they are becoming stiffer. Protein or hemoglobin changes in RBCs, in turn, can disrupt blood flow and possibly lead to vaso-occlusion affecting significantly brain perfusion, as in the case of cerebral malaria. Modeling of these hematologic disorders requires a seamless multiscale approach, where blood cells and their components as well as blood flow in arteries, arterioles, and capillaries are all represented accurately using physiologically consistent parameters. However, not all these parameters can be obtained from first-principles nor from in vivo experiments, but with careful ex vivo or in vitro experiments they can be estimated with adequate accuracy in many cases.

In the past decade or so, there have been many exciting developments in the field of blood rheology, cell mechanics, and tissue mechanics for human diseases, and multiscale modeling has played a critical role in the process. These advances were made possible due to the following three key factors, among others:

1. The development of advanced multiscale modeling techniques (Lin et al. 2013, 2017a) that interface heterogeneous descriptions and corresponding computational solvers across scales. Also, the development of coarse grained versions of molecular dynamics that can simulate much bigger domains and for longer time scales. Finally, large-scale simulations are now being performed almost routinely on most of the labs on inexpensive CPU and GPU clusters providing unprecedented computational power

2. The development of new mechanobiology experimental techniques (Bao and Suresh 2003; Suresh 2007a; Suresh et al. 2005; Diez-Silva et al. 2010) and diverse imaging modalities (Choi et al. 2007; Park et al. 2008, 2010; Hosseini et al. 2016). These tools were made widely available within the past 10–15 years, while not available or hardly accessible to the previous generation of researchers, including optical tweezers (OT) (Dao et al. 2003; Diez-Silva et al. 2010), atomic force microscope (AFM) (Bao and Suresh 2003; Suresh 2007a, b; Carvalho et al. 2013), microfluidics (Bow et al. 2011; Quinn et al. 2011; Du et al. 2013, 2015), diffraction phase microscope (DPM) (Park et al. 2008, 2010), tomography phase microscope (TPM) (Choi et al. 2007) etc. The recent advances have provided diverse and accurate multiscale measurements for static and dynamic experiments that help to obtain input parameters and validate the simulation outputs of the multiscale models. Figure 1 shows a microfluidic platform developed recently at MIT for investigating biophysical/biomechanical
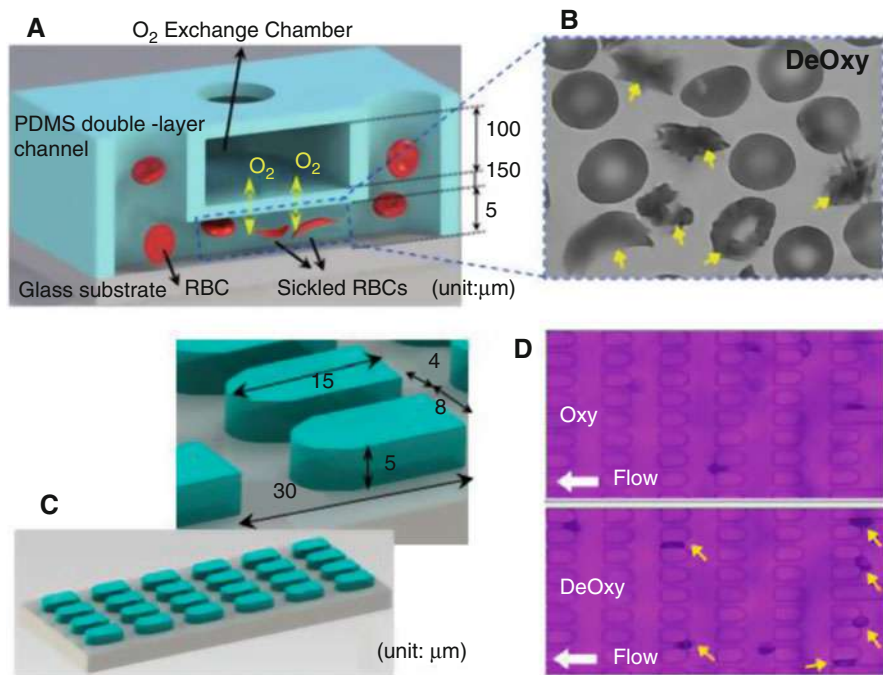
Fig. 1 A microfluidic platform developed recently for investigating biophysical/biomechanical alterations in sickle RBCs under transient hypoxia conditions (Du et al. 2015). (**a**) shows an oxygen-controlled chamber while (**b**) shows the result of hypoxic conditions and the heterogeneity of shapes of sickle RBCs. (**c**) shows a flow cytometer with obstacles designed so that deformable RBCs can easily flow through while stiffer sickled RBCs can get stuck. (**d**) shows some typical results of flowing RBCs under oxygenated (20% O_2) and deoxygenated (2% O_2) conditions, where the yellow arrows indicate sickled RBCs being trapped by the microgates. More details can be found in Du et al. (2015).

alterations in sickle RBCs under transient hypoxia conditions (Du et al. 2015), enabling patient-specific multiscale models as demonstrated in Li et al. (2016, 2017b).

3. The close interdisciplinary collaborations between modeling scientists, experimentalists, and medical doctors in pursuing deep, quantitative understanding of human pathologies (Peng 2011). These interdisciplinary research efforts have been strongly promoted by specific funding initiatives such as the Interagency Modeling and Analysis Group (IMAG) Multiscale Modeling (MSM) Consortium at National Institutes of Health (NIH) in USA, the European Research Council (ERC) supported projects such as "MATHCARD" and "iHEART," as well as international collaborative programs such as Singapore-MIT Alliance for Research and Technology (SMART).

Here we would like to discuss two specific examples where we employed multiscale modeling with microfluidic platforms, advanced phase microscopy, AFM, and optical tweezers synergistically. The first example is about the recent advances in multiscale modeling of an infectious disease – malaria, and the second example is about the latest progresses in a genetic disease – sickle cell disease. More details and other diseases/disorders will be covered in the following chapters included in this section, including modeling of blood clots and lung pathologies.

Malaria: Malaria is still a major infectious disease threatening global public health, with high mortality in young children (Desai et al. 2007; Greenwood et al. 2008; O'Meara et al. 2010; Wells et al. 2015). The disease is transmitted between mosquito hosts and human hosts by *plasmodium* parasites, which infect RBCs. There were a number of key advances recently on the mechanical property changes versus disease states, including the confirmation of RESA protein being the key molecule almost single-handedly responsible for the stiffening of the infected RBCs within the first 24 h (Mills et al. 2007), the precise measurement of up to 10 times progressive stiffening of the infected RBCs within 48 h using multiple independent experimental techniques (OT, microfluidics, DPM) (Suresh et al. 2005; Park et al. 2008; Bow et al. 2011; Mills et al. 2007; Diez-Silva et al. 2012), the quantitative measurement of cytoskeleton damages through the disease states (Shi et al. 2013; Sinha et al. 2015), and protein specific adhesion characteristics (Carvalho et al. 2013; Xu et al. 2013). In parallel, accompanying these new understandings and new experimental data, the multiscale modeling of the healthy and malaria infected RBCs made significant advances as well (Dao et al. 2003, 2006; Quinn et al. 2011; Bow et al. 2011; Li et al. 2005, 2007; Fedosov et al. 2011a, b; Peng et al. 2013; Li et al. 2014; Fedosov et al. 2014; Zhang et al. 2015; Pivkin et al. 2016; Tang et al. 2017), including the RBC membrane model capable of incorporating spectrin tetramer and junction complex connections with lipid membrane (Peng et al. 2013; Tang et al. 2017), the detailed model that can incorporate cytoskeleton and/or membrane defects (Li et al. 2005, 2014; Peng et al. 2013; Zhang et al. 2015; Tang et al. 2017), and the latest OpenRBC whole cell model that includes currently all the cytoskeleton proteins and in future versions the metabolic pathways (Tang et al. 2017). These models were able to accurately reproduce experimental results using OT, microfluidics, DPM, etc., and provided quantitative tools for exploring many mechanistic insights not directly extractable through experiments only (Dao et al. 2003, 2006; Quinn et al. 2011; Bow et al. 2011; Li et al. 2005, 2007, 2014; Fedosov et al. 2011a, b, 2014; Peng et al. 2013; Zhang et al. 2015; Pivkin et al. 2016; Tang et al. 2017).

Sickle Cell Disease: Sickle cell disease is a genetic disease caused by a single amino acid defect in producing hemoglobin (Hb). It was named by Linus Pauling the first *molecular disease* in an early paper (Pauling et al. 1949). The defective hemoglobin, or sickle hemoglobin (HbS), will polymerize and form stiff HbS bundles under low oxygen conditions, resulting in sickle-shaped and stiff RBCs as well as enhanced adhesion. These severely compromised RBC mechanical properties are responsible for the related clinical complications in sickle cell disease, including vaso-occlusive painful crisis, acute chest syndrome, and various

organ/tissue damages in spleen, lung, bone, etc. (Piel et al. 2017; Ware et al. 2017). Some recent new progresses and fundamental understanding in sickle cell disease include the establishment of single-cell based microfluidics platform for dynamic sickling kinetics and capillary obstruction measurements under transient-hypoxia (Du et al. 2015; Li et al. 2017a, b), the detailed characterization of the transient-hypoxia sickling characteristics for different sickle RBC density populations (Du et al. 2015; Li et al. 2016, 2017a), the AFM characterization of sickle RBCs (Maciaszek and Lykotrafitis 2011; Li et al. 2017a), the normoxia shape and mechanical property characterization at single-cell level for different sickle RBC density populations (Xu et al. 2017). With these new experimental techniques and newly available experimental data, the multiscale modeling of sickle RBCs made many significant advances as well (Li et al. 2017a; Lu et al. 2017), including the detailed polymerization model capable of predicting the homogenous and heterogeneous nucleation and growth of the polymerized HbS fibers and fiber bundles (Lu et al. 2016, 2017), the whole cell model capable of incorporating HbS polymerization and resulted shape changes (Lu et al. 2017; Lei and Karniadakis 2012; Li et al. 2012, 2013), and the vaso-occlusion model that incorporate both cell stiffening and cell adhesion (Lei and Karniadakis 2013). These models were able to correctly simulate some critical in vivo vaso-occlusive observations in mice (Kaul and Fabry 2004) and in vitro experiments using the latest microfluidics platform (Du et al. 2015; Li et al. 2017b), and again helped providing many mechanistic insights not directly extractable through experiments only (Li et al. 2017a).

2   Outlook

Multiscale modeling techniques are not unique to biological systems, and advances across other disciplines will find their way to multiscale modeling of human diseases. Bottom-up approaches are still computationally expensive and more *aggressive* coarsening techniques will help to alleviate the cost without sacrificing greatly the accuracy. A particularly useful coarse-graining aspect is the dynamic refinement or de-refinement as shown in Fig. 2 that could target the fine-graining in regions that molecular details matter while representing the rest of the system with sufficient accuracy but at much coarser fidelity. Domain decomposition methods, like the one presented in the next chapter, will enable scaling up from the microscale and mesoscale to the macroscale, hence integrating molecular details with organ-level biomechanics. Precision medicine implies individualized diagnosis and predicting the evolution of a disease, e.g., a possible rupture of a brain aneurysm, but this, in turn, requires a lot of personal data from a patient gathered at routine medical exams and at the time of crises. Therefore, *data-driven multiscale modeling* will play a central role in future modeling of diseases and the great challenge is how to incorporate such data at different scales of multiscale models. Machine and deep learning will play a big role in this data-model interaction for classification, as in the work of Xu et al. (2017), but also for inference and prediction of the evolution of the diseases, even for short term, e.g., in trying to predict glucose
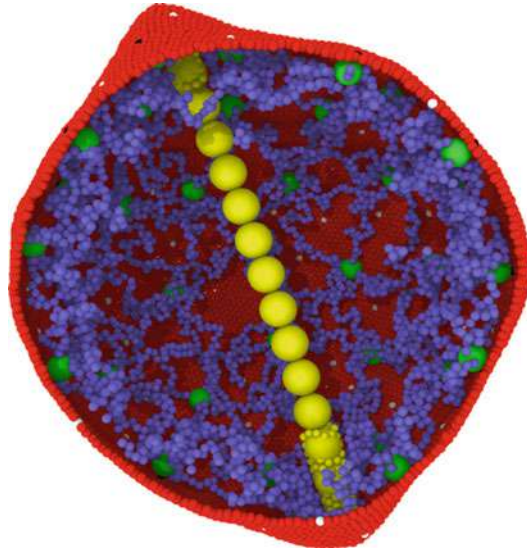
Fig. 2 Multiscale simulation of polymerization inside the RBC: The stiff HbS fiber interacts with the RBC membrane and distorts the RBC to the sickle shape. The HbS fiber is simulated by a hybrid HbS fiber model where two HbS fiber models at different length scales are coupled by applying a mesoscopic adaptive resolution scheme (MARS). In the HbS fiber, one small yellow particle represents one HbS molecule, and 56 HbS molecules are coarse-grained as one large yellow particle. In the RBC membrane, green particles represent actin junctions, blue particles represent spectrin proteins, black particles represent glycophorin proteins, white particles represent the band-3 complex, and red particles represent the lipid particles. (For more details see Lu et al. (2017); Courtesy of Lu Lu and He Li, Brown University)

levels in diabetic patients by monitoring glucose histories and train properly a deep neural network using personalized data, e.g., multi-omics approach and data. In fact, the proper integration of multi-omics with existing multiscale models at the molecular level may be one of the greatest challenges in future generations of multiscale modeling but such a significant advance will make precision medicine at last realizable. At the higher scale levels, several observed quantities from blood pressure, to oxygen concentration, to even monitoring the stiffness of human arteries are becoming possible through the use of wearable biosensors for generation of big data that could eventually help multiscale models to be truly predictive.

3 Brief Introduction of the Chapters

In ▶ Chap. 105, "Domain Decomposition Methods for Multiscale Modeling," Bian and Praprotnik introduce domain decomposition methods that facilitate a smooth transition from one modeling method to another one in adjacent domains. For example, they demonstrate how a continuum formulation, e.g., the Navier-Stokes

equations can be interfaced with the molecular dynamics method applied to a region where the continuum hypothesis breaks down. For more complex fluids that contain molecular structures, they couple molecular dynamics methods with coarse-grain molecular dynamics in the far field. The matching interface conditions can be applied directly to the state variables or indirectly by interpolating the forces across the domains.

In ▶ Chap. 106, "Particle-Based Methods for Mesoscopic Transport Processes," Li Z et al. introduce particle-based methods for mesoscopic transport processes. Biological processes at cellular and subcellular levels involve the transport of many reacting biochemical species, such as enzymes, proteins, and nutrients. The particle-based stochastic models for mesoscopic transport processes, including thermal transport, reactive chemical transport, and ionic transport in mesoscopic systems, are reviewed and discussed.

In ▶ Chap. 107, "Continuum- and Particle-Based Modeling of Human Red Blood Cells," Li XJ et al. address modeling of RBCs by both continuum and particle-based methods. The continuum method is more efficient and can simulate larger domains but the dissipative particle dynamics method used in the particle-based approach can represent molecular details down to the level of transmembrane proteins.

In ▶ Chap. 108, "Computational Models of Eukaryotic Cells in Health and Disease," Lykov K and Pivkin IV present computational models for eukaryotic cells suspended in flow or adherent to a substrate. These models represent a phenomenological top-down approach. The mechanical properties of the cell model components are obtained based on general principles and macroscopic properties of the cell. A future bottom-up approach will allow construction of eukaryotic cell models by relating the mechanical behavior of the whole cell with the bio-chemical properties of its molecular constituents.

In ▶ Chap. 109, "Multiscale Modeling of Malaria-Infected Red Blood Cells," Dasanna AK et al. review malaria disease modeling on the level of single infected RBC, thus quantifying its stiffer mechanics and cytoadherence – the two hallmarks of malaria disease. They discuss recent advances in multiscale models and show how such models are validated and subsequently used to put forward new hypotheses and even to be used for developing new treatment strategies.

In ▶ Chap. 110, "Multiscale Modeling of Sickle Cell Anemia," Li XJ and Lei H review multiscale modeling methods of sickle cell anemia with similar characteristics as malaria but with markedly larger variation of RBC shapes. Unlike its name, roughly 20% of deformed RBCs have the sickle shape whereas the rest can be elongated, granular, maple-leaf and several other shapes. This shape heterogeneity produced by the hemoglobin polymerization process inside sickle RBCs under hypoxic conditions, e.g., in post-capillary venules, affects the rheology of sickle blood and can also lead to different cell-cell and cell-wall interactions that can, in turn, cause total vaso-occlusion, which is responsible for most of the patient crises.

In ▶ Chap. 111, "Multiscale Modeling of Blood Flow-Mediated Platelet Thrombosis," Yazdani A et al. focus on blood flow mediated platelet thrombosis that is common in coronary or carotid arteries but also in prosthetic cardiovascular

devices. The authors describe a bottom-up particle approach starting with molecular dynamics for platelets as well as coarse-grained versions of it and scale up to RBCs and plasma using dissipative particle dynamics. The coagulation cascade that involves multiple advection-diffusion-reaction processes can also be modeled by a new version of dissipative particle dynamics that can deal with transport of species at the mesoscale.

▶ Chapter 112, "Cluster-Guided Multiscale Lung Modeling via Machine Learning," by Lin CL et al. focuses on multiscale modeling of airflow distribution and aerosol transport in the human lung. The scale separation from the trachea to alveoli spans over two orders of magnitude so this is a particularly challenging multiscale problem. The authors discuss methods for modeling lungs both for a specific patient as well as for a group of patients. The entire framework is presented consisting of an imaging-based lung model and a high-fidelity computational fluid dynamics model. For "between patients" modeling, the authors use machine learning to identify homogeneous sub-populations among healthy and diseased populations, in order to bridge individual and population scales.

References

Bao G, Suresh S (2003) Cell and molecular mechanics of biological materials. Nat Mater 2(11):715–725

Bow H et al (2011) A microfabricated deformability-based flow cytometer with application to malaria. Lab Chip 11(6):1065–1073

Carvalho PA, Diez-Silva M, Chen H, Dao M, Suresh S (2013) Cytoadherence of erythrocytes invaded by *Plasmodium falciparum*: quantitative contact-probing of a human malaria receptor. Acta Biomater 9(5):6349–6359

Choi W et al (2007) Tomographic phase microscopy. Nat Methods 4(9):717–719

Dao M, Lim CT, Suresh S (2003) Mechanics of the human red blood cell deformed by optical tweezers. J Mech Phys Solids 51(11–12):2259–2280

Dao M, Li J, Suresh S (2006) Molecularly based analysis of deformation of spectrin network and human erythrocyte. Mater Sci Eng C-Biomim Supramol Syst 26(8):1232–1244

Desai M et al (2007) Epidemiology and burden of malaria in pregnancy. Lancet Infect Dis 7(2): 93–104

Diez-Silva M, Dao M, Han JY, Lim CT, Suresh S (2010) Shape and biomechanical characteristics of human red blood cells in health and disease. MRS Bull 35(5):382–388

Diez-Silva M et al (2012) Pf155/RESA protein influences the dynamic microcirculatory behavior of ring-stage *Plasmodium falciparum* infected red blood cells. Sci Rep 2:614

Du E et al (2013) Electric impedance microflow cytometry for characterization of cell disease states. Lab Chip 13(19):3903–3909

Du E, Diez-Silva M, Kato GJ, Dao M, Suresh S (2015) Kinetics of sickle cell biorheology and implications for painful vasoocclusive crisis. Proc Natl Acad Sci U S A 112(5): 1422–1427

Fedosov DA, Caswell B, Suresh S, Karniadakis GE (2011a) Quantifying the biophysical characteristics of Plasmodium-falciparum-parasitized red blood cells in microcirculation. Proc Natl Acad Sci U S A 108(1):35–39

Fedosov DA, Lei H, Caswell B, Suresh S, Karniadakis GE (2011b) Multiscale modeling of red blood cell mechanics and blood flow in malaria. PLoS Comput Biol 7(12):e1002270

Fedosov DA, Dao M, Karniadakis GE, Suresh S (2014) Computational biorheology of human blood flow in health and disease. Ann Biomed Eng 42(2):368–387

Greenwood BM et al (2008) Malaria: progress, perils, and prospects for eradication. J Clin Investig 118(4):1266–1276

Hosseini P et al (2016) Cellular normoxic biophysical markers of hydroxyurea treatment in sickle cell disease. Proc Natl Acad Sci U S A 113(34):9527–9532

Kaul DK, Fabry ME (2004) In vivo studies of sickle red blood cells. Microcirculation 11(2): 153–165

Lei H, Karniadakis GE (2012) Predicting the morphology of sickle red blood cells using coarse-grained models of intracellular aligned hemoglobin polymers. Soft Matter 8(16):4507–4516

Lei H, Karniadakis GE (2013) Probing vasoocclusion phenomena in sickle cell anemia via mesoscopic simulations. Proc Natl Acad Sci U S A 110(28):11326–11330

Li J, Dao M, Lim CT, Suresh S (2005) Spectrin-level modeling of the cytoskeleton and optical tweezers stretching of the erythrocyte. Biophys J 88(5):3707–3719

Li J, Lykotrafitis G, Dao M, Suresh S (2007) Cytoskeletal dynamics of human erythrocyte. Proc Natl Acad Sci U S A 104(12):4937–4942

Li XJ, Caswell B, Karniadakis GE (2012) Effect of chain chirality on the self-assembly of sickle hemoglobin. Biophys J 103(6):1130–1140

Li XJ, Vlahovska PM, Karniadakis GE (2013) Continuum- and particle-based modeling of shapes and dynamics of red blood cells in health and disease. Soft Matter 9(1):28–37

Li XJ, Peng ZL, Lei H, Dao M, Karniadakis GE (2014) Probing red blood cell mechanics, rheology and dynamics with a two-component multi-scale model. Philos Trans R Soc A-Math Phys Eng Sci 372(2021):20130389

Li XJ et al (2016) Patient-specific blood rheology in sickle-cell anaemia. Interface Focus 6(1):20150065

Li XJ, Dao M, Lykotrafitis G, Karniadakis GE (2017a) Biomechanics and biorheology of red blood cells in sickle cell anemia. J Biomech 50:34–41

Li X, Du E, Dao M, Suresh S, Karniadakis GE (2017b) Patient-specific modeling of individual sickle cell behavior under transient hypoxia. PLoS Comput Biol 13(3):e1005426

Lin CL, Peng GCY, Karniadakis G (2013) Multi-scale modeling and simulation of biological systems preface. J Comput Phys 244:1–3

Lu L, Li XJ, Vekilov PG, Karniadakis GE (2016) Probing the twisted structure of sickle hemoglobin fibers via particle simulations. Biophys J 110(9):2085–2093

Lu L, Li H, Bian X, Li XJ, Karniadakis GE (2017) Mesoscopic adaptive resolution scheme toward understanding of interactions between sickle cell fibers. Biophys J 113(1):48–59

Maciaszek JL, Lykotrafitis G (2011) Sickle cell trait human erythrocytes are significantly stiffer than normal. J Biomech 44(4):657–661

Mills JP et al (2007) Effect of plasmodial RESA protein on deformability of human red blood cells harboring *Plasmodium falciparum*. Proc Natl Acad Sci U S A 104(22):9213–9217

O'Meara WP, Mangeni JN, Steketee R, Greenwood B (2010) Changes in the burden of malaria in sub-Saharan Africa. Lancet Infect Dis 10(8):545–555

Park YK et al (2008) Refractive index maps and membrane dynamics of human red blood cells parasitized by *Plasmodium falciparum*. Proc Natl Acad Sci U S A 105(37):13730–13735

Park YK et al (2010) Metabolic remodeling of the human red blood cell membrane. Proc Natl Acad Sci U S A 107(4):1289–1294

Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia, a molecular disease. Science 110(2865):543–548

Peng GCY (2011) What biomedical engineers can do to impact multiscale modeling (TBME letters special issue on multiscale modeling and analysis in computational biology and medicine: Part-2). IEEE Trans Biomed Eng 58(12):3440–3442

Peng ZL et al (2013) Lipid bilayer and cytoskeletal interactions in a red blood cell. Proc Natl Acad Sci U S A 110(33):13356–13361

Piel FB, Steinberg MH, Rees DC (2017) Sickle cell disease. N Engl J Med 376(16):1561–1573

Pivkin IV et al (2016) Biomechanics of red blood cells in human spleen and consequences for physiology and disease. Proc Natl Acad Sci U S A 113(28):7804–7809

Quinn DJ et al (2011) Combined simulation and experimental study of large deformation of red blood cells in microfluidic systems. Ann Biomed Eng 39(3):1041–1050

Shi H et al (2013) Life cycle-dependent cytoskeletal modifications in *Plasmodium falciparum* infected erythrocytes. PLoS One 8(4):e61170

Sinha A, Chu TTT, Dao M, Chandramohanadas R (2015) Single-cell evaluation of red blood cell bio-mechanical and nano-structural alterations upon chemically induced oxidative stress. Sci Rep 5:9768

Suresh S (2007a) Biomechanics and biophysics of cancer cells. Acta Biomater 3(4):413–438

Suresh S (2007b) Nanomedicine – elastic clues in cancer detection. Nat Nanotechnol 2(12):748–749

Suresh S et al (2005) Connections between single-cell biomechanics and human disease states: gastrointestinal cancer and malaria. Acta Biomater 1(1):15–30

Tang YH et al (2017) OpenRBC: a fast simulator of red blood cells at protein resolution. Biophys J 112(10):2030–2037

Ware RE, de Montalembert M, Tshilolo L, Abboud MR (2017) Sickle cell disease. Lancet 390(10091):311–323

Wells TNC, van Huijsduijnen RH, Van Voorhis WC (2015) Malaria medicines: a glass half full? Nat Rev Drug Discov 14(6):424–442

Xu MJ et al (2017) A deep convolutional neural network for classification of red blood cells in sickle cell anemia. PLoS Comput Biol 13(10):e1005746

Xu XF et al (2013) Probing the Cytoadherence of malaria infected red blood cells under flow. PLoS One 8(5):e64763

Zhang Y et al (2015) Multiple stiffening effects of nanoscale knobs on human red blood cells infected with *Plasmodium falciparum* malaria parasite. Proc Natl Acad Sci U S A 112(19):6068–6073

Xin Bian and Matej Praprotnik

Contents

Abstract

Domain decomposition methods (DDM), which originate from the Schwarz alternating method to solve elliptic partial different equations, are largely extended and prove to have increasing influences on multiscale modeling of materials. We discuss some of the important extensions of the DDM in the fields

X. Bian (✉)
Computational Science and Engineering Laboratory, Clausiusstrasse, ETH Zurich, Switzerland
e-mail: bianx@ethz.ch

M. Praprotnik
Laboratory for Molecular Modeling, National Institute of Chemistry, Ljubljana, Slovenia

Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
e-mail: praprot@cmm.ki.si

of multiscale modeling for soft materials such as simple and complex fluids. To this end, we typically model the fluids in two or more levels of detail, which exploits the computational efficiency of the coarse model and physical accuracy of the fine description. For simple fluids, we take a continuum perspective to couple the molecular dynamics (MD) and Navier-Stokes equations by matching the state variables and/or fluxes across the hybrid interface. For complex fluids, we take a discrete perspective to encompass the complex structure of the molecules and couple the MD with coarse-grained MD by interpolating the forces between the two levels of descriptions.

1   Introduction

Domain decomposition methods (DDM) may represent a generic concept of solving a problem defined on a global domain by first dividing the domain into a set of overlapping/non-overlapping subdomains. Given that the set of the subdomains covers the original domain and a local problem is solved in each subdomain, the solution for the original global problem is available. Therefore, in a nutshell, it is a methodology of "divide and conquer." However, such a generic concept may be interpreted quite differently in various disciplines.

The name of DDM is mostly referred in applied mathematics, where it represents the decomposition of partial differential equation (PDE) or its approximation on the whole domain into a set of coupled problems on interconnecting subdomains. This dates back to the classical Schwarz alternating method proposed in 1870 (Smith et al. 1996), where the solution of an elliptic boundary value problem is obtained by alternatively solving two same elliptic boundary value problems defined in the two overlapping subdomains. For a general classification of DDM, the decomposition of domain may enter (Toselli and Widlund 2005):

1. At the continuum level, where different PDE models are appropriate for different regions. For example, in the applications of fluid-structure interactions, the fluid and solid regions are described by PDEs of fluid and solid mechanics, respectively. Another example is the aerodynamic boundary layer proposed by Ludwig Prandtl in 1904 (Landau and Lifshitz 1987); in the vicinity of bounding surface, viscous effects of the fluid are significant, while far away from the surface, a description of inviscid flow is sufficient.
2. At the discretization level, where different approximation schemes are employed in different regions. For example, a finite difference method is adopted in one region, while a finite element method is preferred in the other region.
3. At the algebraic level, where a large linear (sparse) system of equations results from the discretization of the PDE. A global iterative method may be inefficient so that a strategy from DDM may choose to solve each block of the linear system individually by employing a direct or iterative method. In this context, the DDM is considered as a preconditioner in combination with other acceleration

techniques such as Krylov space methods. The numerical analysis of DDM as a preconditioner is very developed, which may be found in excellent monographs (Smith et al. 1996; Quarteroni and Valli 1999; Toselli and Widlund 2005).

In practice of numerical modeling, the decomposition at these three levels mentioned above may be interconnected.

Utilizing the idea of DDM for multiscale modeling with heterogeneous mathematical/physical descriptions is also not a new concept. In computational chemistry, one of the earliest examples is the QM/MM (quantum mechanics/molecular mechanics) (Warshel and Karplus 1972; Warshel and Levitt 1976), where chemical reactions involving large molecules are handled by quantum mechanics and regions at elsewhere by classical models. Typically, the degrees of freedom for the two descriptions are fixed during the course of the simulation.

A sudden increase of attention in applying DDM for multiscale modeling of fluids started with the paper of O'Connell and Thompson (1995), where one subdomain is simulated by molecular dynamics (MD) with Lennard-Jones potential, while the other subdomain is solved by a finite difference method discretization of the Navier-Stokes (NS) equations. The overlapping region of the two subdomains implements the "handshaking" of the physical models at two scales. Since the inception of this work, many further important contributions emerge, especially on the constraint dynamics of MD's artificial boundary such as the Maxwell buffer (Hadjiconstantinou and Patera 1997; Hadjiconstantinou 1999), flux-exchange method (Flekkøy et al. 2000; Delgado-Buscalioni and Coveney 2003a; Flekkøy et al. 2005; De Fabritiis et al. 2006), least constraint dynamics (Nie et al. 2004; Werder et al. 2005), adaptive forcing (Pivkin and Karniadakis 2006; Fedosov and Karniadakis 2009), and so on. These methodologies are from a continuum perspective of fluid mechanics, and the coupling of two distinct scales is mainly for modeling simple liquids. There exist a few excellent review articles on this topic (Wijesinghe and Hadjiconstantinou 2004; Mohamed and Mohamad 2010; Delgado-Buscalioni 2012).

Another important research line on DDM for multiscale modeling of structured or complex fluids is represented by the adaptive resolution scheme or AdResS from computational physics/chemistry. In one subdomain, a detailed MD model with atomic details is employed, while in the other subdomain, a coarse-grained (CG) model is adopted. The CG model is usually constructed off-line and, if required, is able to reproduce quantities of interest from MD faithfully. A complex molecule may diffuse across the interface between the two subdomains and change degrees of freedom gradually. In the overlapping region of the two subdomains, either force or energy is interpolated smoothly between the two distinct descriptions. Previous excellent technical reviews on this research line are also available (Praprotnik et al. 2008; Delle Site and Praprotnik 2017). Another interesting hybrid approach has been proposed recently in Scukins et al. (2015), where the hybrid MD/continuum hydrodynamics system of liquid water is considered as two completely miscible liquids using two-phase modeling, where one phase corresponding to the MD and the other to fluctuating continuum hydrodynamics.

For modeling biological functions and diseases, especially those being localized such as the endothelial glycocalyx layer (Weinbaum et al. 2007; Deng et al. 2012) and blood clot formation (Fogelson and Neeves 2015), the DDM appear to be very promising because one can treat different subdomains of the system using different models and methods. In this chapter, we discuss some of the important extensions of the DDM in the fields of multiscale modeling for simple and complex fluids. In the following, we shall first describe the DDM for simple fluids, which include some introduction materials on MD and NS equations. Furthermore, in the same section, we illustrate a few effective algorithms to couple the MD and NS descriptions, with a focus on constraining the artificial boundary of MD. Subsequently, we revisit the concepts of DDM for complex fluids or AdResS. Along the discussions of fundamental algorithms, we illustrate a few selected applications. We summarize this chapter in the end with some perspectives.

2 Domain Decomposition Methods for Simple Fluids

2.1 Molecular Dynamics

We consider a system of N identical, spherical, and structureless particles; they have the same mass m and are enclosed in a volume $\mathscr{V}$. The Hamiltonian $\mathscr{H}$ of the system is defined as

$$\mathscr{H}(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^{N} \left[\frac{|\mathbf{p}_i|^2}{2m} + u_i(\mathbf{r}^N) \right], \tag{1}$$

where the first and second terms on the right-hand side are the kinetic and potential energies, respectively. The value of $(\mathbf{r}^N, \mathbf{p}^N)$ defines a phase point, where $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ are $3N$-dimensional coordinates and $\mathbf{p}^N = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N)$ are $3N$-dimensional momenta. In addition, velocity is defined as $\mathbf{v}_i = \mathbf{p}_i/m$. The trajectory of a phase point is determined by Hamilton's equations:

$$\dot{\mathbf{r}}_i = \frac{\partial \mathscr{H}}{\partial \mathbf{p}_i}, \quad \dot{\mathbf{p}}_i = -\frac{\partial \mathscr{H}}{\partial \mathbf{r}_i}. \tag{2}$$

More specifically, we consider an argon-like liquid described by a shifted Lennard-Jones (LJ) potential truncated at $r_{ij} = r_c$

$$u(r) = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 - \left(\frac{\sigma}{r_c}\right)^{12} + \left(\frac{\sigma}{r_c}\right)^6 \right], & r \leq r_c \\ 0, & r > r_c \end{cases} \tag{3}$$

where ε and σ set the characteristic energy and length scales, respectively. Relative distance between particles is $r_{ij} = |\mathbf{r}_{ij}|$ and $\mathbf{e}_{ij} = \mathbf{r}_{ij}/r_{ij}$. The potential is pairwise,

and therefore, the potential energy of particle i is $u_i(\mathbf{r}^N) = 1/2 \sum_j^{N_{\text{neigh}}} u(r_{ij})$, where N_{neigh} is the number of neighboring particles within a spherical region of radius r_c around particle i. Hence, a pairwise force between particles i and j reads

$$\mathbf{F}_{ij} = -\frac{\partial u(r_{ij})}{\partial r_{ij}} \mathbf{e}_{ij}. \tag{4}$$

The MD trajectories evolve by a time integrator, such as the Verlet velocity algorithm (Allen and Tildesley 1989; Tuckerman 2010) with time step δt.

2.2   Navier-Stokes Equations

We consider an isothermal description of the fluids, where the energy or entropy change of the fluids is ignored. We have the conservations of mass and momentum densities as (Landau and Lifshitz 1987)

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0, \tag{5}$$

$$\frac{\partial \rho \mathbf{V}}{\partial t} + \nabla \cdot \Pi = 0, \tag{6}$$

where ρ, $\mathbf{V}$, and Π are density, velocity, and stress fields, respectively. Here velocity is capitalized to differentiate with the MD particle's velocity $\mathbf{v}$. For a compressible Newtonian fluid, the stress reads

$$\Pi = \rho \mathbf{V}\mathbf{V} - p\mathbf{I} - \eta \left(\nabla \mathbf{V} + \nabla \mathbf{V}^T \right) + \mathbf{I} \left(\frac{2}{3}\eta - \zeta \right) \nabla \cdot \mathbf{V}, \tag{7}$$

with $\mathbf{I}$ being the identity tensor and η and ζ the dynamic and bulk viscosities, respectively. The equation of state $p = p(\rho)$ and viscosities are usually obtained from MD simulations in the context of DDM for multiscale modeling. For an incompressible fluid, the term $\nabla \cdot \mathbf{V} = 0$ is dropped.

The Navier-Stokes equations are typically solved by mesh-based methods, such as finite difference, finite volume, and finite element, or other Lagrangian mesh-free particle methods.

2.3   Concurrent Coupling Between Particle Dynamics and Continuum Description

For illustrative purpose, we consider a specific hybrid simulation between MD and NS equations solved by a finite difference method (FDM). To demonstrate the concept, we take an example of unidirectional flow such as the Couette flow and
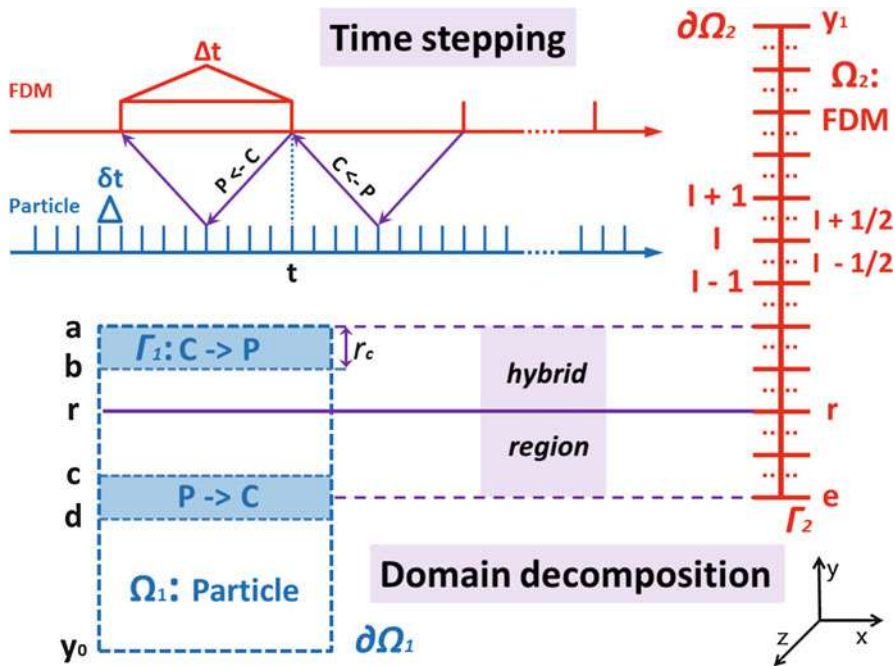
Fig. 1 (Color online) Sketch of a domain decomposition method and time stepping for coupling discrete particle dynamics with finite difference scheme. Subdomain Ω_1 is described by a molecular dynamics (MD) simulation, while subdomain Ω_2 is described by a finite difference method (FDM). The whole domain is bounded by two walls $\partial\Omega_1$ and $\partial\Omega_2$, at $y = y_0$ and y_1, respectively. An artificial boundary $\Gamma_1 : [y_b, y_a]$ applies to the MD, and an artificial boundary $\Gamma_2 : y = y_e$ applies to the FDM. A hybrid reference line is defined at $y = y_r$, at which the global solution is pieced together by combining results from MD below y_r and results from FDM above y_r. FDM runs with time step Δt, and DPD runs with time step δt, where $\Delta t = N_s \delta t$ and $N_s \gg 1$ is a positive integer. Time integrations of FDM and MD are staggered to synchronize. The communication time step between the two simulations is $\Delta t_{\text{comm}} = \Delta t$. The arrows between two time axes indicate information passing. (Modified from Bian et al. (2016))

couple the two descriptions along the y direction. This is best sketched in Fig. 1. The hybrid reference line is defined at y_r, and the global solution combines the solution of the MD below y_r and the solution of the FDM above y_r. The subdomain of MD is a three-dimensional box spanning $[0, L_{MD}] \times [0, L_{MD}]$ in x and z direction. In y direction, it ranges from $y_0 = 0$ to y_a. In the case of a wall-bounded domain, there are solid particles below y_0, and specific molecular interactions at fluid-solid interfaces that are different from fluid-fluid interactions must be taken into account. The FDM domain ranges from y_e to y_1 in y direction. No-slip or partial slip condition may apply at y_1 for the continuum description.

The time integrations of the two simulations are staggered to synchronize the average time progression. To minimize the lag time of information exchange (Bian et al. 2015b), we take communication time step $\Delta t_{\text{comm}} = \Delta t = N_s \delta t$. A typical

choice of N_s is an integer number so large ($N_s \gg 1$) that individual particle's velocity decorrelates after $N_s \delta t$ MD steps. This signifies the timescale separation of the two distinct descriptions of fluids; over each Δt the FDM follows a description of continuum mechanics, meanwhile the relevant molecular process reaches a quasi-stationary state. Therefore, given the intrinsic δt of MD and stability constraint on Δt, scale separation parameter N_s demands a low threshold for the grid size of FDM.

To couple the two scale descriptions, we have to pay more attention to the two artificial boundary regions Γ_1: $C \rightarrow P$ (continuum to particles) and Γ_2: $P \rightarrow C$ (particles to continuum). More specifically, on the communication between the two solvers, the particle simulation integrates with time step δt from $t = t - 3\Delta t/2$ to $t = t - \Delta t/2$. To impose the artificial boundary Γ_2 for the continuum ($P \rightarrow C$) at $t - \Delta t$, it is straightforward to perform simple spatial averaging on particle velocities between $y_d \leq y < y_c$ and thereafter temporal averaging over $t - 3\Delta t/2 < t \leq t - \Delta t/2$ as, for example, for the velocity state variable,

$$\mathbf{V}_e = \frac{1}{N_s} \sum_{j=1}^{N_s} \frac{1}{N_{cd}} \sum_{i=1}^{N_{cd}} \mathbf{v}_i. \tag{8}$$

Here, N_{cd} is the instantaneous number of particles located in cell $P \rightarrow C$. Furthermore, the region $P \rightarrow C$ is centered at $y = y_e$ with a thickness of r_c. Similarly, we may alternatively come up easily with the averaging procedures to impose mass and momentum fluxes at interface y_e of the FDM for compressible fluids as (Flekkøy et al. 2000)

$$\rho \mathbf{V} \cdot \mathbf{e}_y = \frac{1}{v_c} \sum_i m \mathbf{v}_i \cdot \mathbf{e}_y, \tag{9}$$

$$\Pi \cdot \mathbf{e}_y = \frac{1}{v_c} \sum_i \left(m \mathbf{v}_i \mathbf{v}_i + \frac{1}{2} \sum_{j \neq i} \mathbf{F}_{ij} \mathbf{r}_{ij} \right), \tag{10}$$

where $\mathbf{e}_y$ is the unit vector pointing outward at $y = y_e$ from FDM and v_c is the local cell volume centered at $y = y_e$ for spatial averaging. Furthermore, Eqs. (9) and (10) need temporal averaging over $t - 3\Delta t/2 < t \leq t - \Delta t/2$ as is done for Eq. (8). For non-isothermal case, energy flux can be imposed as well at the continuum interface (Delgado-Buscalioni and Coveney 2003a).

For more complex flow scenarios, a simple local-time-and-space averaging of particle quantities may not provide sufficiently smooth boundary conditions for the continuum description and may result in instability of the continuum solutions. Therefore, novel algorithms for noise reduction (Li et al. 1998; Grinberg 2012) are called for. Another alternative is to couple one continuum solver with multi-instance MD simulations, so the concept of ensemble averaging may be exploited (Neumann and Bian 2017).

Once $\mathbf{V}_e$ or the fluxes are updated at $t - \Delta t$, the FDM integrates one step from $t - \Delta t$ to t. Its new state variable solutions $\mathbf{V}_a$ and $\mathbf{V}_b$ at y_a and y_b or fluxes at y_a are passed to the particle simulation and are further utilized in various constraint dynamics (as explained in the next four subsections) to set the artificial boundary $\Gamma_1 : y_b \leq y < y_a$ of the particle simulation from $t - \Delta t/2$ to $t + \Delta t/2$. Meanwhile, the MD simulation integrates with time step δt from $t - \Delta t/2$ to $t + \Delta t/2$. Afterward, again Eq. (8) or Eqs. (9) and (10) are applied to impose state or flux boundary conditions of Γ_2 at t. The alternation of imposing artificial boundaries and the staggered integrations of the two simulations repeat until the end of the hybrid simulation.

Due to the truncation of the subdomain for the MD simulation, an average pressure force $\mathbf{F}^P(y)$ must be imposed at the truncation line $y = y_a$ inward. The best way to devise $\mathbf{F}^P(y)$ so far is to make use of the radial distribution function of particles and MD interacting forces to reflect the average effects outside y_a in the continuum limit. $\mathbf{F}^P$ is position-dependent and applies to each particle i within $y_a - y_i < r_c$ distance from y_a in the interior, that is, particles in Γ_1. This approach proves to induce negligible density oscillations near the artificial boundary (Werder et al. 2005; Lei et al. 2011). However, the imposed averaged force does not completely prevent particles from leaving outside. Therefore, a specular reflection or particle deletion is implemented at y_a. The former approach is more suitable to deal with an incompressible fluid, while the latter approach together with a particle insertion procedure (Delgado-Buscalioni and Coveney 2003b) is more physical to describe a compressible fluid by a grand-canonical ensemble so that the total number of particle in the MD subdomain fluctuates (Flekkøy et al. 2005).

The tricky part is to impose the artificial boundary in $\Gamma_1(C \rightarrow P)$ for MD particles from $t - \Delta t/2$ to $t + \Delta t/2$. There is *no unique way* to achieve this, as many extra degrees of freedom on particles are under-determined. For coupling state variable, the usual strategy is to perform a constraint dynamics on each particle i in Γ_1 in such a way that

$$\frac{1}{N_{\Gamma_1}} \sum_{i=1}^{N_{\Gamma_1}} \mathbf{v}_i = \overline{\mathbf{V}}_{\Gamma_1}, \tag{11}$$

is satisfied at every δt or on average over $N_s \delta t$ and thermal fluctuations are affected as little as possible. Here, N_{Γ_1} is the instantaneous number of particles in Γ_1, and $\overline{\mathbf{V}}_{\Gamma_1}$ is the average velocity of the continuum solutions in the same region. In this case, $\overline{\mathbf{V}}_{\Gamma_1} = (\mathbf{V}_a + \mathbf{V}_b)/2$ at time t. For coupling the fluxes, it is similar that we need to spread fluxes (few degrees of freedom) obtained from the continuum solution to the particles within the artificial boundary Γ_1.

In the following, we shall focus on four popular algorithms of constraint dynamics in Γ_1. Since we deal primarily with a unidirectional flow for illustrative purpose, we use scalar variables instead of vector ones. The constraint of the particle dynamics is always performed in x direction of the flow, and therefore the index for x is omitted. Velocities in y and z directions are not altered by the constraint dynamics.

2.4   Relaxation Dynamics

A relaxation dynamics method was proposed in the first paper of hybrid simulations on liquid flow (O'Connell and Thompson 1995). Effectively it reads as

$$\dot{v}_i = \frac{F_i}{m} + \frac{\varepsilon}{\delta t}\left(\overline{V}_{\Gamma_1} - \frac{1}{N_{\Gamma_1}}\sum_{j=1}^{N_{\Gamma_1}} v_j\right),\tag{12}$$

where F_i is the usual total particle force on particle i. The rest of the terms on the right-hand side act as a relaxing force, which drives each particle i in Γ_1 toward the average velocity $\overline{V}_{\Gamma_1}$ over $\delta t/\varepsilon$ time period. In O'Connell and Thompson (1995), the continuum has the same time step as that of the particle simulation, that is, $\Delta t = \delta t$. Moreover, the relaxation parameter is chosen as $\varepsilon = 0.01$ for the particular properties of the fluid simulated, where the authors argue that a smaller value of ε provides inadequate coupling, while a larger value may lead to excessive damping of thermal fluctuations.

2.5   Maxwell Buffer

The Maxwell buffer was first proposed in Hadjiconstantinou and Patera (1997) as

$$v_i = V_i + \delta v_i,\tag{13}$$

$$V_i = V_b + (V_a - V_b)(y_i - y_b)/\Delta y,\tag{14}$$

where a deterministic component V_i is obtained by a linear interpolation between the values on the two nearest grid points V_a and V_b. Under the assumption of a local equilibrium for MD, the stochastic component δv_i is drawn from the Maxwell-Boltzmann distribution at a given temperature $k_B T$,

$$p(\delta v_i) = \sqrt{\frac{m}{2\pi k_B T}}\exp\left[\frac{-m(\delta v_i)^2}{2k_B T}\right].\tag{15}$$

It is simple to see that Eqs. (13) and (14) satisfy directly the constraint posed in Eq. (11).

2.6   Least Constraint Dynamics

By taking the extremum of the time integral of the Lagrangian for the particles in Γ_1, which is subject to the non-holonomic constraint in Eq. (11), the equation of motion (EoM) for each particle i in Γ_1 is according to (Nie et al. 2004)

$$\dot{v}_i = \frac{F_i}{m} - \frac{1}{N_{\Gamma_1}} \sum_{j=1}^{N_{\Gamma_1}} \frac{F_j}{m} + \frac{1}{\delta t} \left(\overline{V}_{\Gamma_1} - \frac{1}{N_{\Gamma_1}} \sum_{j=1}^{N_{\Gamma_1}} v_j \right). \tag{16}$$

Instead of repeating the derivation for Eq. (16) as in Nie et al. (2004), we may arrive at the same expression from another perspective. To satisfy the constraint in Eq. (11), an extra body force $F_{\Gamma_1}^b$ is introduced *dynamically* at every time step δt such that (Werder et al. 2005)

$$\overline{v}'_{\Gamma_1} = \overline{v}_{\Gamma_1} + \frac{\delta t}{N_{\Gamma_1} m} (F_{\Gamma_1} + F_{\Gamma_1}^b), \tag{17}$$

where $\overline{v}_{\Gamma_1}$ is the average velocity and F_{Γ_1} is the total force in Γ_1 due to particle interactions,

$$\overline{v}_{\Gamma_1} = \frac{1}{N_{\Gamma_1}} \sum_{j=1}^{N_{\Gamma_1}} v_j, \quad F_{\Gamma_1} = \sum_{i=1}^{N_{\Gamma_1}} F_j. \tag{18}$$

To satisfy Eq. (11), we set $\overline{v}'_{\Gamma_1} = \overline{V}_{\Gamma_1}$, and therefore, we obtain an expression for the body force $F_{\Gamma_1}^b$ as

$$F_{\Gamma_1}^b = \frac{N_{\Gamma_1} m}{\delta t} (\overline{V}_{\Gamma_1} - \overline{v}_{\Gamma_1}) - F_{\Gamma_1} = \frac{N_{\Gamma_1} m}{\delta t} \left(\overline{V}_{\Gamma_1} - \frac{1}{N_{\Gamma_1}} \sum_{j=1}^{N_{\Gamma_1}} v_j \right) - \sum_{j=1}^{N_{\Gamma_1}} F_j. \tag{19}$$

The value of $F_{\Gamma_1}^b$ is dynamic and may vary at every δt. If we spread the body force $F_{\Gamma_1}^b$ from Eq. (19) *evenly* on N_{Γ_1} particles in Γ_1, it gives exactly the same EoM as in Eq. (16) for each particle i.

2.7   Flux-Exchange Coupling

The exchange of flux for dense fluids was first proposed in Flekkøy et al. (2000) and extended in Delgado-Buscalioni and Coveney (2003a). Rather than constraining the state variables directly, as done in the previous three methods, this method imposes flux at the truncation line y_a of the subdomain of particle simulation. Therefore, the EoM of particles in Γ_1 reads

$$\dot{v}_i = \frac{F_i}{m} + F^x(y_i), \quad F^x(y_i) = \tau_a^{xy} A\lambda(y_i), \tag{20}$$

where $A = L_x \times L_z$ is the surface area of the truncation and τ_a^{xy} is the shear stress at y_a from the continuum. The distribution function $\lambda(y_i)$ of the shear force on each particle must be normalized

$$\lambda(y_i) = g(y_i)/\sum_{j=1}^{N_{\Gamma_1}} g(y_j), \tag{21}$$

where $g(y)$ is an arbitrary function so that $\lambda(y)$ diverges at $y = y_a$ and decays to zero as y approaches y_b from y_a.

If we assume that there is a locally linear shear flow within Γ_1, which is a reasonable assumption for many flow problems, then the distribution function $\lambda(y)$ may be defined better than an arbitrary one (Ren 2007). By assuming a locally linear shear flow, we can work out the continuum limit of the shear forces on line y_a from particles within Γ_1. By setting the shear force on y_a from particles equal to the continuum solution of FDM at y_a, we have an identity as follows:

$$A\tau_a^{xy} = \int_0^{r_c} A d\gamma^D(h) h\dot{\gamma} dh, \tag{22}$$

where h is the distance of particle i from y_a ($h = y_a - y_i$) and $\dot{\gamma}$ is an arbitrary shear rate, while given the particle interaction potential, $\gamma^D(h)$ can be numerically evaluated (Bian et al. 2016). Equation (22) holds for an arbitrary shear rate $\dot{\gamma}$. Therefore, given a shear stress τ^{xy} from the continuum solution of FDM, the distribution of shear force on each particle i is independent of $\dot{\gamma}$ as

$$F^x(y_i) = B_0 \tau_a^{xy} \gamma^D(h) h, \tag{23}$$

where B_0 is a normalization constant for the particular particle simulation parameters.

2.8   Some Comments on Different Coupling Strategies

In order to bridge the concepts of state and flux coupling, Walther et al. (2012) applied the gradient of velocities averaged from MD simulations to impose the flux boundary condition on the lattice Boltzmann continuum solver. For conservation and stability purposes, we may have also a mixture of coupling state variables and fluxes; we refer to Flekkøy et al. (2005), Ren (2007), Bian et al. (2015b, 2016) and references therein for further discussions.

On the accuracy of the coupling schemes, the quantity of interest has been often the *mean field* variables, such as velocity, density, and stresses, which are sufficient from the continuum perspective. There has not been a consensus as to which of the four methods is most effective, as all the methods introduce above are actually capable of generating accurate mean profiles. For practical convenience, the relaxation dynamics involves an empirical parameter ε, which requires more trial-and-tune simulations. Therefore, the least constraint dynamics is preferred over the relaxation dynamics. The Maxwell buffer and flux exchange have the same easiness as the least constraint dynamics.

However, if we take a microscopic perspective and take the continuum description as a bath to supplement the truncated particle simulation, we may evaluate more closely certain quantities of interest in the particle region. For example, by calculating the fluctuation correlations between field variables of MD, we may compare the performance of the coupling schemes mentioned above (Bian et al. 2015a, 2016, 2018). It turns out that we have two sources of errors in the particle region; one is due to the truncation of the particle domain, and the other is due to the constraint dynamics performed on the particles in the artificial boundary. The first source of error is universal for different coupling strategies, and it may be almost completely removed by implementing an appropriate particle removal and insertion algorithm to mimic the grand-canonical ensemble of the an open particle system (Delgado-Buscalioni and Coveney 2003b; Delle Site and Praprotnik 2017). On the second source of error, the Maxwell buffer completely ignores any intrinsic thermal velocity correlations between particles and simply constrains individual particle's velocity as identical independent Gaussian distribution in the artificial boundary, therefore introducing the most artifacts, whereas the other constraint algorithms are relatively mild and commit much smaller error on this aspect (Bian et al. 2016).

Furthermore, if the overlapping region between particle and continuum subdomain increases, that is, if we can afford to perform extra volume of particle simulation, the contaminations of all the constraint dynamics on the thermal fluctuation correlations have very similar behavior, and the negative effects decrease linearly as the size of the overlap increases (Bian et al. 2016).

Thus far, there has been quite a few applications which were enabled by the DDM for multiscale modeling, bridging the atomistic and continuum descriptions. One of the most apparent classes of application is to model local or singular effects accurately with MD and capture bulk macroscopic flow behavior elsewhere with NS equations. This type of problem includes the fluid-solid boundary conditions with flow slip or epitaxial growth taking place (Thompson and Robbins 1990; O'Connell and Thompson 1995), contact line problem with contact angle as a dynamic variable (Hadjiconstantinou 1999), lid-driven cavity flow with singular stress at the corner (Nie et al. 2006), sound propagation through lipid monolayer (De Fabritiis et al. 2006), platelet deposition on the wall of a brain aneurysm (Grinberg et al. 2013), and so on. In the next section, however, we will describe a triple-scale approach that couples atomistic and continuum hydrodynamics through an intermediate CG description.

3   Domain Decomposition Methods for Complex Fluids: Adaptive Resolution Scheme

Another branch of DDM is related to coupling fine and CG descriptions of soft matter and molecular liquids within a particle-based framework. Similarly to other domain decomposition approaches, the aim of such adaptive resolution simulations is to model the interesting part of the system with the detailed atomistic (AT)

resolution, which is coupled with a CG representation of the remaining part of the system. Among the most advanced multiscale methods for conducting this kind of molecular simulations is the adaptive resolution scheme (AdResS) (Praprotnik et al. 2005, 2008, 2011). The method can successfully couple two or more levels of resolution concurrently present in the system. AdResS enables the particles to change their resolution on the fly, during the course of an MD simulation. The method is suitable for molecular systems where the AT resolution is required only in a spatially localized region, whereas a lower CG level of detail is sufficient for the rest of the system. Such cases are typically found in simulations of biological macromolecules, e.g., DNA. There, the AT resolution is required only for the macromolecule and the solvent in its vicinity, whereas the solvent farther away is adequately treated on a simplified CG level.

AdResS allows molecules to freely move across different regions and change their resolution on the fly according to their position in the system. When a CG molecule leaves the CG domain, it is remapped into the atomistically resolved molecule with a random orientation. To avoid any overlaps of its atoms with the atoms of the neighboring molecules, it is required that introduction of the atomistic degrees of freedom is continuous and not instantaneous. To this end, an interface layer between the AT and CG regions is introduced that allows an atomistic molecule to gradually find an energetically permissible orientation with respect to its neighboring molecules. This transition region, also called a hybrid (HY) region, contains hybrid molecules where both representations are superimposed.

Two levels of resolution are coupled via a force interpolation scheme. The intermolecular force between given molecules α and β is defined as

$$\mathbf{F}_{\alpha\beta}^{\text{AdResS}} = w(X_\alpha)w(X_\beta)\mathbf{F}_{\alpha\beta}^{AT} + [1 - w(X_\alpha)w(X_\beta)]\mathbf{F}_{\alpha\beta}^{CG}, \tag{24}$$

where X_α and X_β are the centers-of-mass positions of the molecules α and β, respectively, and w is a weighting function that governs the transition between different resolution regions and therefore needs to be smooth. This function depends on the position of the mapping point of the molecule and is defined in the following way: $w = 1$ corresponds to the AT region and $w = 0$ to the CG region, whereas the values $0 < w < 1$ correspond to the HY region.

Original variant of the AdResS based on the force coupling above does not allow for a definition of a Hamiltonian and consequently has to be employed in conjunction with a thermostat. A recent Hamiltonian version of the method, i.e., H-AdResS (Potestio et al. 2013b; Everaers, R. 2016; Español et al. 2015), on the other hand, allows for the definition of a global Hamiltonian. This enables one to perform adaptive resolution Monte Carlo simulations (Potestio et al. 2013a). However, as the translation invariance is broken due to the resolution change, this implies that the total linear momentum cannot be conserved by H-AdResS. Since the linear momentum conservation is crucial for hydrodynamics, the original AdResS, which is not Hamiltonian, can therefore preserve linear momentum despite broken translational symmetry and is thus more convenient for coupling with the continuum hydrodynamics, as described below. Recent extensions of AdResS

also involve coupling to a quantum level of description (Poma and Delle Site 2010, 2011; Agarwal and Delle Site 2015, 2016; Delle Site 2018) as well as to open systems that exchange mass, momentum, and energy with their surroundings (Delgado-Buscalioni et al. 2015; Sablić et al. 2016; Wang et al. 2013; Agarwal et al. 2014, 2015; Delle Site 2016; Mukherji and Kremer 2013; Kreis et al. 2015). This allows for conducting MD simulations either in the grand-canonical statistical ensemble or under non-equilibrium conditions (Delgado-Buscalioni et al. 2015; Sablić et al. 2016, 2017b; Delle Site 2018). AdResS has been successfully applied to biomolecular systems (Zavadlav et al. 2017) such as proteins (Zavadlav et al. 2014; Fogarty et al. 2015) and DNA molecules (Zavadlav et al. 2015a, b, 2016b) solvated in multiscale solvents (Bevc et al. 2013; Nagarajan et al. 2013; Zavadlav et al. 2014, 2015b). In connection to hydrodynamics phenomena, AdResS has also been used to couple MD to the dissipative particle dynamics (DPD) method (Hoogerbrugge and Koelman 1992; Español and Warren 1995, 2017; Groot and Warren 1997; Español 1995), the multiparticle collision dynamics (MPC) (Malevanets and Kapral 1999), and smoothed dissipative particle dynamics (SDPD) (Español and Revenga 2003; Vázquez-Quesada et al. 2009). In the DPD, groups of atoms/molecules are lumped to form soft beads that interact via explicit soft conservative, random, and dissipative forces to simulate fluids on a mesoscopic scale with correct hydrodynamics interactions. In the SDPD, the Navier-Stokes equations are numerically solved with a formalism that is reminiscent of MD, whereas in the MPC the system is modeled by particles with continuous positions and velocities and stochastic interparticle interactions. Thus, using AdResS Petsev et al. (2015, 2017) coupled the MD to SDPD, whereas Alekseeva et al. (2016) linked the MD with MPC. The robustness of both hybrid approaches was demonstrated on a Lennard-Jones fluid. AdResS has been also applied to couple MD with DPD water (Zavadlav and Praprotnik 2017). In this chapter, however, we will focus on coupling with the continuum hydrodynamics (Delgado-Buscalioni et al. 2008, 2009, 2015).

One can combine AdResS with the flux-exchange coupling (Flekkøy et al. 2000; Delgado-Buscalioni and Coveney 2003a, 2004; Delgado-Buscalioni et al. 2005; Flekkøy et al. 2005; De Fabritiis et al. 2006; Delgado-Buscalioni and De Fabritiis 2007), described in Sect. 2.7, to link an MD with a continuum hydrodynamic domain, resulting in a triple-scale setup (Delgado-Buscalioni et al. 2008, 2009), as depicted in Fig. 2. In this triple-scale approach, the particle-based domain is simulated by MD simulation. The dynamics of molecules is thus governed by Newton's equations of motion, as explained in Sect. 2.1. On the other hand, continuum description enables the study of macroscopic fluid flows. The fluid is described by Navier-Stokes equations (see Sect. 2.2). The Navier-Stokes equations can be numerically solved in different ways (Koumoutsakos 2005), e.g., the finite volume method, where the continuum region is divided into small cells of volume (De Fabritiis et al. 2006; Delgado-Buscalioni et al. 2008). In such a way, discretized Navier-Stokes equations are then integrated in time using an explicit Euler scheme (Delgado-Buscalioni and De Fabritiis 2007).

The MD and continuum domains share an interface, as shown in Fig. 2. The otherwise independent MD and continuum domains exchange information after
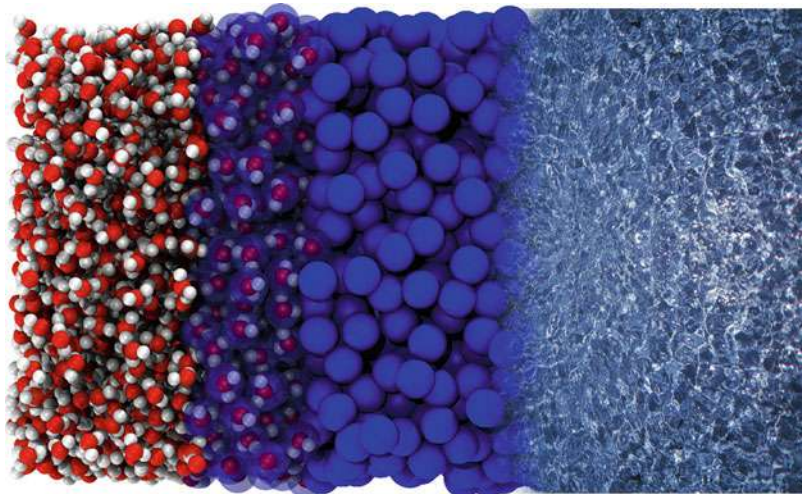
Fig. 2 Domain decomposition of triple-scale liquid water system: coupling of atomistic and continuum hydrodynamics. The MD domain contains atomistic and coarse-grained water molecules, whereas the continuum domain is solved by computational fluid dynamics approach. (Reprinted from Delgado-Buscalioni et al. 2015)

every fixed time interval Δt (Delgado-Buscalioni et al. 2008). Flux balance implies the conservation of mass and momentum across the interface, i.e., both domains should receive equally large but oppositely signed mass and momentum transfer across the interface over each Δt. The momentum flux across the interface is then used to update the flow variables at the continuum boundary cells. In turn, the same (but oppositely signed) flux needs to be imposed into the particle system across the interface (see Sect. 2.3).

The triple-scale method has been applied to simple liquids, i.e., liquid of tetrahedral molecules and water (Delgado-Buscalioni et al. 2008, 2009). Equilibrium structural properties, i.e., the radial distribution function and density profile, in MD domain of the triple-scale system have been validated against reference all-atom MD simulations. Furthermore, the mass fluctuations of the open MD domain were demonstrated to agree with the theoretical grand-canonical predictions (Delgado-Buscalioni et al. 2009). The fluids were also exposed to the Couette and Stokes flows (Delgado-Buscalioni et al. 2008, 2009), and the resulting velocity profiles agreed with continuum hydrodynamics.

Of course, the underlying idea of this triple-scale scheme is that AdResS plays the role of the inserting facility for complex molecules such as star polymers (Delgado-Buscalioni et al. 2015; Sablić et al. 2016, 2017a, b). In the CG domain, a given star polymer, consisting of several tens of monomers, is represented with only one very soft CG bead. We face the problem of inserting new molecules into a dense liquid. Thus, the idea behind the resolution change is that AdResS allows the insertion of molecules of arbitrary size into the system. The CG domains act as

a mass reservoir where large molecules can be easily inserted due to soft effective interactions among CG beads (Delgado-Buscalioni and Coveney 2003a; De Fabritiis et al. 2004; Borg et al. 2014). Then, as the molecules move toward the AT domain, they gain the fine-grained details employing AdResS.

4 Summary and Perspectives

We have presented the DDM for multiscale modeling of simple and complex fluids. For simple fluids we employ a perspective of continuum mechanics, where state field variables and fluxes are fundamental quantities. Therefore, it is natural to couple state variables and/or fluxes between a particle simulation (e.g., MD) simulation with a NS solver. Within this perspective, the MD resolves local effects and provides macroscopic averages as a closure for the NS equations. Other physical quantities in MD that are not needed for NS are considered as unwanted, ignored, or filtered out. Consequently, various constraint dynamics on state variable or fluxes of MD are all effective and do not have apparent difference on the hybrid coupled simulations. If we are interested not only in the macroscopic behavior but also in certain mesoscopic or microscopic physics of the particle simulation in the course of hybrid simulation, the conclusion is different. For example, thermal fluctuations are hallmarks, and one of the driving mechanisms for physical processes at micro-/mesoscopic scales. If we evaluate the correlations of the thermal fluctuations in the particle simulations in the context of a hybrid simulation, we find that the constraint algorithm of Maxwell buffer ignores the local correlations completely, while other three constraint dynamics are relatively better at preserving the natural properties of the particle simulations.

We have also briefly presented the particle-based multiscale linear-momentum-preserving AdResS for coupling fine and CG molecular representations. AdResS is suitable to simulate fluids on the micro-/mesoscopic scale, where hydrodynamics plays an important role. In this spirit, AdResS has been recently applied to link MD and DPD methods for simulation of water at ambient conditions (Zavadlav and Praprotnik 2017). As the DPD water model corresponds to several water molecules, the supramolecular coupling is enabled by a recently developed clustering algorithm SWINGER that assembles, disassembles, and reassembles clusters as needed during the course of the simulation (Zavadlav et al. 2016a). This allows for a seamless coupling between standard atomistic MD and DPD models. In the future, this framework could be applied to important problems such as abnormal rheological and biomechanical properties of red blood cells encountered in disease states (Chang et al. 2016; Altenhoff et al. 2007; Rossinelli et al. 2015; Fedosov et al. 2010, 2011).

In this chapter, we focus on the DDM for fluids, and for readers interested in DDM for solid materials, there are quite a few excellent technical reviews, e.g., Miller and Tadmor (2009). There are other multiscale modeling methods tightly related to DDM, such as *CONFFESSIT* (Laso and Öttinger 1993) and *heterogeneous multiscale method* (Ren and Weinan 2005; Weinan et al. 2007; Borg et al. 2013; Yasuda and Yamamoto 2010), where a macroscopic model is selected for

the whole domain and a microscopic solver is utilized to provide fine details such as boundary conditions or constitutive relations wherever needed. Along the research line of adaptive mesh refinement, Garcia et al. also develop the algorithm refinement technique at the finest scale of the mesh to switch on microscopic description such as the direction simulation of Monte Carlo (Garcia et al. 1999). Kevrekidis and colleagues propose space-selective microscopic descriptions without a macroscopic equation (therefore, named as *equation-free*) (Kevrekidis and Samaey 2009). By employing novel gap-tooth interpolation in space and projection in time, the equation-free framework is able to perform system-level tasks and predictions.

There are also a few high-performance software packages free for usage (Tang et al. 2015; Neumann and Bian 2017; Halverson et al. 2013). Together with the algorithm developments, they will foster further applications of DDM for multiscale modeling.

References

Agarwal A, Delle Site L (2015) Path integral molecular dynamics within the grand canonical-like adaptive resolution technique: simulation of liquid water. J Chem Phys 143:094102

Agarwal A, Delle Site L (2016) Grand-canonical adaptive resolution centroid molecular dynamics: implementation and application. Comput Phys Commun 206:26

Agarwal A, Wang H, Schütte C, Delle Site L (2014) Chemical potential of liquids and mixtures via adaptive resolution simulation. J Chem Phys 141:034102

Agarwal A, Zhu J, Hartmann C, Wang H, Delle Site L (2015) Molecular dynamics in a grand ensemble: Bergmann-Lebowitz model and adaptive resolution simulation. New J Phys 17:083042

Alekseeva U, Winkler RG, Sutmann G (2016) Hydrodynamics in adaptive resolution particle simulations: multiparticle collision dynamics. J Comput Phys 314:14–34

Allen MP, Tildesley DJ (1989) Computer simulation of liquids. Clarendon Press, Oxford

Altenhoff AM, Walther JH, Koumoutsakos P (2007) A stochastic boundary forcing for dissipative particle dynamics. J Comp Phys 225:1125–1136

Bevc S, Junghans C, Kremer K, Praprotnik M (2013) Adaptive resolution simulation of salt solutions. New J Phys 15:105007

Bian X, Li Z, Deng M, Karniadakis GE (2015a) Fluctuating hydrodynamics in periodic domains and heterogeneous adjacent multidomains: thermal equilibrium. Phys Rev E 92:053302

Bian X, Li Z, Karniadakis GE (2015b) Multi-resolution flow simulations by smoothed particle hydrodynamics via domain decomposition. J Comput Phys 297:132–155

Bian X, Deng M, Tang YH, Karniadakis GE (2016) Analysis of hydrodynamic fluctuations in heterogeneous adjacent multidomains in shear flow. Phys Rev E 93:033312

Bian X, Deng M, Karniadakis GE (2018) Analytical and computational studies of correlations of hydrodynamic fluctuations in shear flow. Commun Comput Phys 23:93–117

Borg MK, Lockerby DA, Reese JM (2013) A multiscale method for micro/nano flows of high aspect ratio. J Comput Phys 233:400–413

Borg MK, Lockerby DA, Reese JM (2014) The fade mass-stat: a technique for inserting or deleting particles in molecular dynamics simulations. J Chem Phys 140(7):074110

Chang H, Li X, Li H, Karniadakis GE (2016) Md/dpd multiscale framework for predicting morphology and stresses of red blood cells in health and disease. PLOS Comput Bio 12:e1005173

De Fabritiis G, Delgado-Buscalioni R, Coveney PV (2004) Energy controlled insertion of polar molecules in dense fluids. J Chem Phys 121(24):12139–12142

De Fabritiis G, Delgado-Buscalioni R, Coveney PV (2006) Multiscale modeling of liquids with molecular specificity. Phys Rev Lett 97:134501

Delgado-Buscalioni R (2012) Tools for multiscale simulation of liquids using open molecular dynamics. In: Engquist B, Runborg O, Tsai YHR (eds) Numerical analysis of multiscale computations, vol 82. Springer, Berlin/Heidelberg, pp 145–166

Delgado-Buscalioni R, Coveney PV (2003a) Continuum-particle hybrid coupling for mass, momentum, and energy transfers in unsteady fluid flow. Phys Rev E 67:046704

Delgado-Buscalioni R, Coveney PV (2003b) USHER: an algorithm for particle insertion in dense fluids. J Chem Phys 119:978–987

Delgado-Buscalioni R, Coveney P (2004) Hybrid molecularcontinuum fluid dynamics. Phil Trans R Soc A 362(1821):1639–1654

Delgado-Buscalioni R, De Fabritiis G (2007) Embedding molecular dynamics within fluctuating hydrodynamics in multiscale simulations of liquids. Phys Rev E 76:036709

Delgado-Buscalioni R, Flekkøy EG, Coveney PV (2005) Fluctuations and continuity in particle-continuum hybrid simulations of unsteady flows based on flux-exchange. EPL 69(6):959

Delgado-Buscalioni R, Kremer K, Praprotnik M (2008) Concurrent triple-scale simulation of molecular liquids. J Chem Phys 128:114110

Delgado-Buscalioni R, Kremer K, Praprotnik M (2009) Coupling atomistic and continuum hydrodynamics through a mesoscopic model: application to liquid water. J Chem Phys 131:244107

Delgado-Buscalioni R, Sablić J, Praprotnik M (2015) Open boundary molecular dynamics. Eur Phys J Special Top 224:2331–2349

Delle Site L (2016) Formulation of Liouville's theorem for grand ensemble molecular simulations. Phys Rev E 93:022130

Delle Site L (2018) Grand canonical adaptive resolution simulation for molecules with electrons: a theoretical framework based on physical consistency. Comput Phys Commun 222:94–101

Delle Site L, Praprotnik M (2017) Molecular systems with open boundaries: theory and simulation. Phys Rep 693:1–56

Deng M, Li X, Liang H, Caswell B, Karniadakis GE (2012) Simulation and modeling of slip flow over surfaces grafted with polymer brushes and glycocalyx fibres. J Fluid Mech 711:192–211

Español P (1995) Hydrodynamics from dissipative particle dynamics. Phys Rev E 52(2):1734–1742

Español P, Revenga M (2003) Smoothed dissipative particle dynamics. Phys Rev E 67(2):026705

Español P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. Europhys Lett 30(4):191–196

Español P, Warren PB (2017) Perspective: dissipative particle dynamics. J Chem Phys 146(15):150901

Español P, Delgado-Buscalioni R, Everaers R, Potestio R, Donadio D, Kremer K (2015) Statistical mechanics of hamiltonian adaptive resolution simulations. J Chem Phys 142:064115

Everaers, R (2016) Thermodynamic translational invariance in concurrent multiscale simulations of liquids. Eur Phys J Special Top 225:1483–1503

Fedosov DA, Karniadakis GE (2009) Triple-decker: interfacing atomistic-mesoscopic-continuum flow regimes. J Comput Phys 228:1157–1171

Fedosov DA, Caswell B, Karniadakis GE (2010) A multiscale red blood cell model with accurate mechanics, rheology, and dynamics. Biophys J 98:2215–2225

Fedosov DA, Pan W, Caswell B, Gompper G, Karniadakis GE (2011) Predicting human blood viscosity in silico. Proc Natl Acad Sci USA 108:11772–11777

Flekkøy EG, Wagner G, Feder J (2000) Hybrid model for combined particle and continuum dynamics. Europhys Lett 52(3):271–276

Flekkøy EG, Delgado-Buscalioni R, Coveney PV (2005) Flux boundary conditions in particle simulations. Phys Rev E 72:026703

Fogarty AC, Potestio R, Kremer K (2015) Adaptive resolution simulation of a biomolecule and its hydration shell: structural and dynamical properties. J Chem Phys 142:195101

Fogelson AL, Neeves KB (2015) Fluid mechanics of blood clot formation. Ann Rev Fluid Mech 47(1):377–403

Garcia AL, Bell JB, Crutchfield WY, Alder BJ (1999) Adaptive mesh and algorithm refinement using direct simulation Monte Carlo. J Comput Phys 154(1):134–155

Grinberg L (2012) Proper orthogonal decomposition of atomistic flow simulations. J Comput Phys 231(16):5542–5556

Grinberg L, Fedosov DA, Karniadakis GE (2013) Parallel multiscale simulations of a brain aneurysm. J Comput Phys 244:131–147

Groot RD, Warren PB (1997) Dissipative particle dynamics: bridging the gap between atomistic and mesoscopic simulation. J Chem Phys 107(11):4423–4435

Hadjiconstantinou NG (1999) Hybrid atomisticcontinuum formulations and the moving contact-line problem. J Comput Phys 154(2):245–265

Hadjiconstantinou NG, Patera AT (1997) Heterogeneous atomistic-continuum representations for dense fluid systems. Int J Mod Phys C 08(04):967–976

Halverson JD, Brandes T, Lenz O, Arnold A, Bevc S, Starchenko V, Kremer K, Stuehn T, Reith D (2013) Espresso++: a modern multiscale simulation package for soft matter systems. Comput Phys Commun 184(4):1129–1149

Hoogerbrugge PJ, Koelman JMVA (1992) Simulating microsopic hydrodynamics phenomena with dissipative particle dynamics. Europhys Lett 19(3):155–160

Kevrekidis IG, Samaey G (2009) Equation-free multiscale computation: algorithms and applications. Annu Rev Phys Chem 60(1):321–344

Koumoutsakos P (2005) Multiscale flow simulations using particles. Ann Rev Fluid Mech 37:457–487

Kreis K, Fogarty A, Kremer K, Potestio R (2015) Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations. Eur Phys J Special Top 224:2289–2304

Landau LD, Lifshitz EM (1987) Fluid mechanics. Course of theoretical physics, vol 6, 2nd edn. Pergamon Press, Oxford

Laso M, Öttinger HC (1993) Calculation of viscoelastic flow using molecular models: the CONNFFESSIT approach. J Non-Newton Fluid Mech 47:1–20

Lei H, Fedosov DA, Karniadakis GE (2011) Time-dependent and outflow boundary conditions for dissipative particle dynamics. J Comput Phys 230:3765–3779

Li J, Liao D, Yip S (1998) Coupling continuum to molecular-dynamics simulation: reflecting particle method and the field estimator. Phys Rev E 57:7259–7267

Malevanets A, Kapral R (1999) Mesoscopic model for solvent dynamics. J Chem Phys 110(17):8605–8613

Miller RE, Tadmor EB (2009) A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. Model Simul Mater Sci Eng 17(5):053001

Mohamed K, Mohamad A (2010) A review of the development of hybrid atomistic-continuum methods for dense fluids. Microfluid Nanofluid 8(3):283–302

Mukherji D, Kremer K (2013) Coil-globule-coil transition of PNIPAm in aqueous methanol: coupling all-atom simulations to semi-grand canonical coarse-grained reservoir. Macromolecules 46(22):9158–9163

Nagarajan A, Junghans C, Matysiak S (2013) Multiscale simulation of liquid water using a four-to-one mapping for coarse-graining. J Chem Theory Comput 9:5168–5175

Neumann P, Bian X (2017) MaMiCo: transient multi-instance molecular-continuum flow simulation on super-computers. Comput Phys Commun 220:390–402

Nie XB, Chen SY, Weinan E, Robbins MO (2004) A continuum and molecular dynamics hybrid method for micro- and nano-fluid flow. J Fluid Mech 500:55–64

Nie X, Robbins MO, Chen S (2006) Resolving singular forces in cavity flow: multiscale modeling from atomic to millimeter scales. Phys Rev Lett 96:134501

O'Connell ST, Thompson PA (1995) Molecular dynamics-continuum hybrid computations: a tool for studying complex fluid flows. Phys Rev E 52:R5792–R5795

Petsev ND, Leal LG, Shell MS (2015) Hybrid molecular-continuum simulations using smoothed dissipative particle dynamics. J Chem Phys 142(4):044101

Petsev ND, Leal LG, Shell MS (2017) Coupling discrete and continuum concentration particle models for multiscale and hybrid molecular-continuum simulations. J Chem Phys 147:234112

Pivkin IV, Karniadakis GE (2006) Controlling density fluctuations in wall-bounded dissipative particle dynamics systems. Phys Rev Lett 96:206001

Poma A, Delle Site L (2010) Classical to path-integral adaptive resolution in molecular simulation: towards a smooth quantum-classical coupling. Phys Rev Lett 104:250201

Poma A, Delle Site L (2011) Adaptive resolution simulation of liquid para-hydrogen: testing the robustness of the quantum-classical adaptive coupling. Phys Chem Chem Phys 13:10510

Potestio R, Español P, Delgado-Buscalioni R, Everaers R, Kremer K, Donadio D (2013a) Monte Carlo adaptive resolution simulation of multicomponent molecular liquids. Phys Rev Lett 111:060601

Potestio R, Fritsch S, Español P, Delgado-Buscalioni R, Kremer K, Everaers R, Donadio D (2013b) Hamiltonian adaptive resolution simulation for molecular liquids. Phys Rev Lett 110:108301

Praprotnik M, Delle Site L, Kremer K (2005) Adaptive resolution molecular-dynamics simulation: changing the degrees of freedom on the fly. J Chem Phys 123(22):224106

Praprotnik M, Delle Site L, Kremer K (2008) Multiscale simulation of soft matter: from scale bridging to adaptive resolution. Annu Rev Phys Chem 59:545–571

Praprotnik M, Poblete S, Kremer K (2011) Statistical physics problems in adaptive resolution computer simulations of complex fluids. J Stat Phys 145:946–966

Quarteroni A, Valli A (1999) Domain decomposition methods for partial differential equations. Oxford science publications, Oxford

Ren W (2007) Analytical and numerical study of coupled atomistic-continuum methods for fluids. J Comput Phys 227(2):1353–1371

Ren W, Weinan EW (2005) Heterogeneous multiscale method for the modeling of complex fluids and micro-fluidics. J Comput Phys 204(1):1–26

Rossinelli D, Tang YH, Lykov K, Alexeev D, Bernaschi M, Hadjidoukas P, Bisson M, Joubert W, Conti C, Karniadakis G, Fatica M, Pivkin I, Koumoutsakos P (2015) The in-silico lab-on-a-chip: petascale and high-throughput simulations of microfluidics at cell resolution. In: Proceedings of the international conference for high performance computing, networking, storage and analysis, SC'15, vol 2, ACM, New York, pp 1–12

Sablić J, Praprotnik M, Delgado-Buscalioni R (2016) Open boundary molecular dynamics of sheared star-polymer melts. Soft Matter 12:2416–2439

Sablić J, Delgado-Buscalioni R, Praprotnik M (2017a) Application of the eckart frame to soft matter: rotation of star polymers under shear flow. Soft Matter 13:6988–7000

Sablić J, Praprotnik M, Delgado-Buscalioni R (2017b) Deciphering the dynamics of star molecules in shear flow. Soft Matter 13:4971–4987

Scukins A, Nerukh D, Pavlov E, Karabasov S, Markesteijn A (2015) Multiscale molecular dynamics/hydrodynamics implementation of two dimensional "mercedes benz" water model. Euro Phys J Special Top 224(12):2217–2238

Smith B, Bjørstad P, Gropp W (1996) Domain decomposition: parallel multilevel methods for elliptic partial differential equations. Cambridge University Press, New York

Tang YH, Kudo S, Bian X, Li Z, Karniadakis GE (2015) Multiscale universal interface: a concurrent framework for coupling heterogeneous solvers. J Comput Phys 297:13–31

Thompson PA, Robbins MO (1990) Shear flow near solids: epitaxial order and flow boundary conditions. Phys Rev A 41:6830–6837

Toselli A, Widlund OB (2005) Domain decomposition methods–algorithms and theory. Springer, Berlin/Heidelberg

Tuckerman ME (2010) Statistical mechanics: theory and molecular simulation. Oxford University Press, Oxford

Vázquez-Quesada A, Ellero M, Español P (2009) Consistent scaling of thermal fluctuations in smoothed dissipative particle dynamics. J Chem Phys 130(3):034901

Walther JH, Praprotnik M, Kotsalis EM, Koumoutsakos P (2012) Multiscale simulation of water flow past a C540 fullerene. J Comput Phys 231(7):2677–2681

Wang H, Hartmann C, Schütte C, Delle Site L (2013) Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique. Phys Rev X 3:011018

Warshel A, Karplus M (1972) Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. J Am Chem Soc 94(16):5612–5625

Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mole Bio 103(2):227–249

Weinan EW, Engquist B, Li X, Ren W, Vanden-Eijnden E (2007) Heterogeneous multiscale method: a review. Commun Comput Phys 2(3):367–450

Weinbaum S, Tarbell JM, Damiano ER (2007) The structure and function of the endothelial glycocalyx layer. Ann Rev Biomed Eng 9(1):121–167

Werder T, Walther JH, Koumoutsakos P (2005) Hybrid atomistic-continuum method for the simulation of dense fluid flows. J Comput Phys 205:373–390

Wijesinghe HS, Hadjiconstantinou NG (2004) Discussion of hybrid atomistic-continuum methods for multiscale hydrodynamics. Inter J Multi Comput Eng 2(2):189–202

Yasuda S, Yamamoto R (2010) Multiscale modeling and simulation for polymer melt flows between parallel plates. Phys Rev E 81:036308

Zavadlav J, Praprotnik M (2017) Adaptive resolution simulations coupling atomistic water to dissipative particle dynamics. J Chem Phys 147:114110

Zavadlav J, Melo MN, Marrink SJ, Praprotnik M (2014) Adaptive resolution simulation of an atomistic protein in MARTINI water. J Chem Phys 140:054114

Zavadlav J, Melo MN, Marrink SJ, Praprotnik M (2015a) Adaptive resolution simulation of polarizable supramolecular coarse-grained water models. J Chem Phys 142:244118

Zavadlav J, Podgornik R, Praprotnik M (2015b) Adaptive resolution simulation of a DNA molecule in salt solution. J Chem Theory Comput 11:5035–5044

Zavadlav J, Marrink SJ, Praprotnik M (2016a) Adaptive resolution simulation of supramolecular water: the concurrent making, breaking, and remaking of water bundles. J Chem Theory Comput 12:4138–4145

Zavadlav J, Podgornik R, Melo MN, Marrink SJ, Praprotnik M (2016b) Adaptive resolution simulation of an atomistic DNA molecule in MARTINI salt solution. Eur Phys J Spec Top 225:1595–1607

Zavadlav J, Bevc S, Praprotnik M (2017) Adaptive resolution simulations of biomolecular systems. Eur Biophys J 46:821–835

Particle-Based Methods for Mesoscopic Transport Processes

106

Zhen Li, Wenxiao Pan, and Alexandre M. Tartakovsky

Contents

Abstract

Biological phenomena at cellular and subcellular levels involve transport of reacting biochemical species, including proteins, enzymes, and nutrients. At the mesoscopic scale, stochastic effects can play an important role and dominate the biological processes. Therefore, the continuum deterministic description, which ignores fluctuations, becomes no longer accurate. To this end, mesoscopic methods with stochastic terms are attracting more attention as a promising approach for tackling challenging problems in cellular biology and bioengineering. This chapter describes particle-based stochastic models for mesoscopic transport processes, including thermal transport, reactive chemical transport, and ionic transport in mesoscopic systems.

Z. Li (✉)
Department of Mechanical Engineering, Clemson University, Clemson, SC, USA
e-mail: zli7@clemson.edu

W. Pan
Department of Mechanical Engineering, University of Wisconsin-Madison, Madison, WI, USA
e-mail: wpan9@wisc.edu

A. M. Tartakovsky
Computational Mathematics Group, Pacific Northwest National Laboratory, Richland, WA, USA
e-mail: alexandre.tartakovsky@pnnl.gov

1 Introduction

At the mesoscopic scale, thermal fluctuations and discrete features play important roles in biological processes, and the classical approaches based on deterministic transport equations are inadequate to predict the time evolution of the mesoscale biological system (Hellander 2008). Deterministic equations are formulated as a system of (generally, nonlinear) ordinary differential equations (ODEs) and partial differential equations (PDEs) for the dynamics of mean-field concentrations of the biochemical species. However, at cellular and subcellular levels, the biological processes present considerable variations from cell to cell (Chen et al. 2017). The variations arise from the fact that the cells are so small that the discrete features and stochastic effects become crucial and may play a significant role in the relevant dynamics. Therefore, to accurately model biological processes at the mesoscopic scale, a better approach involves developing discrete particle-based methods suitable for stochastic processes, where the discreteness of the biochemical species and the randomness of transports and reactions are explicitly accounted for. From a modeling point of view, stochastic dynamics is considerably harder to simulate and analyze than the mean-field behavior (Schnoerr et al. 2017).

For mesoscopic processes driven by thermal noises, a popular approach that accounts for the effects of thermal fluctuations is the Markovian stochastic process, where stochastic terms induced by equilibrium or non-equilibrium fluctuations can be modeled in terms of Gaussian white noise (Ortiz de Zárate and Sengers 2006). A Markovian approximation is valid when the correlation time of fluctuations is much shorter than the characteristic timescale of the dynamics of fluid particles/voxels. This is true as long as the size of fluid particles/voxels is much larger than the biochemical molecule size. As the size of fluid particle/voxel becomes comparable with the molecule size, it is anticipated that the macroscopic or mesoscopic modeling framework based on equilibrium/non-equilibrium statistics will fail to describe the underlying system dynamics accurately. In this regard, microscale methods with much higher computational cost, such as the reaction-diffusion master equation (RDME) (Drawert et al. 2016; Schnoerr et al. 2017), must be employed to attain insight into the discrete features of biochemical reactive transport.

This chapter focuses on recent advances in particle-based methods, including dissipative particle dynamics (DPD), smoothed particle hydrodynamics (SPH), and smoothed dissipative particle dynamics (SDPD), for mesoscale transport processes. In these methods, the stochastic terms induced by thermal fluctuations are modeled as Markov processes. Developing these methods has provided not only a deeper understanding of the transport processes in the presence of fluctuations but also a framework for modeling non-equilibrium processes of mesoscopic systems using a local equilibrium approximation and investigating the energetic fluctuations of mesoscopic transport in biological systems and beyond. Methods for thermal transport at the mesoscopic scale are introduced in Sect. 2. The methods for modeling mesoscopic reactive transport are discussed in Sect. 3. Section 4 presents the methods for modeling mesoscopic ionic transport, and Sect. 5 concludes with a brief summary and discussion.

2   Mesoscopic Thermal Transport

Compared to healthy tissues, unhealthy tissues (especially tumors and cancers) usually have lower pH values and higher temperatures stemming from microenvironment abnormalities (Taghizadeh et al. 2015). The differences between unhealthy and healthy tissues can be used to design targeted therapies. For example, based on the fact that tumor microenvironments are slightly hyperthermic, thermosensitive polymers and liposomes can be used as targeted drug carriers to effectively deliver drugs to tumor cells without harming healthy cells (Perez-Herrero and Fernandez-Medarde 2015). However, developing thermosensitive drug delivery techniques with controllable drug release in non-isothermal processes requires a fundamental understanding of the mesoscopic thermal transport and knowledge of the role thermal fluctuations play in biological processes. This section introduces two mesoscopic modeling methods that can provide such fundamental understanding and knowledge.

First, Avalos and Mackie (1997) and Español (1997) developed an energy-conserving DPD (eDPD) model that augments the classical DPD model with the internal energy as an additional property of each DPD particle. The classical DPD model was developed based on equilibrium thermodynamics (Español and Warren 1995), where the dissipative and random forces satisfying the fluctuation-dissipation theorem act as a thermostat to maintain the system at a constant temperature. Thus, the classical DPD method is limited to modeling isothermal systems and can neither sustain temperature gradients nor model thermal transport in non-isothermal processes (Li et al. 2017). In contrast, in the eDPD method, each eDPD particle is considered as a coarse-grained particle associated with an internal energy in addition to other quantities, such as position and momentum. The stochastic differential equations governing the dynamics of eDPD particles include the conservations of momentum and energy in the form of (Li et al. 2014):

$$
\begin{aligned}
m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} &= m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = \mathbf{F}_i = \sum_{i \neq j} (\mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R) \,, \\
C_v \frac{\mathrm{d}T_i}{\mathrm{d}t} &= q_i = \sum_{i \neq j} (q_{ij}^C + q_{ij}^V + q_{ij}^R) \,,
\end{aligned}
\tag{1}
$$

where m, t, $\mathbf{r}_i$, $\mathbf{v}_i$, and $\mathbf{F}_i$ denote mass, time, position, velocity, and force vectors, respectively. T_i represents the temperature, C_v the thermal capacity, and q_i the net heat flux of particle i. The summations of force and heat flux are carried out over all particles within a cutoff radius r_{cf} from particle i for forces and a cutoff radius r_{ct} for heat fluxes.

The three components of $\mathbf{F}_i$, including conservative, dissipative, and random forces, are given by $\mathbf{F}_{ij}^C = a_{ij} w_C(r_{ij}) \mathbf{e}_{ij}$, $\mathbf{F}_{ij}^D = -\gamma_{ij} w_D(r_{ij}) (\mathbf{e}_{ij} \cdot \mathbf{v}_{ij}) \mathbf{e}_{ij}$, and $\mathbf{F}_{ij}^R \mathrm{d}t = \sigma_{ij} w_R(r_{ij}) \mathrm{d}W_{ij} \mathbf{e}_{ij}$, respectively (Groot and Warren 1997). The coefficients a, γ, and σ determine the strength of each force. The weight functions $w_C(r)$,

$w_D(r)$, and $w_R(r)$ are defined with a cutoff radius r_{cf} beyond which these weight functions vanish. $dW_{ij} = dW_{ji}$ are independent increments of the Wiener process. Using the assumption of local thermodynamic equilibrium, Español (1997) derived a Fokker-Planck equation (FPE) for the stochastic equation (1). The FPE solution gives the relationship between the dissipative force and the random force, which requires $\sigma_{ij}^2 = 4\gamma_{ij}k_B T_i T_j/(T_i + T_j)$ and $w_D(r) = w_R^2(r)$. It indicates that the fluctuation-dissipation theorem is applied locally based on the particle temperature T_i instead of the system's thermodynamic temperature, stressing the local nature of the eDPD model (Avalos and Mackie 1997). The heat fluxes between two particles i and j, accounting for the thermal conduction q_{ij}^C, viscous heating q_{ij}^V, and random heat flux q_{ij}^R, are given by (Li et al. 2014)

$$q_{ij}^C = k_{ij} w_{CT}(r_{ij}) \left(T_i^{-1} - T_j^{-1} \right) , \tag{2a}$$

$$q_{ij}^V = \frac{1}{2C_v} \left\{ w_D(r_{ij}) \left[\gamma_{ij} \left(\mathbf{e}_{ij} \cdot \mathbf{v}_{ij} \right)^2 - m_i^{-1}\sigma_{ij}^2 \right] - \sigma_{ij} w_R(r_{ij}) \left(\mathbf{e}_{ij} \cdot \mathbf{v}_{ij} \right) \xi_{ij} \right\} , \tag{2b}$$

$$q_{ij}^R = \beta_{ij} w_{RT}(r_{ij}) \xi_{ij}^e , \tag{2c}$$

where the thermal conduction q_{ij}^C contains T^{-1} rather than T because the thermodynamic quantity conjugated to the internal energy is the inverse of the temperature instead of the temperature itself (Español 1997). The parameters k_{ij} and β_{ij} determine the strength of the thermal conduction and the random heat flux, respectively. In particular, k_{ij} plays the role of a thermal conductivity given as $k_{ij} = C_v^2 \kappa (T_i + T_j)^2/4k_B$, where κ is interpreted as a mesoscale heat friction coefficient. The relationships $\beta_{ij}^2 = 2k_B k_{ij}$ and $w_{CT}(r) = w_{RT}^2(r)$ are required to satisfy the fluctuation-dissipation theorem. A common choice of the weight functions is $w_{CT}(r) = w_{RT}^2(r) = (1 - r/r_{ct})^{s_T}$ with a cutoff radius r_{ct}, where the case of $s_T = 2.0$ corresponds to the typical quadratic weighting function (Groot and Warren 1997).

For an eDPD particle, the characteristic scale of the kinetic energy related to its momentum is $k_B T$, while the characteristic scale of the internal energy related to its temperature is $C_v T$. A scaling factor $k_B T/C_v T$ is required to convert kinetic energy into internal energy. Therefore, the heat flux q_{ij}^V due to viscous heating given by Eq. (2b) has a factor $1/C_v$ when the kinetic energy $k_B T$ is taken as the energy unit. Also, the factor 2 in the denominator guarantees that the heat generated by nonconservative interactions is distributed evenly to both particles of a pair. Through numerical experimentation, Li et al. (2014) demonstrated that Eqs. (1) and (2) conserve energy.

The transport properties, such as diffusivity and viscosity, do not need to be prescribed in eDPD. Groot and Warren (1997) and Marsh et al. (1997) have derived the expressions of diffusivity and kinematic viscosity in terms of DPD parameters, which are given as

$$D = \frac{3k_B T}{4\pi\gamma\rho} \left(\int_0^\infty r^2 g(r) w_D(r) dr \right)^{-1}, \quad \nu = \frac{D}{2} + \frac{2\pi\gamma\rho}{15} \int_0^\infty r^4 g(r) w_D(r) dr, \tag{3}$$

where $g(r)$ is the radial distribution function. A rough analytical prediction of D and ν can be obtained by substituting $w_D(r) = w_R^2(r) = (1 - r/r_{cf})^s$ and $g(r) = 1.0$ into Eq. (3), which leads to $D = 3k_B T \prod_{i=1}^3 (s + i)/8\pi\gamma\rho r_{cf}^3$ and $\nu = D/2 + 16\pi\gamma\rho r_{cf}^5/5 \prod_{i=1}^5 (s + i)$.

The DPD fluid's thermal conductivity also can be calculated in terms of the model parameters. When the energy transport is dominated by thermal conduction, the macroscopic thermal conductivity λ can be calculated as (Mackie et al. 1999)

$$\lambda = \frac{2\pi}{3} \frac{\rho^2}{T^2} \int_0^\infty r^4 k w_{CT}(r) g(r) dr , \tag{4}$$

where $k = C_v^2 \kappa T^2/k_B$ and $w_{CT}(r)$ is the weight function. Given $w_{CT}(r) = (1 - r/r_{ct})^{s_T}$ and $g(r) = 1.0$, an analytical estimate for the thermal conductivity λ can be obtained as $\lambda = 16\pi\rho^2\kappa C_v^2 r_{ct}^5/k_B \prod_{i=1}^5 (s_T + i)$. In particular, when the typical quadratic weight function $(1 - r/r_{ct})^2$ is adopted for $w_{CT}(r)$, $\lambda = 2\pi\rho^2\kappa C_v^2 r_{ct}^5/315 k_B$. Let $\mathrm{Pr} = \rho\nu C_v/\lambda$ be the Prandtl number, which is a dimensionless number whose value can be obtained from experiments. By replacing λ with Pr, the result is a formula for determining the mesoscale heat friction κ given by $\kappa = 315 k_B \nu/2\pi\rho \mathrm{Pr} C_v r_{ct}^5$. Notably, this formula is obtained from the typical quadratic weight function. When a different weight function is employed, a similar formula could be derived from Eq. (4).

The expression of Eq. (3) indicates that both the diffusivity D and kinematic viscosity ν of the eDPD fluid increase linearly with the temperature if other variables in Eq. (3) remain constant. However, for most of the liquids, including water, ethanol, and glycerin, the diffusivity increases, but the kinematic viscosity decreases with increasing temperature. To correct the temperature-dependent properties (diffusivity, viscosity, and thermal conductivity) of these liquids, Li et al. (2014) analyzed the sensitivity of D, ν, and λ with respect to the model parameters and defined the exponent of the weighting function s as a function of temperature. They took liquid water as an example and demonstrated that the eDPD model can produce the correct temperature-dependent diffusivity and viscosity, as well as Prandtl numbers, for various temperatures, which is consistent with available experimental data. For details on the eDPD model parameterization, refer to Li et al. (2014) and Lei et al. (2017).

Alternatively to the bottom-up eDPD model, a top-down particle-based model for mesoscopic thermal transport can be derived from macroscopic PDEs governing the conservations of continuity, momentum, and energy. By introducing thermal fluctuations into an SPH discretization of the (deterministic) Navier-Stokes equations, Español and Revenga (2003) developed the SDPD model. The following SPH discretization was used:

$$m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = \sum_j \left[\frac{P_i}{d_i^2} + \frac{P_j}{d_j^2} \right] F_{ij} \mathbf{r}_{ij} - \frac{5\eta}{3} \sum_j \frac{F_{ij}}{d_i d_j} \left[\mathbf{v}_{ij} + (\mathbf{v}_{ij} \cdot \mathbf{e}_{ij})\mathbf{e}_{ij} \right] ,$$

$$T_i \frac{\mathrm{d}S_i}{\mathrm{d}t} = \frac{5\eta}{6} \sum_j \frac{F_{ij}}{d_i d_j} \left[\mathbf{v}_{ij}^2 + (\mathbf{v}_{ij} \cdot \mathbf{e}_{ij})^2 \right] - 2\kappa \sum_j \frac{F_{ij}}{d_i d_j} (T_i - T_j) ,$$

$$(5)$$

where P and T are the pressure and temperature given by the local equilibrium assumption. The number density d of the fluid particles is computed as $d_i = \sum_j W(|\mathbf{r}_i - \mathbf{r}_j|)$, where $W(r)$ is a compact bell-shaped smoothing kernel (Monaghan 2005). The geometrical factor F_{ij} is given by $F_{ij} = -\nabla W(r_{ij})/r_{ij}$. The transport coefficients are the shear viscosity η and the thermal conductivity κ. The relationship between pressure and density should be specified via an equation of state. The popular choices in SDPD simulations are $p = c^2 \rho$ and $p = p_0(\rho/\rho_0)^\gamma + b$ (Vázquez-Quesada et al. 2009). The first term in the entropy equation is the viscous heating term that captures the physical mechanism by which the energy dissipated by viscous forces transforms into the internal energy. The last term is the heat conduction term induced by the temperature difference between particles.

Using the GENERIC (General Equation for Non-Equilibrium Reversible-Irreversible Coupling) formalism (Grmela and Öttinger 1997; Öttinger and Grmela 1997), thermal fluctuations are then systematically introduced to satisfy the first and second laws of thermodynamics. The velocity and entropy random terms are given by (Español and Revenga 2003)

$$m_i \mathrm{d}\tilde{\mathbf{v}}_i = \sum_j A_{ij} \mathrm{d}\hat{\mathbf{W}}_{ij} \cdot \mathbf{e}_{ij} ,$$

$$T_i \mathrm{d}\tilde{S}_i = -\frac{1}{2} \sum_j A_{ij} \mathrm{d}\hat{\mathbf{W}}_{ij} : \mathbf{e}_{ij} \mathbf{v}_{ij} + \sum_j C_{ij} \mathrm{d}W_{ij}^H ,$$

$$(6)$$

respectively, where $\mathrm{d}\hat{\mathbf{W}}_{ij} = (\mathrm{d}\mathbf{W}_{ij} + \mathrm{d}\mathbf{W}_{ij}^T)/2$ is the symmetric part of a matrix of independent increments of the Wiener process $\mathrm{d}\mathbf{W}_{ij}$. $\mathrm{d}W_{ij}^H$ is an independent increment of the Wiener process for each pair of particles, giving rise to the heat conduction terms. Using the fluctuation-dissipation theorem, the parameters A_{ij} and C_{ij} are related to viscosity η and thermal conductivity κ as

$$A_{ij} = \left[\frac{40}{3} \eta k_B \frac{T_i T_j}{T_i + T_j} \frac{F_{ij}}{d_i d_j} \right]^{1/2} , \quad C_{ij} = \left[4\kappa k_B T_i T_j \frac{F_{ij}}{d_i d_j} \right]^{1/2} . \quad (7)$$

Thus, the equations of an SDPD model are obtained by adding the stochastic terms in Eq. (6) to the SPH equations in Eq. (5). Kordilla et al. (2014) derived similar stochastic equations by directly discretizing the fluctuating hydrodynamics equations (Ortiz de Zárate and Sengers 2006) with the SPH method.

Although derived using different approaches, the "bottom-up" eDPD and "top-down" SDPD methods have many similarities. Both SDPD and eDPD momentum equations have a conservative force that comes from the gradient of potential energy, a dissipative force induced by viscous friction, and a stochastic term representing thermal fluctuations. SDPD and eDPD energy equations include a heat conduction term to reduce temperature differences between neighboring fluid particles, a viscous heating term to account for the internal energy generated by the viscous dissipation, and a stochastic term induced by thermal fluctuations. In each method, the stochastic contributions in velocities and energies directly relate to the dissipative terms via the fluctuation-dissipation theorem.

Because eDPD was derived from coarse-graining microscopic dynamics (Español et al. 2016), phenomenological expressions for the particle interactions in Eq. (1) can be easily derived to account for complex physics. In the last decade, the bottom-up eDPD approach has been successfully applied to diverse problems involving non-isothermal processes. For example, Cao et al. (2013) used eDPD to simulate natural convection in an eccentric annulus and investigated the effects of various factors on the streamlines and temperature distributions (shown in Fig. 1a). Tang et al. (2016) applied the eDPD method to modeling thermosensitive polymers and simulated the thermally induced self-assembly process of thermoresponsive micelles and vesicles (Fig. 1b). Other examples include heat-stiffening polymer nanocomposites (Cudjoe et al. 2017), thermophoretic microswimmers (Fedosov et al. 2015), natural convection in microchannels (Abu-Nada 2010), and thermoresponsive microgels (Li et al. 2015a).

Due to the similar form of the equations, it should be possible to include complex physics in SDPD using the eDPD coarse-graining procedure. To date, most SDPD applications are concerned with complex fluids and colloidal suspensions (Litvinov et al. 2008; Bian et al. 2012; Lei et al. 2016). The advantage of SDPD for such applications is that the properties of each fluid component, including density, viscosity, and speed of sound, are prescribed directly as model parameters and do not need to be estimated from other model parameters as with eDPD.

3   Mesoscopic Diffusive and Reactive Transport

Many biological processes depend on the concentrations of specific proteins, ions, or other biochemical factors (Anand et al. 2003). For example, proteins in an aqueous solution diffuse in a living cell due to Brownian motion and some collisions of appropriate proteins may lead to chemical reactions. Thus, diffusion and reaction are two fundamental transport processes in biological systems.

For modeling mesoscopic diffusive and reactive transport, Li et al. (2015b) developed a transport dissipative particle dynamics (tDPD) model by defining concentration carried by each DPD particle in addition to other quantities, such as position and momentum. The time evolution of the position and solute concentration of a tDPD particle i with unit mass $m_i \equiv 1$ is governed by the mass and momentum conservation laws, which can be described by the following set of equations:
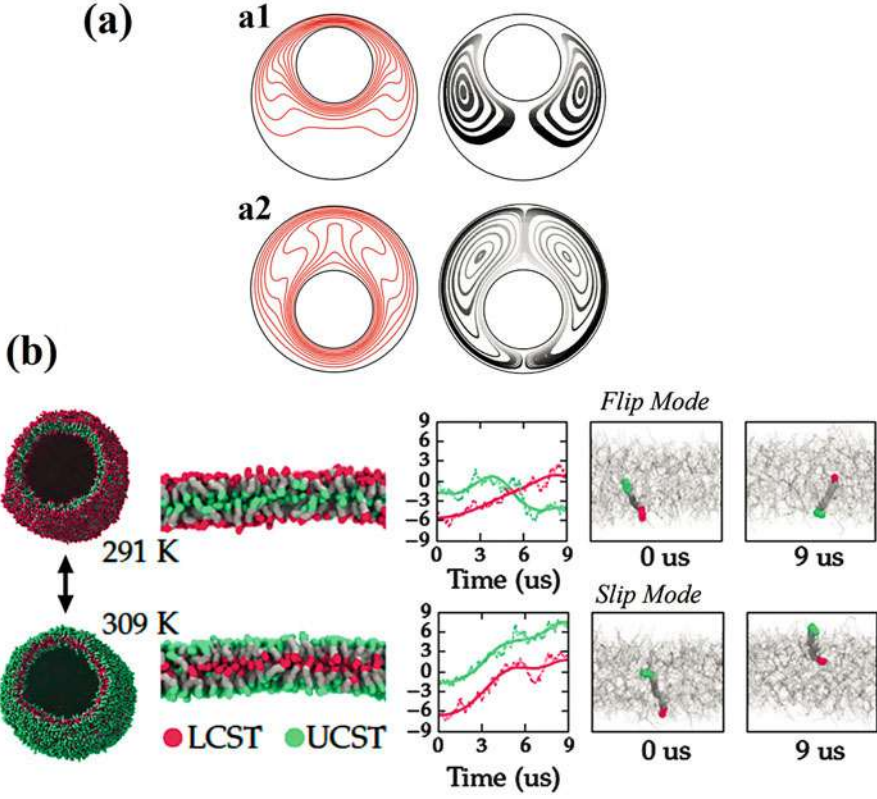
Fig. 1 (**a**) Thermal (left) and flow (right) fields of the natural convection in an eccentric annulus with two different eccentricities. (Adapted from Cao et al. 2013). (**b**) Thermally induced inversion of a unilamellar vesicle formed by triblock copolymers and a proper orthogonal decomposition (POD) analysis showing two dominant molecular movement modes, i.e., flip and slip, during membrane inversion. (Adapted from Tang et al. 2016)

$$\frac{\mathrm{d}^2\mathbf{r}_i}{\mathrm{d}t^2} = \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = \mathbf{F}_i = \sum_{i \neq j}(\mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R) + \mathbf{F}_i^{\mathrm{ext}},$$

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = Q_i = \sum_{i \neq j}(Q_{ij}^D + Q_{ij}^R) + Q_i^S,$$

(8)

where t, $\mathbf{r}_i$, $\mathbf{v}_i$, and $\mathbf{F}_i$ denote time and position, velocity, and force vectors, respectively. $\mathbf{F}_i^{\mathrm{ext}}$ is the force on particle i from an external force field. Just as in the classical DPD model, the pairwise interaction between tDPD particles i and j consists of the conservative force $\mathbf{F}_{ij}^C = a_{ij}w_C(r_{ij})\mathbf{e}_{ij}$, dissipative force $\mathbf{F}_{ij}^D = -\gamma_{ij}w_D(r_{ij})(\mathbf{e}_{ij} \cdot \mathbf{v}_{ij})\mathbf{e}_{ij}$, and random force $\mathbf{F}_{ij}^R\mathrm{d}t = \sigma_{ij}w_R(r_{ij})\mathrm{d}W_{ij}\mathbf{e}_{ij}$. C_i represents the concentration of one species defined as the number of a chemical

species carried by a tDPD particle i and Q_i, the corresponding concentration flux. The net concentration flux on a particle i accounts for the Fickian flux Q_{ij}^D, the random flux Q_{ij}^R, and a source term Q_{ij}^S generated by local chemical reactions. Because tDPD particles have unit mass, this definition of concentration is equivalent to the concentration in terms of chemical species per unit mass. Then, the volume concentration, i.e., chemical species per unit volume, is ρC_i, where ρ is the number density of tDPD particles. Of note, C_i can be a vector $\mathbf{C}_i$ containing N components, i.e., $\{C_1, C_2, \ldots, C_N\}_i$, when N chemical species are considered.

The driving force for diffusion of each species is the gradient of chemical potential $Q = -DC\nabla\mu/RT$, where D is the diffusion coefficient, R the universal gas constant, T the absolute temperature, and μ the chemical potential given by $\mu = \mu_0 + RT\ln C$ for dilute solutions (Ortiz de Zárate and Sengers 2006). By substituting the chemical potential μ into the expression for Q, the diffusion driving force is found to be proportional to the concentration gradient ∇C, which corresponds to a concentration difference between two neighboring tDPD particles. It follows that the Fickian and random fluxes in the tDPD model are given by

$$Q_{ij}^D = -\kappa_{ij} w_{DC}(r_{ij})\left(C_i - C_j\right), \quad Q_{ij}^R \mathrm{d}t = \varepsilon_{ij} w_{RC}(r_{ij})\mathrm{d}W_{ij}^{cc}, \tag{9}$$

where κ_{ij} and ε_{ij} determine the strength of the Fickian and random fluxes. $\mathrm{d}W_{ij}^{cc} = \mathrm{d}W_{ji}^{cc}$ are independent increments of the Wiener process. $w_{DC}(r)$ and $w_{RC}(r)$ are weight functions with a cutoff radius r_{cc}. The Fickian friction parameter κ plays the analogous role for concentration differences between tDPD particles, as γ does for momentum. In general, the concentration friction κ is a $N \times N$ matrix when the interdiffusivities of N different chemical species are involved. However, considering N chemical species in dilute solution and neglecting the interdiffusivities of different species, the system is then reduced to a set of uncoupled diffusion equations with independent diffusivities between species. Hence, κ becomes a diagonal matrix (Balluffi et al. 2005).

By applying the local-equilibrium assumption to the tDPD system, the random term Q_{ij}^R is related to the dissipative term Q_{ij}^D by satisfying the fluctuation-dissipation theorem (Li et al. 2015b)

$$\varepsilon_{ij}^2 = m_s^2 \kappa_{ij} \rho (C_i + C_j), \quad w_{DC}(r_{ij}) = w_{RC}^2(r_{ij}), \tag{10}$$

where m_s is the mass of a single solute molecule, while C_i and C_j are the respective concentrations on particles i and j. For detailed derivations for obtaining Eq. (10), refer to Li et al. (2015b). In general, the mass of a single solute molecule m_s is much smaller than that of a tDPD particle m, which is often chosen as the mass unit. Consequently, the magnitude of ε is small for $m_s \ll m$, which indicates that the contribution of the random flux Q_{ij}^R to the total diffusion coefficient D is negligible unless m_s becomes comparable to m in nanoscale systems.

The macroscopic properties, including viscosity and diffusivity of a tDPD system, are output properties rather than input parameters. Due to the random movements of tDPD particles generated by stochastic forces, the effective dif-

fusion coefficient D consists of both the random diffusion D^{ξ} and the Fickian diffusion D^F. In general, the random contribution D^{ξ} is a combined result of the random movements of tDPD particles and random flux Q_{ij}^R in Eq. (9). However, the variance of random flux Q_{ij}^R has a small prefactor m_s^2 as given by Eq. (10). Thus, the contribution of the random flux Q_{ij}^R to D^{ξ} is negligible in practical applications. In the derivations, it is assumed that D^{ξ} is induced by the random movements of tDPD particles. In particular, for a tDPD system in thermal equilibrium, the diffusion coefficient D^{ξ} induced by the random movements of tDPD particles and the macroscopic diffusion coefficient D^F due to the Fickian flux can be calculated by Li et al. (2015b)

$$D^{\xi} = \frac{3k_B T}{4\pi\gamma\rho \cdot \int_0^{r_c} r^2 w_D(r)g(r)dr} \ , \quad D^F = \frac{2\pi\kappa\rho}{3} \int_0^{r_{cc}} r^4 w_{DC}(r)g(r)dr \ , \tag{11}$$

where r_c is the cutoff radius for forces and r_{cc} is the cutoff radius for concentration flux. Let $w_D(r) = (1-r/r_c)^{s_1}$ and $w_{DC}(r) = (1-r/r_{cc})^{s_2}$ be the weight functions of the dissipative force $\mathbf{F}_{ij}^D$ and the Fickian flux Q_{ij}^D, respectively. When the radial distribution function of ideal gas $g(r) = 1.0$ is employed, both D^{ξ} and D^F can be evaluated analytically as $D = D^{\xi} + D^F = 3k_B T \prod_{i=1}^{3}(s_1+i)/8\pi\gamma\rho r_c^3 + 16\pi\kappa\rho r_{cc}^5 / \prod_{i=1}^{5}(s_2+i)$, where s_1 and s_2 are the exponents of $w_D(r)$ and $w_{DC}(r)$.

Because the particle-based tDPD method satisfies the conservation of concentration automatically, it provides an economical way to solve advection-diffusion-reaction (ADR) equations with a large number of species. Xu et al. (2011) used a two-dimensional model to investigate biofilm growth in a narrow channel and demonstrated the effects of flow velocity, growth parameter, and hydrodynamic interaction on the biofilm growth regime and morphology. Li et al. (2015b) employed tDPD to study the dynamic process of blood coagulation, modeled by a set of 23 coupled ADR equations for the evolution of 25 biological reactants involved in a combined model of intrinsic and extrinsic pathways of blood coagulation process and fibrinolysis. They reported the tDPD simulation correctly produced the thrombin burst followed by a drop and provided qualitatively correct evolution of fibrin concentration initialized by an injured vessel wall in flowing blood (shown in Fig. 2a).

Alternatively, mesoscopic mass transport can be modeled by a top-down approach: numerically solving the fluctuating hydrodynamics equations, including the continuity equation $d\rho/dt = -\rho(\nabla \cdot \mathbf{v})$ and momentum conservation equation $\rho d\mathbf{v}/dt = -\nabla P + \nabla \cdot \tau + \nabla \cdot \mathbf{s}$ with a random stress tensor $\mathbf{s}$ and the stochastic advection-diffusion equation $\rho dC/dt = \nabla\cdot(D\nabla C)+\nabla\cdot\mathbf{J}$ with a random flux vector $\mathbf{J}$ (Ortiz de Zárate and Sengers 2006). Using the fluctuation-dissipation theorem, the random stress in the momentum equation is related to the viscous stress, while the random concentration flux in the advection-diffusion equation relates to the Fickian diffusion. Kordilla et al. (2014) used SPH discretization to solve these equations. Similar to the tDPD model, the effective diffusion coefficient D in the SPH-based model (Kordilla et al. 2014) consists of a deterministic Fickian coefficient D^F and stochastic contribution D^{ξ}.
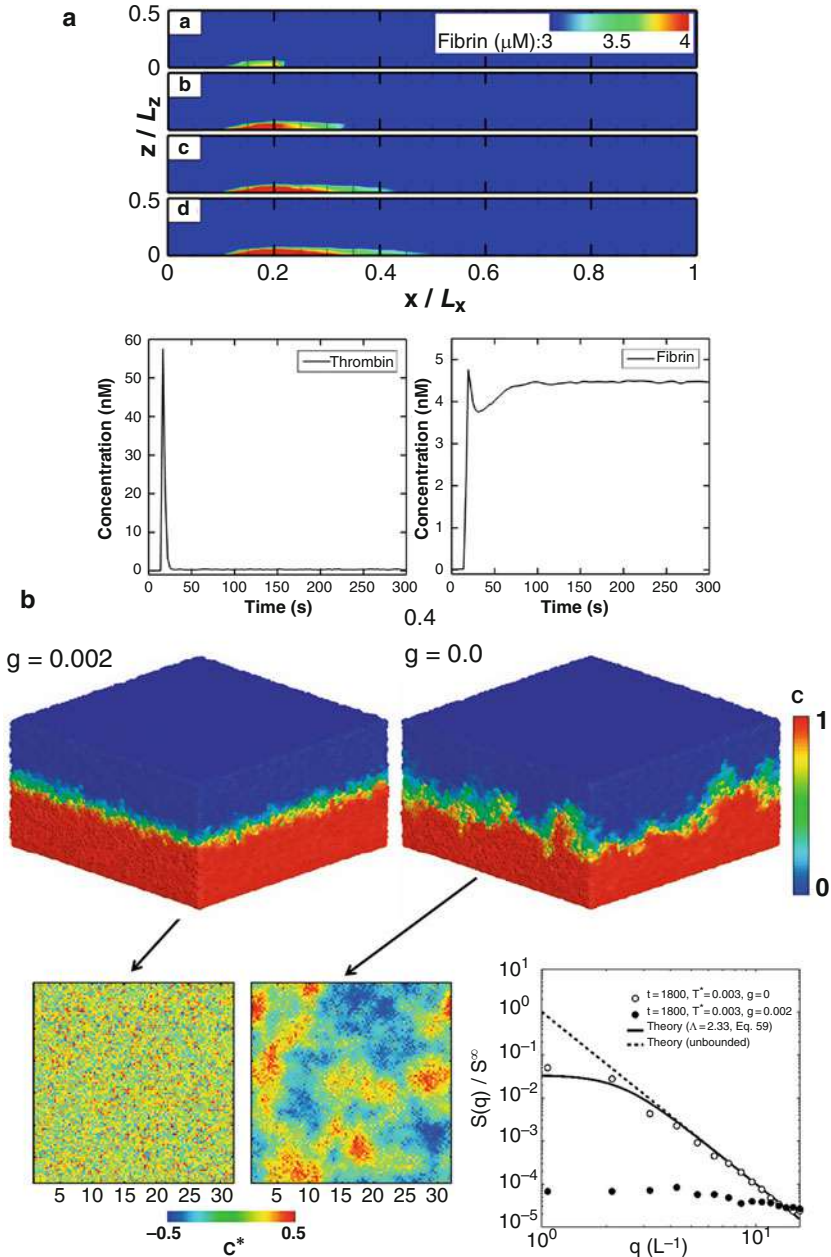
Fig. 2 (**a**) Time evolution of the fibrin concentration field during the dynamic process of blood coagulation in flowing blood and the concentrations of thrombin and fibrin at the center of the injured wall region. (Adapted from Li et al. 2015b). (**b**) Formation of "giant fluctuations" of the front between light and heavy fluids with and without gravity, where the light fluid is on the top of the heavy fluid. The map shows the concentration field, and the plot shows corresponding power spectra of the concentration field. (Adapted from Kordilla et al. 2014)

Kordilla et al. ([2014](#)) applied this SPH-based model for studying the formation of "giant fluctuations" of the front between light and heavy fluids with and without gravity, where the light fluid is on top of the heavy fluid (shown in Fig. [2b](#)). Their results indicate that the presence of gravity significantly reduces front perturbations for all considered values of Fickian diffusion D_F, but the effect of gravity becomes less pronounced as D^F increases. When the Fickian diffusion D_F becomes significantly larger than D_ξ, the stochastic diffusion reduces to a deterministic diffusion, and fluctuations completely disappear. Kordilla et al. ([2014](#)) also demonstrated that in the absence of gravity, the SPH-based model recovers the characteristic q^{-4} divergence of the interface power spectrum and its scale-invariant nature, which are in good agreement with the experiments and analytical solutions.

4   Mesoscopic Ionic Transport

Ionic transport is ubiquitous in biological processes and is extremely important in the vital activity of all organisms. The concentration gradient of potassium ion K^+, sodium ion Na^+, and calcium ion Ca^{2+} across the cell membrane is the basis for the transmission of excitation in organisms. It is not surprising that ionic transport is implicated in numerous diseases. For example, cystic fibrosis is an autosomal recessive disease, which is induced by defective ion transport and characterized by hyperabsorption of Na^+ in the airway epithelia (Kunzelmann and Mall [2003](#)). Transport of Ca^{2+} from muscle cells leads to relaxation of muscles, while the entry of these ions into the cytoplasm upon excitation produces muscle contraction (Rüegg [1992](#)). An inherited skeletal muscle disorder, named Brody disease, is clinically characterized by exercise-induced muscle stiffness, rooted from abnormal Ca^{2+} transport (Guglielmi et al. [2013](#)). Also, the flux of ions across the membranes of neurons changes the electrochemical gradient and results in the production of an electrical signal sent between neurons in the brain (Purves et al. [2004](#)). Disruption of ionic homeostasis significantly affects the neuron activities. Alzheimer's disease is associated with increased intracellular Na^+ and K^+ levels in brain regions, which is induced by imbalanced ionic transport contributed to the pathophysiology of Alzheimer's disease (Vitvitsky et al. [2012](#)). Understanding the functional role and mechanism of ionic transport at cellular/subcellular levels is of crucial physiological importance to develop new diagnostic tests for diseases and to discover novel drug molecules (Modi et al. [2012](#)).

However, simulating fluctuating electrohydrodynamic phenomena with fully coupled hydrodynamics and electrostatics with long-range Coulomb interactions is challenging. Here, a variant of DPD method is used to tackle this challenge. It extends the classical DPD method that has proven accurate and effective in modeling mesoscopic fluctuating hydrodynamics (Groot and Warren [1997](#)) and introduces ions represented by explicit charged particles in a DPD system. The electrostatic interactions between these ions can be computed by two approaches: (1) all interactions between the charged particles are summed in real space, and (2)

all charged particles are projected onto a local electrostatic field that is governed by the Poisson's equation $\nabla(\varepsilon \nabla \varphi) = -\beta e^2 \rho$ and solved separately.

As a consequence of the soft interactions in the DPD model, charged particles may overlap and form infinitely strongly bound ion pairs through electrostatic interactions. To avoid overlapping of charged DPD particles, Groot (2003) distributed the charge within an electrostatic smearing radius R_e using a charge distribution $f(r) = 3(1 - r/R_e)/\pi R_e^3$, while González-Melchor et al. (2006) used a Slater-type charge distribution $f(r) = \exp(-2r/\lambda)/\pi \lambda^3$ with a decay length λ. With explicit ions, the Ewald summation method (Ewald 1921) is a widely used route to calculate electrostatic interactions in particle systems. Because point charges interact according to Coulomb's law, the total electrostatic energy for a periodic cubic system of side L is given by

$$U(\mathbf{r}) = \frac{1}{4\pi \varepsilon_0 \varepsilon_r} \sum_i \sum_{j>i} \sum_{\mathbf{r}} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|}, \tag{12}$$

where $\mathbf{n} = (n_x, n_y, n_z)$ with n_x, n_y, and n_z being integer numbers. The sum over $\mathbf{n}$ takes into account the periodic images. The Ewald treatment decomposes the long-range electrostatic energy in a real space and the reciprocal space contributions, so both the real and reciprocal parts can be computed by short-ranged sums as an approach to $1/r$, capturing the full long-range nature of electrostatic interactions (González-Melchor et al. 2006).

Different from a direct calculation of the electrostatic forces, Groot (2003) introduced a lattice to the DPD system and spread out the charges over the lattice nodes. Then, the long-range portion of the interaction potential was calculated by solving the Poisson equation on the grid based on a particle-particle particle-mesh (PPPM) algorithm by transferring quantities (charges and forces) from the particles to the mesh and vice versa. It has been noted that the PPPM method works efficiently in DPD if the grid size equals the particle size. Because the mesh defines a coarse-graining length for electrostatic interactions, correlation effects on length scales shorter than the mesh size cannot be properly accounted for. Explicit treatment of the ions is computationally expensive, especially for systems with high ionic concentrations. To this end, Medina et al. (2015) developed a Condiff-DPD model with ions represented by the concentration of the ionic species. They used DPD equations for fluid particles and Brownian pseudo particles to describe the evolution of the ionic concentrations. Instead of direct interactions, the pseudo-ions and DPD particles are coupled through a lattice, where the Poisson equation is solved using the PPPM algorithm. Compared to "explicit-ion" simulations whose computational costs depend on the number of ions, the cost for electrostatic calculations in a simulation with "implicit ions" is dominated by the cost for solving the Poisson equation. Using the PPPM method, the cost primarily depends on the number of mesh points.

Although, the particle-to-mesh and then mesh-to-particle mapping/redistribution can solve the Poisson equation for particle-based systems (Groot 2003), its depen-

dence on a grid may contradict the original motivation for using a Lagrangian method, and additional computational complexity and inefficiencies are introduced. To abandon grids and use a unifying Lagrangian description for mesoscopic electrokinetic phenomena, Deng et al. (2016) developed a charged DPD, or cDPD, model, where the Poisson equation is solved on moving cDPD particles rather than grids. Specifically, cDPD describes the solvent explicitly in a coarse-graining sense as DPD particles, while the ion species are described semi-implicitly, i.e., using a Lagrangian description of ionic concentration fields, associated with each moving cDPD particle, which provides a natural coupling between fluctuating electrostatics and hydrodynamics.

The state vector of a cDPD particle can be written as $(\mathbf{r}, \mathbf{v}, c_\alpha, \phi)$, which is not only characterized by its position $\mathbf{r}$ and velocity $\mathbf{v}$ as in the classical DPD model but also by ionic species concentration c_α (with α representing the αth ion type) and electrostatic potential ϕ on the particle. A cDPD particle is then viewed as a coarse-grained fluid volume, which contains the solvent and other charged species. Exchange of the concentration flux of charged species occurs between neighboring cDPD particles, much like the momentum exchange in the classical DPD model. The time evolution of the state vector of a cDPD particle i with unit mass $m_i \equiv 1$ is governed by the following set of stochastic differential equations:

$$
\begin{aligned}
\frac{d^2\mathbf{r}_i}{dt^2} = \frac{d\mathbf{v}_i}{dt} = \mathbf{F}_i &= \sum_{i \neq j}(\mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R + \mathbf{F}_{ij}^E), \\
\frac{dc_{\alpha i}}{dt} = q_{\alpha i} &= \sum_{i \neq j}(q_{\alpha ij}^D + q_{\alpha ij}^E + q_{\alpha ij}^R),
\end{aligned}
\tag{13}
$$

where the total force $\mathbf{F}_i$ consists of three pairwise forces as in classical DPD method, i.e., the conservative force $\mathbf{F}_{ij}^C = a_{ij}w_C(r_{ij})\mathbf{e}_{ij}$, dissipative force $\mathbf{F}_{ij}^D = -\gamma_{ij}w_D(r_{ij})(\mathbf{e}_{ij} \cdot \mathbf{v}_{ij})\mathbf{e}_{ij}$, and random force $\mathbf{F}_{ij}^R = \sigma_{ij}w_R(r_{ij})\xi_{ij}\mathbf{e}_{ij}$, as well as an electrostatic force $\mathbf{F}^E$ coupling hydrodynamics and electrokinetics, given by

$$
\mathbf{F}_{ij}^E = \lambda_{ij}\rho_{ei}\mathbf{E}_{ij}, \quad \mathbf{E}_{ij} = (\phi_i - \phi_j)\omega_E(r_{ij})\hat{\mathbf{r}}_{ij},
\tag{14}
$$

where the coupling parameter λ_{ij} is linearly related to the macroscopic dimensionless coupling parameter $\Lambda = c_0 k_B T \tau^2/(\rho_0 r_0^2)$ with c_0 and ρ_0 as the reference concentration and reference mass density, which are chosen as the respective salt concentration and mass density of bulk solution. r_0 and τ are the unit length and time in DPD. $\rho_{ei} = \sum_\alpha z_\alpha c_{\alpha i}$ is the total charge density within the ith particle, and $\mathbf{E}_{ij}$ is the relative electric field determined by the difference of electrostatic potential field between particles i and j and a weighting function $\omega_E(r)$. Although the electrostatic forces are essentially not pairwise, i.e., $\mathbf{F}_{ij}^E \neq \mathbf{F}_{ji}^E$, a zero total force condition $\sum_{i,j}\mathbf{F}_{ij}^E = 0$ is always satisfied because of the charge neutrality condition $\sum_i \sum_\alpha z_\alpha c_{\alpha i} = 0$, which guarantees the global momentum conservation if there is no external electrostatic field.

Transport equations of ionic species are implemented in the cDPD model by introducing extra degrees of freedom and corresponding evolution equations associated with each cDPD particle. Diffusion of ionic species is driven by three flux terms, i.e., the Fickian flux q_{ij}^D induced by the concentration gradient, electrostatic flux q_{ij}^E induced by electrostatic potential gradient, and random flux q_{ij}^R due to thermal fluctuations, which can be written as

$$
\begin{aligned}
q_{\alpha ij}^D &= -\kappa_{\alpha ij}\,(c_{\alpha i} - c_{\alpha j})\,\omega_{qD}(r_{ij}), \\
q_{\alpha ij}^E &= -\kappa_{\alpha ij}\,z_\alpha\,\bar{c}_{\alpha ij}\,(\phi_i - \phi_j)\,\omega_{qD}(r_{ij}), \\
q_{\alpha ij}^R &= \xi_{\alpha ij}\,\omega_{qR}(r_{ij})\,\theta_{ij}\,\delta t^{-1/2},
\end{aligned}
\tag{15}
$$

where $\bar{c}_{\alpha ij} = (c_{\alpha i} + c_{\alpha j})/2$ is the average ionic concentration, $\kappa_{\alpha ij}$ is the diffusion coefficients of the αth species, and $\omega_{qD}(r_{ij})$ is a weighting function. The parameters in q_{ij}^D and q_{ij}^R are related via the generalized fluctuation-dissipation theorem, i.e., $\omega_{qD}(r) = \omega_{qR}^2(r)$ and $\xi_{\alpha ij}^2 = \kappa_{\alpha ij}(c_{\alpha i} + c_{\alpha j})/c_0 r_0^3$, which suggests that the variance of random flux is proportional to r_0^{-3}. Thus, the random flux cannot be neglected in micro-/nanoscales where $r_0 \ll 10^{-6} m$.

The electrostatic potential ϕ on each cDPD particle is determined by solving the Poisson equation at every time step. Consider the dimensionless Poisson equation rescaled by the DPD units $\nabla(\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})) = -\Gamma\rho_e(\mathbf{r})$, where $\varepsilon(\mathbf{r})$ is the local relative permittivity and $\Gamma = e^2 c_0 r_0^2/\varepsilon_0 k_B T$. The Poisson equation should be solved at every time step using, for example, a successive over-relaxation iteration scheme. Then, the electrostatic potential ϕ_i on the ith cDPD particle is obtained iteratively as

$$
\phi_i^k = \phi_i^{k-1} + \vartheta\left[\chi\rho_{ei} - \sum_{j\neq i}\bar{\varepsilon}_{ij}\phi_{ij}^k\omega_\phi(r_{ij})\right],
\tag{16}
$$

where χ is linearly related to the macroscopic parameter Γ that characterizes the strength of the electrostatic interactions, $\bar{\varepsilon}_{ij} = (\varepsilon_i + \varepsilon_j)/2$ is the average permittivity, $\omega_\phi(r)$ is a weight function, and ϑ is a relaxation factor. The value of ϑ can be adaptively changed during the iteration process, and the value of ϕ_i at the previous time step can be taken as an initial estimate of ϕ_i^{k-1} to achieve faster convergence. Then, the iteration of Eq. (16) is performed until the absolute differences $|\phi_i^k - \phi_i^{k-1}|$ are below a predefined tolerance.

Both the DPD model based on explicit ions (Groot 2003) and the cDPD model based on semi-implicit ions (Deng et al. 2016) have been applied to fluctuating electrohydrodynamic phenomena investigations. Posel et al. (2014) applied Groot's DPD model with explicit ions (Groot 2003) to study pH-dependent self-assembly of poly(2-vinylpyridine)-block-poly(ethylene oxide) diblock copolymers in aqueous media. They quantified the dependences of the apparent radius of gyration and

Fig. 3 (**a**) Apparent radius of gyration $(R_G)_{\langle A_S \rangle_w}$ and weight-averaged association number $\langle A_S \rangle_w$ as functions of the pH^{ap}, where the effect of the salt is depicted by the shift in the radius of gyration. (Adapted from Posel et al. 2014). (**b**) Mass density distribution for both positive (blue squares)- and negative (red circles)-charged polyelectrolytes drifting in electroosmotic flow between two positively charged surfaces. (Adapted from Deng et al. 2016)
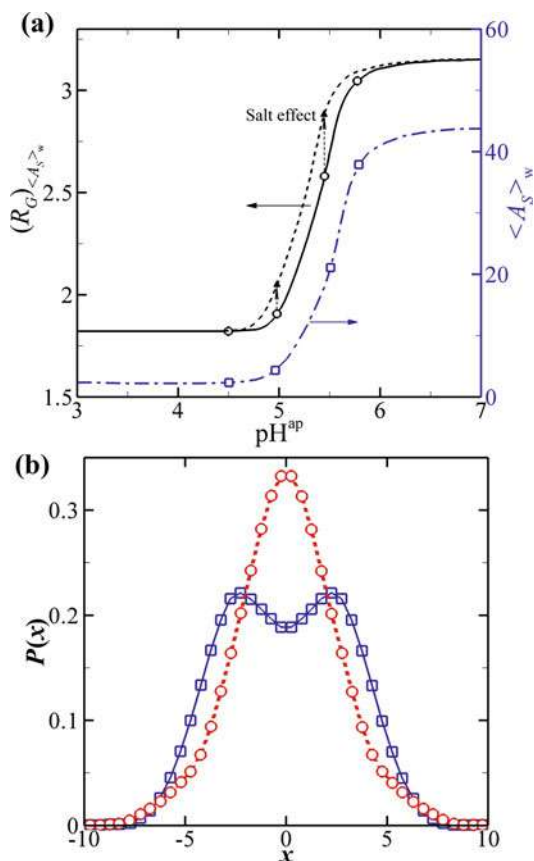
weight-averaged association number on pH, shown in Fig. 3a, and simulated the micellization process that coincides credibly with experimental observations. Deng et al. (2016) used the cDPD model to simulate the dilute polyelectrolyte solution drifting by electroosmotic flow in a microchannel. Unlike the migration of an uncharged polymer in the channel flow, they reported that negative-charged polyelectrolytes are likely to remain at the channel center because of the electrostatic repulsion from the charged surface, while the positive-charged polyelectrolytes drift away from the center with double peaks (Fig. 3b). Moreover, Zhou et al. (2013) investigated the response of a charged colloid and its surrounding microion cloud to an alternating electric field. Moshfegh and Jabbarzadeh (2016) simulated the electroosmotic flow in narrow nanochannels via explicit electrostatic interactions computed using the Ewald summation method. The authors acknowledge they currently are not aware of any top-down mesoscopic particle-based methods for ionic transport that consistently incorporate thermal fluctuations.

5   Summary

This chapter has considered different particle-based methods for mesoscopic transport processes, including thermal transport, reactive biochemical transport, and ionic transport in mesoscopic systems. These mesoscopic methods consistently incorporate thermal fluctuations based on non-equilibrium thermodynamics and are capable of describing certain mesoscopic features that deterministic macroscopic methods cannot model. The equations of a particle-based mesoscopic model can be derived from either bottom-up coarse-graining of microscopic dynamics or top-down discretization of macroscopic PDEs. Regardless of how mesoscopic models are obtained, they have similar structures, representing relevant physical mechanisms that dominate mesoscopic processes. Because mesoscopic methods are grounded in both macroscale and micro-/nanoscale models, they can play an important role in multiscale modeling of diseases and relevant bioengineering applications by seamlessly bridging the gap between microscopic molecular biology and macroscopic bulk behavior.

Despite considerable developments in recent years, particle-based mesoscopic methods are still evolving. New variations in models, novel theoretical interpretations, and innovative numerical algorithms often appear in literature. These methods have been applied to a large number of problems, including polymer and colloidal suspensions, multiphase fluids, biological materials, cell dynamics, and blood rheology, and new applications in different areas are still emerging. In the near future, these methods likely will be used even more widely for tackling challenging problems in biorheology, cellular biology, and bioengineering.

Additional research is needed in several areas, including the parameterization of mesoscopic models. Tuning model parameters manually by trial and error only works for simple systems. For problems with high-dimensional parameter spaces, Bayesian and/or machine-learning methods hold significant promise for inferring the model parameters (Lei et al. 2017). Coupling mesoscopic methods with other micro-/macroscopic methods to handle multiscale problems is another emerging research area. With concurrent coupling algorithms, i.e., domain decomposition (Tang et al. 2015) or adaptive resolution scheme (Praprotnik et al. 2006), in the lower limits, the mesoscopic methods can be coupled with microscale techniques, while in the upper bounds, the mesoscopic methods can be coupled with discretized PDEs. However, in concurrently coupled system, preserving correct fluctuations across the interface of heterogeneous solvers remains a challenging problem. Moreover, there is a need to decrease the computational cost of mesoscale methods, which could be achieved, for example, by improving the integration algorithms (Leimkuhler and Shang 2015) and efficient parallel implementations with graphics processing units (Blumers et al. 2017). Domain decomposition is routinely used to improve the scalability of particle codes. Parallel-in-time algorithms should be also investigated to further increase mesoscale method performance.

References

Abu-Nada E (2010) Natural convection heat transfer simulation using energy conservative dissipative particle dynamics. Phys Rev E 81(5):056704

Anand M, Rajagopal K, Rajagopal KR (2003) A model incorporating some of the mechanical and biochemical factors underlying clot formation and dissolution in flowing blood. J Theor Med 5(3-4):183–218

Avalos JB, Mackie AD (1997) Dissipative particle dynamics with energy conservation. Europhys Lett 40(2):141–146

Balluffi RW, Allen SM, Carter WC (2005) Kinetics of materials. Wiley, Hoboken

Bian X, Litvinov S, Qian R, Ellero M, Adams NA (2012) Multiscale modeling of particle in suspension with smoothed dissipative particle dynamics. Phys Fluids 24(1):012002

Blumers AL, Tang YH, Li Z, Li X, Karniadakis GE (2017) GPU-accelerated red blood cells simulations with transport dissipative particle dynamics. Comput Phys Commun 217:171–179

Cao ZH, Luo K, Yi HL, Tan HP (2013) Energy conservative dissipative particle dynamics simulation of natural convection in eccentric annulus. Int J Heat Mass Trans 65:409–422

Chen M, Li F, Wang S, Cao Y (2017) Stochastic modeling and simulation of reaction-diffusion system with Hill function dynamics. BMC Syst Biol 11:21

Cudjoe E, Khani S, Way AE, Hore MJA, Maia J, Rowan SJ (2017) Biomimetic reversible heat-stiffening polymer nanocomposites. ACS Cent Sci 3(8):886–894

Deng M, Li Z, Borodin O, Karniadakis GE (2016) cDPD: a new dissipative particle dynamics method for modeling electrokinetic phenomena at the mesoscale. J Chem Phys 145:144109

Drawert B, Hellander S, Trogdon M, Yi TM, Petzold L (2016) A framework for discrete stochastic simulation on 3D moving boundary domains. J Chem Phys 145(18):184113

Español P (1997) Dissipative particle dynamics with energy conservation. Europhys Lett 40(6):631–636

Español P, Revenga M (2003) Smoothed dissipative particle dynamics. Phys Rev E 67:026705

Español P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. Europhys Lett 30(4):191–196

Español P, Serrano M, Pagonabarraga I, Zúñiga I (2016) Energy-conserving coarse-graining of complex molecules. Soft Matter 12(21):4821–4837

Ewald PP (1921) Die berechnung optischer und elektrostatischer gitterpotentiale. Ann Phys 369(3):253–287

Fedosov DA, Sengupta A, Gompper G (2015) Effect of fluid-colloid interactions on the mobility of a thermophoretic microswimmer in non-ideal fluids. Soft Matter 11(33):6703–6715

González-Melchor M, Mayoral E, Velázquez ME, Alejandreb J (2006) Electrostatic interactions in dissipative particle dynamics using the Ewald sums. J Chem Phys 125:224107

Grmela M, Öttinger HC (1997) Dynamics and thermodynamics of complex fluids. I. Development of a general formalism. Phys Rev E 56(6):6620–6632

Groot RD (2003) Electrostatic interactions in dissipative particle dynamics-simulation of polyelectrolytes and anionic surfactants. J Chem Phys 118(24):11265–11277

Groot RD, Warren PB (1997) Dissipative particle dynamics: bridging the gap between atomistic and mesoscopic simulation. J Chem Phys 107(11):4423–4435

Guglielmi V, Voermans NC, Gualandi F, Van Engelen BG, Ferlini A, Tomelleri G, Vattemi G (2013) Fourty-four years of brody disease: it is time to review. J Genet Syndr Gene Ther 4(9):1–8

Hellander A (2008) Efficient computation of transient solutions of the chemical master equation based on uniformization and quasi-Monte Carlo. J Chem Phys 128(15):154109

Kordilla J, Pan W, Tartakovsky A (2014) Smoothed particle hydrodynamics model for Landau-Lifshitz-Navier-Stokes and advection-diffusion equations. J Chem Phys 141:224112

Kunzelmann K, Mall M (2003) Pharmacotherapy of the ion transport defect in cystic fibrosis. Am J Respir Med 2(4):299–309

Lei H, Baker NA, Wu L, Schenter GK, Mundy CJ, Tartakovsky AM (2016) Smoothed dissipative particle dynamics model for mesoscopic multiphase flows in the presence of thermal fluctuations. Phys Rev E 94:023304

Lei H, Yang X, Li Z, Karniadakis GE (2017) Systematic parameter inference in stochastic mesoscopic modeling. J Comput Phys 330:571–593

Leimkuhler B, Shang X (2015) On the numerical treatment of dissipative particle dynamics and related systems. J Comput Phys 280:72–95

Li Z, Tang YH, Lei H, Caswell B, Karniadakis GE (2014) Energy-conserving dissipative particle dynamics with temperature-dependent properties. J Comput Phys 265:113–127

Li Z, Tang YH, Li X, Karniadakis GE (2015a) Mesoscale modeling of phase transition dynamics of thermoresponsive polymers. Chem Commun 51(55):11038–11040

Li Z, Yazdani A, Tartakovsky A, Karniadakis GE (2015b) Transport dissipative particle dynamics model for mesoscopic advection-diffusion-reaction problems. J Chem Phys 143:014101

Li Z, Bian X, Li X, Deng M, Tang YH, Caswell B, Karniadakis GE (2017) Dissipative particle dynamics: foundation, evolution, implementation, and applications. Birkhäuser, Cham, pp 255–326

Litvinov S, Ellero M, Hu X, Adams NA (2008) Smoothed dissipative particle dynamics model for polymer molecules in suspension. Phys Rev E 77:066703

Mackie AD, Avalos JB, Navas V (1999) Dissipative particle dynamics with energy conservation: Modelling of heat flow. Phys Chem Chem Phys 1(9):2039–2049

Marsh CA, Backx G, Ernst MH (1997) Static and dynamic properties of dissipative particle dynamics. Phys Rev E 56(2):1676–1691

Medina S, Zhou J, Wang ZG, Schmid F (2015) An efficient dissipative particle dynamics-based algorithm for simulating electrolyte solutions. J Chem Phys 142(2):024103

Modi N, Winterhalter M, Kleinekathofer U (2012) Computational modeling of ion transport through nanopores. Nanoscale 4(20):6166–6180

Monaghan JJ (2005) Smoothed particle hydrodynamics. Rep Prog Phys 68(8):1703–1759

Moshfegh A, Jabbarzadeh A (2016) Fully explicit dissipative particle dynamics simulation of electroosmotic flow in nanochannels. Microfluid Nanofluid 20(4):67

Ortiz de Zárate JM, Sengers JV (2006) Hydrodynamic fluctuations in fluids and fluid mixtures. Elsevier, Amsterdam

Öttinger HC, Grmela M (1997) Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism. Phys Rev E 56(6):6633–6655

Perez-Herrero E, Fernandez-Medarde A (2015) Advanced targeted therapies in cancer: drug nanocarriers, the future of chemotherapy. Eur J Pharm Biopharm 93:52–79

Posel Z, Limpouchova Z, Sindelka K, Lisal M, Prochazka K (2014) Dissipative particle dynamics study of the pH-dependent behavior of poly(2-vinylpyridine)-block-poly(ethylene oxide) diblock copolymer in aqueous buffers. Macromolecules 47(7):2503–2514

Praprotnik M, Delle Site L, Kremer K (2006) Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. Phys Rev E 73:066701

Purves D, Augustine GJ, Fitzpatrick D, Hall WC, LaMantia AS, McNamara JO, Williams SM (2004) Neuroscience, 3rd edn. Sinauer Associates, Inc., Sunderland, Massachusetts U.S.A.

Rüegg JC (1992) Calcium in muscle contraction: cellular and molecular physiology. 2nd edn. Springer-Verlag Berlin

Schnoerr D, Sanguinetti G, Grima R (2017) Approximation and inference methods for stochastic biochemical kinetics – a tutorial review. J Phys A Math Theor 50(9):1–60

Taghizadeh B, Taranejoo S, Monemian SA, Moghaddam ZS, Daliri K, Derakhshankhah H, Derakhshani Z (2015) Classification of stimuli-responsive polymers as anticancer drug delivery systems. Drug Deliv 22(2):145–155

Tang YH, Kudo S, Bian X, Li Z, Karniadakis GE (2015) Multiscale universal interface: a concurrent framework for coupling heterogeneous solvers. J Comput Phys 297:13–31

Tang YH, Li Z, Li X, Deng M, Karniadakis GE (2016) Non-equilibrium dynamics of vesicles and micelles by self-assembly of block copolymers with double thermoresponsivity. Macromolecules 49(7):2895–2903

Vázquez-Quesada A, Ellero M, Español P (2009) Consistent scaling of thermal fluctuations in smoothed dissipative particle dynamics. J Chem Phys 130:034901

Vitvitsky VM, Garg SK, Keep RF, Albin RL, Banerjee R (2012) Na^+ and K^+ ion imbalances in Alzheimer's disease. Biochim Biophys Acta Mol Basis Dis 1822(11):1671–1681

Xu Z, Meakin P, Tartakovsky A, Scheibe TD (2011) Dissipative-particle-dynamics model of biofilm growth. Phys Rev E 83(6):066702

Zhou J, Schmitz R, Dunweg B, Schmid F (2013) Dynamic and dielectric response of charged colloids in electrolyte solutions to external electric fields. J Chem Phys 139:024901

Continuum- and Particle-Based Modeling of Human Red Blood Cells

107

Xuejin Li, Huijie Lu, and Zhangli Peng

Contents

Abstract

Computational modeling and simulations can tackle a broad range of morphological, mechanical, and rheological problems relevant to blood and blood cells. Here, we review some continuum-based and particle-based computational approaches toward the modeling of healthy and diseased red blood cells (RBCs) with focus on the most recent contributions, including the three-level multiscale RBC model coupled with the boundary integral method of surrounding flows and two-component RBC models with explicit descriptions of lipid bilayer, cytoskeleton, and transmembrane proteins.

X. Li (✉)
Department of Engineering Mechanics, Zhejiang University, Zhejiang,
People's Republic of China

Division of Applied Mathematics, Brown University, Providence, RI, USA
e-mail: xuejin_li@zju.edu.cn

H. Lu · Z. Peng
Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN, USA
e-mail: hlu3@nd.edu; zpeng3@nd.edu

1   Introduction

Blood is a bodily fluid that delivers oxygen and nutrients to living cells and takes away metabolic waste products from those same cells. It is composed of blood cells suspended in blood plasma, which makes up about 55% of the volume of whole blood. Blood plasma is the liquid component of blood, which contains mostly water and other substances like dissolved proteins, glucose, and clotting factors. Blood cells are mainly red blood cells (also called erythrocytes or RBCs), white blood cells (also called leukocytes or WBCs), and plasma (also called thrombocytes). Taking together, these three kinds of blood cells add up to a total 45% of the volume of whole blood. In particular, RBCs are the most common type of blood cells, with each cubic millimeter (1 mm^3) of blood containing about 4–6 million RBCs.

A normal RBC is a nucleus-free cell; it is essentially a cell membrane encapsulating a hemoglobin solution. Upon maturation, it adopts a unique biconcave shape of approximately $8.0\,\mu m$ in diameter and $2.0\,\mu m$ in thickness (Fig. 1a). The RBC membrane consists of three basic components: a lipid bilayer, a cytoskeletal network, and transmembrane proteins such as glycophorins, band 3, and some substoichiometric glycoproteins (e.g., CD44, CD47) (Fig. 1b). The deformability of an RBC is determined by the elasticity, geometry, and viscosity of the RBC membrane. A normal RBC is highly deformable allowing it to squeeze through capillaries as small as 3 microns in diameter without any damage. Several pathological conditions, including malaria, sickle cell disease, hereditary spherocytosis, and elliptocytosis, can alter the shape and deformability of circulating RBCs, leading to possible vascular obstruction. Over the past few decades, computational models, such as continuum-based and particle-based RBC models, have been proven to be an important tool in accurately resolving a broad range of biological problems associated with RBCs (Abkarian and Viallat 2008; Vlahovska et al. 2011; Li et al. 2013, 2017b). Actually, there is a great variety of modeling approaches since no universal solution for all blood flow related problems. In this chapter, we overview the computational approaches toward the modeling of RBCs focusing on the most recent contributions.
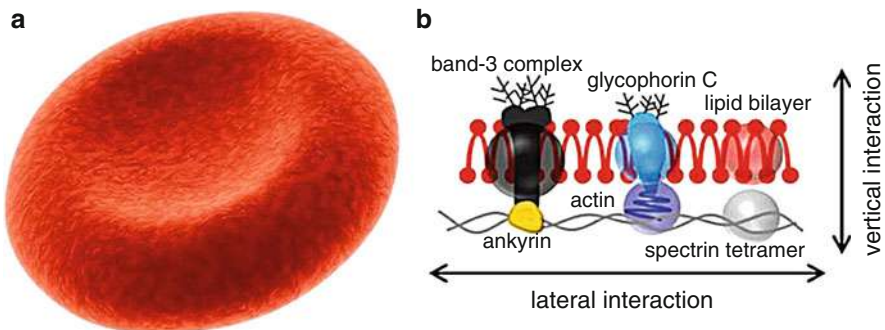
Fig. 1 Schematic representation of the morphology of (**a**) a normal RBC and (**b**) its two-component membrane structure (Li et al. 2017a)

For the most problems, the surrounding blood plasma plays a significant role in the RBC dynamics. The accurate modeling of the fluid-structure interaction between the RBCs and the surrounding flow is crucial to predict the physics. There are three major groups of numerical methods in the existing literature to study motion and deformation of RBCs in flow (Pozrikidis 2010; Barthès-Biesel 2016; Freund 2014). In the first group of methods, boundary integral equations are used to represent the Stokes flow and coupled to the membrane (Walter et al. 2010; Pozrikidis 1992), which is usually modeled as meshes or spectral elements (Zhao et al. 2010) either by enforcing the membrane equilibrium equation directly (Pozrikidis 1992) or by the finite element method based on the virtual work principle (Walter et al. 2010; Peng et al. 2011). This type of methods has been known to be very precise in various studies of simple shear flow and plane hyperbolic flow (Barthès-Biesel 2016; Freund 2014). In the second group of the immerse-boundary type methods, the fluid domain is solved using either the finite difference method (Yazdani et al. 2011) or the lattice Boltzmann method (LBM) (Sui et al. 2008), and the capsule membrane is solved using continuum methods such as finite element method (FEM). The force from capsule membrane nodes is distributed to the fluid domain as Dirac delta functions, which are approximated by smooth functions sharply varying over a few fixed grid meshes. The precision of these methods may be reduced as the methods do not treat the membrane as physical boundaries but as approximated Dirac functions. In the third group, particle methods such as dissipative particle dynamics (DPD) (Pivkin and Karniadakis 2008; Peng et al. 2013), multiparticle collision dynamics (MPCD) (Noguchi and Gompper 2005; McWhirter et al. 2009), and smoothed particle hydrodynamic (SPH) method (Hosseini and Feng 2012) are used, and the capsule and RBC membranes are modeled as triangular networks of particles with bond, angle, and dihedral interactions (Discher et al. 1998). In addition to these three groups, phase-field methods (Du et al. 2004) and arbitrary Lagrangian Eulerian (ALE) method using FEM (Barber et al. 2011; Ni et al. 2015) are used as well to study capsules and vesicles.

2   Continuum-Based Models

Continuum-based models provided useful insights in many aspects of RBC mechanics by adopting the robust and accurate numerical methods developed in the well-established continuum mechanics community (Barthès-Biesel 2016). The cell membrane is typically 7.5–10 nm in thickness; hence it is usually treated as a two-dimensional viscoelastic material embedded in a three-dimensional space. Traditionally, continuum-based RBC models treat RBC membrane and intracellular hemoglobin solution as homogeneous materials, using boundary integral method (BIM) (Ramanujan and Pozrikidis 1998; Lac et al. 2004; Zhao et al. 2010; Veerapaneni et al. 2011), immersed boundary method (IBM) (Peskin 2002; Doddi and Bagchi 2009; Yazdani and Bagchi 2011; Fai et al. 2013), and fictitious domain method (FDM) (Shi et al. 2014; Hao et al. 2015). These continuum-based models describe the modeling system using locally averaged variables, such as density, stress, and velocity, with ordinary and partial differential equations governing flow

dynamics and fluid mechanics. Blood flow in microcirculation normally falls in the Stokes regime where inertial effects are negligible, hence making BIM a popular technique for addressing multiphase flow problems. The BIM method exploits the fact that the equations of fluid motion are linear and can be recast into an integral equation for the evolution of the interface (Ramanujan and Pozrikidis 1998; Lac et al. 2004; Zhao et al. 2010; Veerapaneni et al. 2011). Among other continuum solvers, the IBM has become very effective in addressing fluid-structure problems (Peskin 2002; Fai et al. 2013). There are several numerical extensions of IBM depending on the choice of the structural and fluid formulations. For example, the front-tracking method (Doddi and Bagchi 2009; Yazdani and Bagchi 2011) used a first-order finite element triangulation of cell membrane, while the full Navier-Stokes equation has been solved using projection splitting scheme. More recently, FDM for solving coupled fluid-structure systems by introducing a Lagrange multiplier over the solid domain has been used (Shi et al. 2014; Hao et al. 2015). This multiplier acts as a penalty term that imposes the kinematic constraint in the solid domain. Another computational challenge is to include the membrane thermal undulations; progress in this direction has been made by introducing a stochastic formulation for IBM (Atzberger et al. 2007).

These continuum-based RBC models allow the study of blood flow dynamics on macroscopic length and time scales. However, the RBC membrane is considered as a uniform continuum media in these continuum models, and the detailed molecular structure was not considered. Recently, multiscale RBC models that include sufficient molecular details have been developed. For example, Peng et al. have developed a three-level multiscale model of RBCs and coupled it with a boundary element model to study the dynamic response of RBCs in tube and shear flows (Peng et al. 2010). For the RBC membrane, we built a three-level multiscale model to simulate its viscoelastic behaviors at different length scales, including single protein scale, protein complex scale, and cell scale. For the fluids surrounding the RBC, we used a boundary element approach to study the fluid dynamics based on the Stokes equation of viscous flow (Peng et al. 2011).

The dynamics of RBCs involves physics at different length scales. For example, in the whole-cell level, RBCs may undergo tank-treading motions in shear flow in the micro meter scale (Fischer et al. 1978; Tran-Son-Tay et al. 1984). In the protein complex scale, junctional complexes of RBCs may experience fluctuations due to mode switching (Lee and Discher 2001; Zhu et al. 2007). In the protein level, tension may induce spectrin unfolding in the nanoscale (Rief et al. 1999; Zhu et al. 2007). We developed a multiscale framework to investigate the physics in different length scales, including three models at three different length scales, and coupled them together using a hierarchical multiscale approach, as shown in Fig. 2.

At Level I, Zhu and Asaro developed a model of spectrin using Monte-Carlo simulation (Zhu and Asaro 2008). Spectrin is the major RBC cytoskeleton protein that forms a triangular network. Rief et al. applied AFM to stretch a single spectrin and found that it can be modeled accurately using a nonlinear constitutive model called worm-like chain (WLC) (Rief et al. 1999; Discher et al. 1998; Li et al. 2005). Besides that, it was found that if the spectrin is stretched beyond its contour
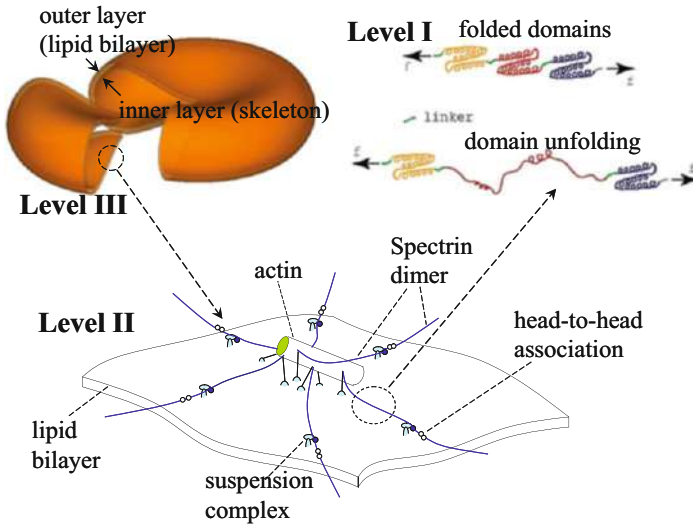
Fig. 2 Two-component RBC model using three-level hierarchical multiscale method (Peng et al. 2010)

length, multiple protein domains may unfold to increase the contour length. In order to capture the full mechanical behavior of the spectrin, Zhu and Asaro took into account its domain unfolding feature by employing a Monte-Carlo method, as the probability of the unfolding is a function of the loading (Zhu and Asaro 2008). By simulating the unfolding of the spectrin, Zhu and Asaro got consistent force-displacement curves as in the AFM experiment (Rief et al. 1999). In addition, the effect of stretching rate is also quantified. After these force-displacement curves are obtained, the information is passed as a database to the next level model (Level II).

At Level II, Zhu et al. developed a 3D junctional complex model with molecular details using Brownian dynamics (BD), which is a basic protein complex unit in the RBC cytoskeleton (Zhu et al. 2007). In the junctional complex, an actin protofilament is associated with six or five spectrins. Zhu et al. built an exact geometry model based on the state-of-the-art understanding of the molecular structure of a junctional complex, including the binding sites between spectrins and actins (Sung and Vera 2003). They simulated thermal fluctuations of the RBC membrane using this model (Zhu et al. 2007), and predicted the area and shear moduli of the cytoskeleton, as functions of membrane deformation (Zhu et al. 2007; Peng et al. 2010), which can be passed to the next cell level model (Level III).

At Level III, Peng et al. developed a whole-cell model of RBCs using finite element method (FEM) and considered the normal and tangential interactions between the lipid bilayer and the cytoskeleton (Peng et al. 2010, 2011). The cytoskeleton may slide against the lipid bilayer (Fischer 1992; Dodson and Dim-itrakopoulos 2010), since the cytoskeleton is connected to the lipid bilayer through transmembrane proteins such as band 3 and glycophorin C, which can move freely

within the lipid bilayer. In order to describe the bilayer-cytoskeletal interaction accurately, we modeled the RBC membrane as two distinct layers. We employed an inner layer to present the cytoskeleton with finite area and shear stiffnesses, and negligible bending stiffness, and an outer layer to represent the lipid bilayer with zero shear stiffness, huge area stiffness, and significant bending stiffness. Vertical elastic interaction and tangential viscous friction between the lipid bilayer and the cytoskeleton are simulated using a contact algorithm based on the penalty method (Malone and Johnson 1994). The bilayer-cytoskeletal friction coefficient is estimated based on the Stokes-Einstein relation and experimentally measured diffusivities of band 3 and glycophorin C in the lipid bilayer (Kapitza et al. 1984; Kodippili et al. 2009; Tomishige 1998). Reduced integration shell elements based on Mindlin theory were used to simulate both the outer and the inner layers (Peng et al. 2010).

Besides the elasticity, we also considered the viscosities of the lipid bilayer and the cytoskeleton, which play significant roles in dynamics processes such as tank-treading. For example, it has been demonstrated that without considering the membrane viscosity, the predicted tank-treading frequency cannot match the experimental measurement (Fedosov et al. 2011a; Peng et al. 2011). Different from the existing study (Fedosov et al. 2011a), we consider the viscosities of the lipid bilayer and the cytoskeleton separately using a generalized Voigt-Kelvin stress-strain relation (Evans and Skalak 1980). Furthermore, we modeled the sliding between the bilayer and the skeleton as a viscous friction force. The viscous fluids surrounding the cell are incorporated in the following boundary element model of Stokes flow.

We applied a boundary element method to simulate the Stokes flows inside and outside of the RBC and coupled it with FEM to study the fluid-structure interaction. We used a staggered algorithm with explicit time integration (Peng et al. 2011). For the interface dynamics in the Stokes flow with zero Reynolds number, we applied the boundary integral equation (Pozrikidis 1992) so that the velocity $\mathbf{v}^f$ is given as

$$
\begin{aligned}
\mathbf{v}^f(\mathbf{x}_0) \;=\; & \frac{2}{1+\Lambda}\bar{\mathbf{v}}^f(\mathbf{x}_0) \\
& -\frac{1}{4\pi\eta_1(\Lambda+1)}\int_{\Gamma^{fb}}\mathbf{G}(\mathbf{x},\mathbf{x}_0)\cdot\Delta\mathbf{t}^f(\mathbf{x})d\Gamma(\mathbf{x}) \\
& +\frac{1-\Lambda}{4\pi(1+\Lambda)}\!\!\iint_{\Gamma^{fb}}\mathbf{v}^f(\mathbf{x})\cdot\mathbf{T}(\mathbf{x},\mathbf{x}_0)\cdot\mathbf{n}(\mathbf{x})d\Gamma(\mathbf{x}),
\end{aligned}
\tag{1}
$$

where Γ^{fb} is the membrane surface and $\bar{\mathbf{v}}^f$ is the prescribed undisturbed velocity field of the shear flow. The vector $\Delta\mathbf{t}^f = \mathbf{t}^{f,1} - \mathbf{t}^{f,2}$ is the discontinuity in the interfacial surface traction, where $\mathbf{t}^{f,1}$ is the traction in the outside surface $\Gamma^{fb,1}$ of the interface and $\mathbf{t}^{f,2}$ is the traction in the inside surface $\Gamma^{fb,2}$ of the interface. The surface traction is related to the nodal force through the principle of virtual work (Walter et al. 2010). $\iint$ denotes the principal value integration.

G is the Green's function for velocity. Its components are

$$G_{ij}(\mathbf{x}, \mathbf{x}_0) = \frac{\delta_{ij}}{|\mathbf{x} - \mathbf{x}_0|} + \frac{(x_i - x_{0_i})(x_j - x_{0_j})}{|\mathbf{x} - \mathbf{x}_0|^3}, \qquad (2)$$

where δ_{ij} is Kronecker's delta. **T** is the Green's function for stress. It can be written as

$$T_{ijk}(\mathbf{x}, \mathbf{x}_0) = -6\frac{(x_i - x_{0_i})(x_j - x_{0_j})(x_k - x_{0_k})}{|\mathbf{x} - \mathbf{x}_0|^5}. \qquad (3)$$

Extensive validations have been conducted in each level to verify the models, such as optical tweezers stretching (Peng et al. 2010), micropipette aspiration (Peng et al. 2010), tank-treading, and tumbling of RBCs in shear flow (Peng et al. 2011). More importantly, the effect of stress-free cytoskeleton state on the tank-treading behavior has been compared with the available experiments in details (Dupire et al. 2012; Peng et al. 2014, 2015). Besides validations, we predicted correlation between the occurrence of spectrin unfolding and increase in the mechanical load upon individual skeleton-bilayer pinning points and related it to the vesiculation process (Knowles et al. 1997). The simulation results also show that during tank-treading, the protein density variation is insignificant for healthy RBCs, but significant for cells with a smaller bilayer–cytoskeletal friction coefficient, which may be the case in hereditary spherocytosis (Walensky et al. 2003). We also predicted two different modes of motions for RBCs in shear flow (Peng and Zhu 2013) and studied the effect of stress-free state on the tank-treading motion (Peng et al. 2014, 2015; Dupire et al. 2012). We showed that the cell maintains its biconcave shape during tank-treading motions under low shear rate flows, by employing a spheroidal stress-free state in the cytoskeleton (Peng et al. 2014). Furthermore, we numerically confirmed the hypothesis that, as the stress-free state approaches a sphere, the threshold shear rates corresponding to the establishment of tank treading decrease. By comparing with the experimental measurements (Dupire et al. 2012), our study suggests that the stress-free state of RBCs is a spheroid that is close to a sphere, rather than the biconcave shape applied in existing models (Peng et al. 2014). In addition, we also quantified the stability phase diagram of different motion modes in high shear rate flows and explored the effect of stress-free state on the phase diagram (Peng et al. 2015).

3   Particle-Based Models

Particle-based models treat both the fluid and the cell membrane as particulate materials, and they can model critical biophysical processes involving the RBCs at cellular and subcellular scales, such as membrane fluctuations and the defective membrane structure. Several particle-based RBC models, including molecular-detailed composite RBC models (Li and Lykotrafitis 2012, 2014; Tang et al. 2017)
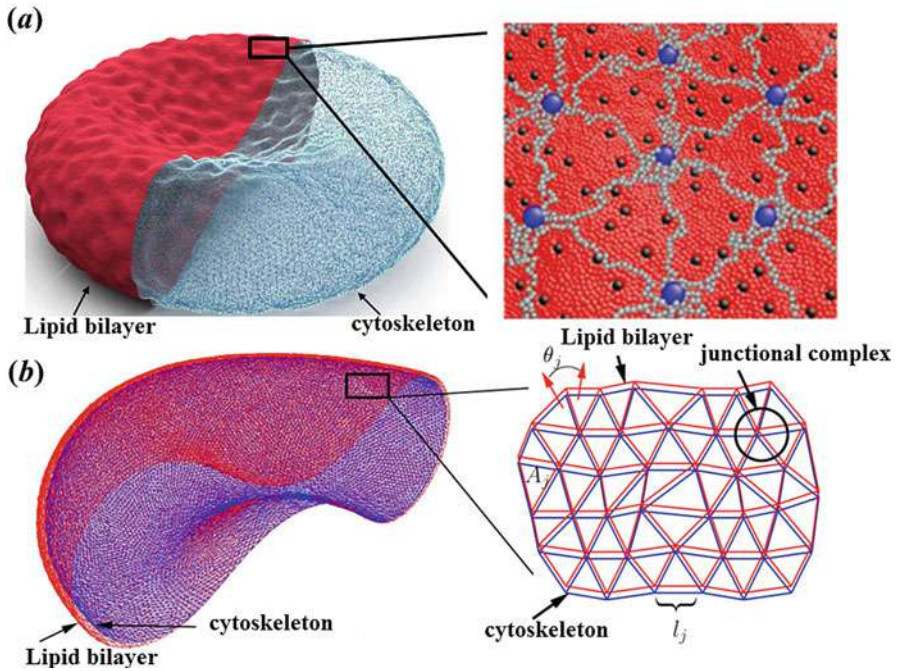
Fig. 3 Two-component RBC model at protein resolution (Tang et al. 2017; Li and Lykotrafitis 2014) and coarse-grained level (Peng et al. 2013)

and coarse-grained RBC models (Discher et al. 1998; Li et al. 2005; Noguchi and Gompper 2005; Pivkin and Karniadakis 2008; Pan et al. 2010; Fedosov et al. 2010; Peng et al. 2013), have been developed and employed to quantify the structural, mechanical, and rheological properties of RBCs in health and disease.

Generally, the molecular-detailed composite RBC models describe explicitly lipid bilayer, cytoskeleton, and transmembrane proteins using coarse-grained molecular dynamics (CGMD) (Li and Lykotrafitis 2012, 2014). They are able to capture the membrane-related structural problems in RBCs. Specifically, the spectrin tetramer in the composite membrane model is represented by a chain of 39 beads (gray particles in Fig. 3a) connected by spring bonds (Li et al. 2007). The corresponding potential has the form,

$$V_{cy}^{s-s} = \frac{1}{2}k_0(r - r_{eq}^{s-s})^2 \tag{4}$$

where k_0 and r_{eq}^{s-s} are the spring constant and equilibrium distance between two spectrin particles, respectively. Spectrin particles that are not connected by the spring potential interact with each other via the repulsive term of Lennard-Jones potential, given by

$$V_{\text{rep}}(r_{ij}) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \epsilon & r_{ij} < r_{\text{eq}}^{s-s}, \\ 0 & r_{ij} \geq r_{\text{eq}}^{s-s} \end{cases} \tag{5}$$

where ϵ is the energy unit and σ_{ij} is the length unit, and r_{ij} is the distance between spectrin particles. To couple the lipid bilayer and spectrin network, actin junctional complexes are connected to the glycophorin C, and the middle beads of the spectrin network are bonded to the band-3 complexes which are specifically rendered in yellow particles. These bonds are modeled as harmonic springs, given by

$$V_{\text{cy}}^{a-s} = \frac{1}{2} k_0 (r - r_{\text{eq}}^{a-s})^2, \tag{6}$$

where $r_{\text{eq}}^{a-s} = 10\,\text{nm}$ is the equilibrium distance between an actin and a spectrin particle.

These molecular-detailed composite membrane models have been successfully to study some issues associated with RBC membrane defects, such as the multiple stiffening effects of nanoknobs on malaria-infected RBCs (Zhang et al. 2015; Chang et al. 2016), the vesiculation of mature RBCs in hereditary elliptocytosis and spherocytosis (Li and Lykotrafitis 2015), and the diffusion of transmembrane proteins in normal and defective RBC membrane (Li et al. 2016). However, modeling only a portion of the RBC membrane does not efficiently depict the whole-cell characteristics strongly related to RBC mechanics and rheology.

More recent efforts have focused on this approach, leading to a two-component whole-cell model (OpenRBC) at protein resolution with explicit descriptions of lipid bilayer, cytoskeleton, and transmembrane proteins (Fig. 3a) (Tang et al. 2017). In the OpenRBC, the modeled RBC is generated from a triangular mesh of cytoskeleton that resembles the biconcave disc shape of a normal RBC, where the lipid molecules and mobile band-3 particles are randomly placed on triangular faces. The modeled RBC system is also optimized by a velocity quenching algorithm to avoid the overlap of particles. It can simulate an entire RBC with a resolution down to single protein level, which offers unique tools for the qualitative and quantitative predictions of dynamic and mechanical properties of healthy and pathological RBCs beyond available experimental measurements, such as the vesiculation of RBCs in hereditary spherocytosis and elliptocytosis (Tang et al. 2017), and rupture of RBCs in splenic passage (Li et al. 2018). However, at present, they are computationally very expensive for large-scale applications such as modeling whole blood involving a large number of RBCs. On the other hand, coarse-grained whole-cell model provides the opportunity to significantly reduce the computational complexity. For example, a three-dimensional multiscale RBC (MS-RBC) model has been developed and successfully applied to RBC simulations at different length scales (Pivkin and Karniadakis 2008; Fedosov et al. 2010).

In the MS-RBC model, the cell membrane is modeled by a 2D triangulated network with N_v vertices connected by springs, where each vertex is represented by a DPD particle. The RBC membrane model takes into account the elastic energy,

bending energy, and constraints of fixed surface area and enclosed volume, which is defined as

$$V = V_s + V_b + V_a + V_v, \tag{7}$$

where V_s is the elastic energy that mimics the elastic spectrin network, given by

$$V_s = \sum_{i \in \text{springs}} \left[\frac{k_B T l_m}{4p} \frac{3x_i^2 - 2x_i^3}{1 - x_i} \right] + \sum_{\alpha \in \text{triangles}} \frac{1}{A_\alpha} \left[\frac{3\sqrt{3}k_B T l_m^3 x_0^4}{64p} \frac{4x_0^2 - 9x_0 + 6}{(1 - x_0^2)} \right], \tag{8}$$

where $k_B T$ is the energy unit and A_α is the area of triangle α formed by three vertices. Also, $x_i = l_i/l_m$, $x_0 = l_0/l_m$, where l_i is the length of spring i, l_0 and l_m are the equilibrium spring length and maximum spring extension, and p is the persistence length.

The cell membrane viscoelasticity is imposed by introducing a viscous force on each spring, which has the form,

$$\mathbf{F}_{ij}^D = -\gamma^T \mathbf{v}_{ij} - \gamma^C (\mathbf{v}_{ij} \cdot \mathbf{e}_{ij}) \mathbf{e}_{ij}, \tag{9}$$

$$\mathbf{F}_{ij}^R dt = \sqrt{2k_B T} \left(\sqrt{2\gamma^T} d\overline{\mathbf{W}}_{ij}^S + \sqrt{3\gamma^C - \gamma^T} \frac{tr[d\mathbf{W}_{ij}]}{3} \mathbf{1} \right) \cdot \mathbf{e}_{ij}, \tag{10}$$

where γ^T and γ^C are dissipative parameters, $\mathbf{v}_{ij}$ is the relative velocity of spring ends, and $d\overline{\mathbf{W}}_{ij}^S = d\mathbf{W}_{ij}^S - tr[d\mathbf{W}_{ij}^S]\mathbf{1}/3$ is the traceless symmetric part of a random matrix representing the Wiener increment.

The bending resistance of the RBC membrane is modeled by

$$V_b = \sum_{\alpha, \beta \text{ pair}} k_b \left[1 - \cos(\theta_{\alpha\beta} - \theta_0) \right], \tag{11}$$

where k_b is the bending modulus constant, $\theta_{\alpha\beta}$ is the instantaneous angle between two adjacent triangles having common edge, and θ_0 is the spontaneous angle. In addition, the RBC model includes the area and volume conservation constraints, which mimic the area-incompressibility of the lipid bilayer and the incompressibility of the interior fluid, respectively. The corresponding energy terms are given by

$$V_a = \frac{k_a k_B T (A - A_0)^2}{2l_0^2 A_0}, \quad V_v = \frac{k_v k_B T (V - V_0)^2}{2l_0^3 V_0}, \tag{12}$$

where k_a and k_v are the area and volume constraint coefficients. Here A_0 and V_0 are the equilibrium area and volume of a cell, respectively.

The MS-RBC model is multiscale, as the RBC can be represented on the spectrin level, where each spring in the network corresponds to a single spectrin tetramer with the equilibrium distance between two neighboring actin connections of ~75 nm. On the other hand, for more efficient computations, the RBC network can also be highly coarse-grained with the equilibrium spring lengths of up to 500~600 nm. In addition, being constructed from a coarse-grained molecular dynamics (CGMD) approach, the MS-RBC model can naturally include membrane thermal fluctuations (Fedosov et al. 2011b; Chang et al. 2017). Such formulations are compatible with coarse-grained mechanical descriptions of the RBC membrane with the advantage of including the viscosity of the RBC membrane without additional cost. Thus, the particle-based whole-cell models can resolve subcellular and cellular scales.

A two-component particle-based whole-cell model, which separately accounts for the lipid bilayer and cytoskeleton but implicitly includes the transmembrane proteins, has been developed and implemented using DPD (Fig. 3b) (Peng et al. 2013). In this two-component whole-cell model, the cell membrane is modeled by two distinct components, i.e., the lipid bilayer and the spectrin network. Specifically, through the DPD approach, each component is constructed by a two-dimensional triangulated network, which is analogous to the one-component whole-cell model; however, the lipid bilayer of the two-component whole-cell model has no shear stiffness at healthy state but only bending stiffness and a very large local area stiffness, whereas the cytoskeleton has no bending stiffness but possesses a finite shear stiffness. In addition, this unique two-component whole-cell model takes into account the bilayer-cytoskeleton interaction potential, which is expressed as a summation of harmonic potentials given by

$$V_{\text{int}} = \sum_{j,j' \in 1 \dots N_{bs}} \frac{k_{bs}(d_{jj'} - d_{jj',0})^2}{2},$$ (13)

where k_{bs} and N_{bs} are the spring constant and the number of bond connections between the lipid bilayer and the cytoskeleton, respectively. $d_{jj'}$ is the distance between the vertex j of the cytoskeleton and the corresponding projection point j' on the lipid bilayer, with the corresponding unit vector $\mathbf{n}_{jj'}$; $d_{jj',0}$ is the initial distance between the vertex j and the point j', which is set to zero in the current simulations. Physical view of the local bilayer-cytoskeleton interactions includes the major connections via band-3 complex and ankyrin, as well as the secondary connections via glycophorin C and actin junctions; here we consider them together as an effective bilayer-cytoskeleton interaction and model it as a normal elastic force, $\mathbf{f}_{jj'}^E$, and a tangential friction force, $\mathbf{f}_{jj'}^F$. The corresponding elastic force on the vertex j of the cytoskeleton is given by

$$\mathbf{f}_{jj'}^E = \begin{cases} k_{bs}(d_{jj'} - d_{jj',0})\mathbf{n}_{jj'} & d_{jj'} < d_c \\ 0 & d_{jj'} \geq d_c \end{cases}$$ (14)

where $d_c \approx 0.2\,\mu m$ is the cutoff distance. The tangential friction force between the two components is represented by

$$\mathbf{f}_{jj'}^{F} = -f_{bs}[\mathbf{v}_{jj'} - (\mathbf{v}_{jj'} \cdot \mathbf{n}_{jj'})\mathbf{n}_{jj'}], \tag{15}$$

where f_{bs} is the tangential friction coefficient and $\mathbf{v}_{jj'}$ is the difference between the two velocities.

The two-component whole-RBC model is capable of simulating the interactions between the lipid bilayer and the cytoskeletal network of human RBCs (Peng et al. 2013; Li et al. 2014; Chang et al. 2016). Furthermore, it is computationally more efficient than the two-component OpenRBC model for blood flow modeling. Herein, the two-component whole-cell RBC model has been used in simulations of blood flow and quantified the existence of bilayer-cytoskeletal slip for RBC membrane in pathological state (Peng et al. 2013; Li et al. 2014).

4   Summary

To summarize the continuum-based and particle-based models of RBCs, we list the main features and applications of some existing mesoscale RBC models in Table 1. The main differences are the molecular details considered and the length scales these models can reach. The two-component CG model focused on the detailed bilayer-cytoskeletal interaction and transmembrane protein diffusion, but it is too expensive to employ it to study a whole cell with current methodology. The two-component DPD model studied the whole-cell behavior and explored the bilayer-cytoskeletal interaction in a simplified way. The three-level multiscale approach is developed to bridge three models in different length scales, but due to the information passing, some information is lost, such as the thermal fluctuations in the whole-cell level.

There is still plenty of room to improve these models. For example, only one-way coupling has been done for the three-level multiscale RBC model, and two-way coupling can be added to study how the deformation in the macroscale can influence the dynamics of the proteins in the microscale. In addition, future effort will be put on how to improve the efficiency of the two-component CGMD RBC model in order to apply it in the whole-cell scale and couple it with fluid motion. Such simulations could shed light on the coupling of biology, chemistry, and mechanics (the "triple-point"). To make model predictions more reliable, it would require further mesoscopic validation and reliability testing of these models against experimental and clinical studies. Moreover, smoothed dissipative particle dynamics (SDPD) and smoothed particles hydrodynamics (SPH) have been also applied to study RBCs models (Hosseini and Feng 2012; Fedosov et al. 2014; Müller et al. 2015). Compared to conventional DPD, SDPD adopt advantages from smoothed particles hydrodynamics (SPH) (Van Liedekerke et al. 2013), such as thermal consistency and specification of viscosity.

Table 1 Comparison of different RBC models

RBC models	Main characteristics	Examples of application
Two-component CGMD composite model (Li and Lykotrafitis 2012, 2014; Zhang et al. 2015)	One-particle-thick bilayer model Small piece of membrane	Transmembrane protein diffusion RBC membrane defects RBC membrane vesiculation
Three-level multiscale model (Peng et al. 2010, 2011)	FLM, LD and MC Spanning from protein to cell	Micropipette vesiculation Dynamics junctional complexes Spectrin unfolding
OpenRBC model (Tang et al. 2017)	One-particle-thick bilayer model Bilayer-cytoskeletal interaction Whole-cell level	RBC vesiculation RBC lysis
One-component DPD model (Pivkin and Karniadakis 2008; Fedosov et al. 2010)	Coupled with fluids Systematic coarse-graining Whole-cell level	Membrane thermal fluctuations Tank-treading motion Microfluidics
Two-component DPD model (Peng et al. 2013; Li et al. 2014)	Bilayer-cytoskeletal interaction Whole-cell level	Bilayer-cytoskeletal slip Bilayer tethering Membrane thermal fluctuations
Two-component MPCD model (Noguchi and Gompper 2005; McWhirter et al. 2009)	Coarse-grained model Whole-cell level	Tube flow Tank-treading motion
LBM model (Zhang et al. 2008; Reasor et al. 2012)	Fluid-structure interaction Whole-cell level	Blood rheology Cell margination
Continuum model (Ramanujan and Pozrikidis 1998; Dao et al. 2003; Yazdani and Bagchi 2011; Freund 2013)	Uniform property No molecular details	Large-scale blood flow modeling Tank-treading Optical tweezers stretching

Since the RBC is a model system for cell mechanics and biological membranes, RBC models can be applied in other systems as well. For example, recently we extended the two-component RBC DPD model to study the blood vessel walls by considering the isotropic matrix and the anisotropic collagen fibers separately (Witthoft et al. 2016). Furthermore, we can also apply RBC models to study the lipid bilayer and the actin cortex network or the intermediate filament network in general eukaryotic cells, because some of them share similar structures as the spectrin network of RBCs, such as the auditory outer hair cells and the nucleus lamina (Alberts et al. 2002; Boal 2012).

References

Abkarian M, Viallat A (2008) Vesicles and red blood cells in shear flow. Soft Matter 4:653

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Molecular biology of the cell. Garland Science, New York

Atzberger PJ, Kramer PR, Peskin CS (2007) A stochastic immersed boundary method for fluid-structure dynamics at microscopic length scales. J Comput Phys 224:1255–1292

Barber JO, Restrepo JM, Secomb TW (2011) Simulated red blood cell motion in microvessel bifurcations: Effects of cell–cell interactions on cell partitioning. Cardiovasc Eng Technol 2(4):349–360

Barthès-Biesel D (2016) Motion and deformation of elastic capsules and vesicles in flow. Annu Rev Fluid Mech 48:25–52

Boal D (2012) Mechanics of the Cell, 2nd edn. Cambridge University Press, Cambridge

Chang HY, Li X, Li H, Karniadakis GE (2016) MD/DPD multiscale framework for predicting morphology and stresses of red blood cells in health and disease. PLoS Comput Biol 12(10):e1005173

Chang HY, Li X, Karniadakis GE (2017) Modeling of biomechanics and biorheology of red blood cells in type 2 diabetes mellitus. Biophys J 113(2):481–490

Dao M, Lim CT, Suresh S (2003) Mechanics of the human red blood cell deformed by optical tweezers. J Mech Phys Solids 51:2259–2280

Discher DE, Boal DH, Boey SK (1998) Simulations of the erythrocyte cytoskeleton at large deformation. ii. micropipette aspiration. Biophys J 75:1584–1597

Doddi SK, Bagchi P (2009) Three-dimensional computational modeling of multiple deformable cells flowing in microvessels. Phys Rev E 79(4):046318

Dodson WR III, Dimitrakopoulos P (2010) Tank-treading of erythrocytes in strong shear flows via a nonstiff cytoskeleton-based continuum computational modeling. Biophys J 99:2906–2916

Du Q, Liu C, Wang X (2004) A phase field approach in the numerical study of the elastic bending energy for vesicle membranes. J Comput Phys 198(2):450–468

Dupire J, Socol M, Viallat A (2012) Full dynamics of a red blood cell in shear flow. Proc Natl Acad Sci USA 109:20808–20813

Evans E, Skalak P (1980) Mechanics and thermodynamics of biomembranes. CRC Press, Boca Raton

Fai TG, Griffith BE, Mori Y, Peskin CS (2013) Immersed boundary method for variable viscosity and variable density problems using fast constant-coefficient linear solvers I: Numerical method and results. SIAM J Sci Comput 35(5):B1132–B1161

Fedosov DA, Caswell B, Karniadakis GE (2010) A multiscale red blood cell model with accurate mechanics, rheology, and dynamics. Biophys J 98:2215–2225

Fedosov DA, Caswell B, Karniadakis GE (2011a) Wall shear stress-based model for adhesive dynamics of red blood cells in malaria. Biophys J 100:2084–2093

Fedosov DA, Lei H, Caswell B, Suresh S, Karniadakis GE (2011b) Multiscale modeling of red blood cell mechanics and blood flow in malaria. PLoS Comput Biol 7:e1002270

Fedosov DA, Peltomaki M, Gompper G (2014) Deformation and dynamics of red blood cells in flow through cylindrical microchannels. Soft Matter 10:4258–4267

Fischer TM (1992) Is the surface area of the red cell membrane skeleton locally conserved? Biophys J 61:298–305

Fischer TM, Stohr-Liesen M, Schmid-Schonbein H (1978) The red cell as a fluid droplet: tank tread-like motion of the human erythrocyte membrane in shear flow. Science 202:894–896

Freund JB (2014) Numerical simulation of flowing blood cells. Annu Rev Fluid Mech 46(1):67–95

Freund JB (2013) The flow of red blood cells through a narrow spleen-like slit. Phys Fluids 25:110807

Hao W, Xu Z, Liu C, Lin G (2015) A fictitious domain method with a hybrid cell model for simulating motion of cells in fluid flow. J Comput Phys 280:345–362

Hosseini S, Feng J (2012) How malaria parasites reduce the deformability of infected red blood cells. Biophy J 103:1–10

Kapitza H, Rupped D, Galla H, Sackmann E (1984) Lateral diffusion of lipids and glycophorin in solid phosphatidylcholine bilayers. The role of structural defects. Biophys J 45:577

Knowles DW, Tilley L, Mohandas N, Chasis JA (1997) Erythrocyte membrane vesiculation: model for the molecular mechanism of protein sorting. Proc Natl Acad Sci USA 94:12969–12974

Kodippili G, Spector J, Sullivan C, Kuypers F, Labotka R, Gallagher P, Ritchie K, Low P (2009) Imaging of the diffusion of single band 3 molecules on normal and mutant erythrocytes. Blood 113:6237

Lac E, Barthes-Biesel D, Pelekasis N, Tsamopoulos J (2004) Spherical capsules in three-dimensional unbounded stokes flows: effect of the membrane constitutive law and onset of buckling. J Fluid Mech 516:303–334

Lee JC, Discher DE (2001) Deformation-enhanced fluctuations in the red cell skeleton with theoretical relations to elasticity, connectivity, and spectrin unfolding. Biophys J 81:3178–3192

Li H, Lykotrafitis G (2012) Two-component coarse-grained molecular-dynamics model for the human erythrocyte membrane. Biophys J 102:75–84

Li H, Lykotrafitis G (2014) Erythrocyte membrane model with explicit description of the lipid bilayer and the spectrin network. Biophys J 107:642–653

Li H, Lykotrafitis G (2015) Vesiculation of healthy and defective red blood cells. Phys Rev E 92(1):012715

Li H, Zhang Y, Ha V, Lykotrafitis G (2016) Modeling of band-3 protein diffusion in the normal and defective red blood cell membrane. Soft Matter 12(15):3643–3653

Li H, Lu L, Li X, Buffet P, Dao M, Karniadakis GE, Suresh S (2018) Mechanics of diseased red blood cells in human spleen and consequences for hereditary blood disorders. Proc Natl Acad Sci USA 115:9574–9579

Li J, Dao M, Lim CT, Suresh S (2005) Spectrin-level modeling of the cytoskeleton and optical tweezers stretching of the erythrocyte. Biophys J 88:3707–3719

Li J, Lykotrafitis G, Dao M, Suresh S (2007) Cytoskeletal dynamics of human erythrocyte. Proc Natl Acad Sci USA 104:4937–4942

Li X, Vlahovska PV, Karniadakis GE (2013) Continuum- and particle-based modeling of shapes and dynamics of red blood cells in health and disease. Soft Matter 9:28–37

Li X, Peng Z, Lei H, Dao M, Karniadakis GE (2014) Probing red blood cell mechanics, rheology and dynamics with a two-component multi-scale model. Philos Trans A Math Phys Eng Sci 372:20130389

Li X, Dao M, Lykotrafitis G, Karniadakis GE (2017a) Biomechanics and biorheology of red blood cells in sickle cell anemia. J Biomech 50:34–41

Li X, Li H, Chang HY, Lykotrafitis G, Karniadakis GE (2017b) Computational biomechanics of human red blood cells in hematological disorders. ASME J Biomech Eng 139:021008

Malone JG, Johnson NL (1994) A parallel finite-element contact/impact algorithm for nonlinear explicit transient analysis 1. the search algorithm and contact mechanics. Int J Numer Methods Eng 37:559–590

McWhirter JL, Noguchi H, Gompper G (2009) Flow-induced clustering and alignment of vesicles and red blood cells in microcapillaries. Proc Natl Acad Sci USA 106(15):6039–6043

Müller K, Fedosov D, Gompper G (2015) Smoothed dissipative particle dynamics with angular momentum conservation. J Comput Phys 281:301–315

Ni A, Cheema TA, Park CW (2015) Numerical study of RBC motion and deformation through microcapillary in alcohol plasma solution. Open J Fluid Dyn 05(01):26–33

Noguchi H, Gompper G (2005) Shape transitions of fluid vesicles and red blood cells in capillary flows. Proc Natl Acad Sci USA 102:14159–14164

Pan W, Caswell B, Karniadakis GE (2010) Rheology, microstructure and migration in brownian colloidal suspensions. Langmuir 26(1):133–142

Peng Z, Zhu Q (2013) Deformation of the erythrocyte cytoskeleton in tank treading motions. Soft Matter 9:7617–7627

Peng Z, Asaro R, Zhu Q (2010) Multiscale modeling of erythrocyte membranes. Phys Rev E 81:031904

Peng Z, Asaro R, Zhu Q (2011) Multiscale modelling of erythrocytes in stokes flow. J Fluid Mech 686:299–337

Peng Z, Li X, Pivkin IV, Dao M, Karniadakis GE, Suresh S (2013) Lipid bilayer and cytoskeletal interactions in a red blood cell. Proc Natl Acad Sci USA 110:13356–13361

Peng Z, Mashayekh A, Zhu Q (2014) Erythrocyte responses in low shear rate flows – effects of non-biconcave stress-free state in cytoskeleton. J Fluid Mech 742:96–118

Peng Z, Salehyar S, Zhu Q (2015) Stability of the tank treading modes of erythrocytes and its dependence on cytoskeleton reference states. J Fluid Mech 771:449–467

Peskin CS (2002) The immersed boundary method. Acta Numer 11:479–517

Pivkin IV, Karniadakis GE (2008) Accurate coarse-grained modeling of red blood cells. Phys Rev Lett 101(11):118105

Pozrikidis C (1992) Boundary integral and singularity methods for linearized viscous flow. Cambridge University Press, Cambridge

Pozrikidis C (2010) Flow-induced deformation of two-dimensional biconcave capsules. In: Computational hydrodynamics of capsules and biological cells. CRC Press, Boca Raton, pp 1–33

Ramanujan S, Pozrikidis C (1998) Deformation of liquid capsules enclosed by elastic membranes in simple shear flow: large deformations and the effect of fluid viscosities. J Fluid Mech 361:117–143

Reasor DA, Clausen JR, Aidun CK (2012) Coupling the lattice-Boltzmann and spectrin-link methods for the direct numerical simulation of cellular blood flow. Int J Numer Methods Fluids 68:767–781

Rief M, Pascual J, Saraste M, Gaub HE (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. J Mol Biol 286:553–561

Shi L, Pan TW, Glowinski R (2014) Three-dimensional numerical simulation of red blood cell motion in poiseuille flows. Int J Numer Methods Fluids 76(7):397–415

Sui Y, Low H, Chew Y, Roy P (2008) Tank-treading, swinging, and tumbling of liquid-filled elastic capsules in shear flow. Phys Rev E 77(1):016310

Sung LA, Vera C (2003) Protofilament and hexagon: a three-dimensional mechanical model for the junctional complex in the RBC membrane skeleton. Ann Biomed Eng 31:1314–1326

Tang YH, Lu L, Li H, Evangelinos C, Grinberg L, Sachdeva V, Karniadakis GE (2017) Openrbc: a fast simulator of red blood cells at protein resolution. Biophys J 112(10):2030–2037

Tomishige M (1998) Regulation mechanism of the lateral diffusion of band 3 in erythrocyte membranes by the membrane skeleton. J Cell Biol 142:989–1000

Tran-Son-Tay R, Sutera S, Rao P (1984) Determination of red blood cell membrane viscosity from rheoscopic observations of tank-treading motion. Biophys J 46:65–72

Van Liedekerke P, Odenthal T, Smeets B, Ramon H (2013) Solving microscopic flow problems using stokes equations in sph. Comput Phys Commun 184:1686–1696

Veerapaneni SK, Rahimian A, Biros G, Zorin D (2011) A fast algorithm for simulating vesicle flows in three dimensions. J Comput Phy 230(14):5610–5634

Vlahovska P, Young Y, Danker G, Misbah C (2011) Dynamics of a non-spherical microcapsule with incompressible interface in shear flow. J Fluid Mech 678:221–247

Walensky LD, Mohandas N, Lux SE (2003) Disorders of the red blood cell membrane. In: Handin RI, Lux SE, Stossel TP (eds) Blood: principles and practice of hematology, 2nd edn. Lippincott Williams & Wilkins, Philadelphia, pp 1709–1858

Walter J, Salsac A, Barthès-Biesel D, Tallec PL (2010) Coupling of finite element and boundary integral methods for a capsule in a Stokes flow. Int J Numer Methods Eng 83:829–850

Witthoft A, Yazdani A, Peng Z, Bellini C, Humphrey JD, Karniadakis GE (2016) A discrete mesoscopic particle model of the mechanics of a multi-constituent arterial wall. J R Soc Interface 13:20150964

Yazdani A, Bagchi P (2011) Phase diagram and breathing dynamics of a single red blood cell and a biconcave capsule in dilute shear flow. Phys Rev E 84:026314

Yazdani A, Kalluri R, Bagchi P (2011) Tank-treading and tumbling frequencies of capsules and red blood cells. Phys Rev E 83:046305

Zhang, JF, Johnson PC, Popel AS (2008) Red blood cell aggregation and dissociation in shear flows simulated by lattice Boltzmann method. J Biomech 41:47–55

Zhang Y, Huang C, Kim S, Golkaram M, Dixon MW, Tilley L, Li J, Zhang S, Suresh S (2015) Multiple stiffening effects of nanoscale knobs on human red blood cells infected with plasmodium falciparum malaria parasite. Proc Natl Acad Sci USA 112(19): 6068–6073

Zhao H, Isfahania AHG, Olsonc LN, Freund JB (2010) A spectral boundary integral method for flowing blood cells. J Comput Phys 229:3726–3744

Zhu Q, Asaro R (2008) Spectrin folding vs. unfolding reactions and RBC membrane stiffness. Biophys J 94:2529–2545

Zhu Q, Vera C, Asaro R, Sche P, Sung L (2007) A hybrid model for erythrocyte membrane: a single unit of protein network coupled with lipid bilayer. Biophys J 93:386–400

Computational Models of Eukaryotic Cells in Health and Disease

108

Kirill Lykov and Igor V. Pivkin

Contents

Abstract

Eukaryotic cells play a crucial role in many processes in health and disease in the human body. The importance of an understanding of cell behavior from the mechanical point of view triggered the development of many computational models. In this chapter, we present recent progress in this field. We focus primarily on computational models of suspended in flow and adherent to substrate eukaryotic cells.

1 Introduction

Cells are building blocks of living beings, and there is a wide variety of cell types each tailored for a specific purpose in the body. Cells composed of the cell membrane, complex 3D cytoskeletal network, and various organelles, including the nucleus, are called eukaryotic cells. These cells play roles in many processes of a mechanical nature in health and disease, e.g., cell migration, wound healing, and

K. Lykov (✉) · I. V. Pivkin
Faculty of Informatics, Institute of Computational Science, Universita della Svizzera Italiana, Lugano, Switzerland
e-mail: kirill.lykov@usi.ch; igor.pivkin@usi.ch

cancer metastasis. The mechanics of the eukaryotic cell as a whole is determined by the properties of its subcellular components.

The plasma membrane separates the body of the cell from the extracellular environment. It is composed of different lipids and cholesterol with the inclusion of various embedded proteins. The volume inside the cell membrane is filled with the protein-rich fluid, called cytosol. Although the plasma membrane has a mechanism for the fluid and solute exchange with the surrounding, it is mostly impermeable, which leads to volume conservation of a cell. Incompressibility of the lipid layers results in plasma membrane surface area conservation and bending resistance.

The space between the plasma membrane and cell organelles is occupied, besides of cytosol, by a complex filament network called cytoskeleton. It is a dynamic structure, which maintains the integrity of the cell. In addition, it plays an essential role in many cell processes: migration, adhesion, division, stress resistance, and mechanotransduction. The cytoskeleton is composed of interconnected filaments of three main types: actin filaments, microtubules, and intermediate filaments (Suresh 2007). Actin filaments are semiflexible biopolymers, which exhibit the highest resistance to deformations until some critical stress value. If the stress is higher than this value, they are fluidized. Intermediate filaments can resist moderate deformations, and they do not fluidized under high values of shear stress, hence, providing structural integrity to the cell. Microtubules do not contribute significantly to the cell stiffness. Being interconnected with filaments of other types, they stabilize the cytoskeleton. They also might contribute to the compression resistance under high pressure (Brangwynne et al. 2006).

The nucleus is the largest eukaryotic organelle, and it is stiffer than the cell itself (Guilak et al. 2000; Friedl et al. 2011). Mechanics-wise, the most important components of the nucleus are nucleus envelope and chromatin network. The nucleus envelope is composed of two phospholipid bilayers with attached lamin meshwork. These bilayers act as a barrier between cytoplasm and nucleus internal structure (Denais and Lammerding 2014). In comparison to the cell membrane, nucleus membrane exposes weaker area and volume constraints allowing fluid to get in and out (Rowat et al. 2006). The nuclear lamina, which is often considered to be the main contributor to the nucleus stiffness (Gerlitz and Bustin 2012), is 2D meshwork attached to the inner bilayer. According to a recent study, lamina levels control nuclear strain stiffening at large extensions (Stephens et al. 2017). Contrary, chromatin network, which occupies the nucleus volume, governs response to small deformations (Stephens et al. 2017).

Mechanical properties of eukaryotic cells depend on the cell type as well as the state of the cell. Cells can be in suspended state, when they are traveling in the blood stream, or in the adherent state, when they are attached to the extracellular matrix. Another source of alternations in cell structure and mechanics is various diseases (Nematbakhsh and Lim 2015), which might alter the properties of any of the cell components. One example is softening of the cell membrane by HIV before invasion into the cell by the insertion of fusion peptide (Agrawal et al. 2016). Due to the dominant role of cytoskeleton in the cell mechanics, it is not surprising that the progression of numerous diseases leads to the cytoskeleton properties alteration.

The most prominent example is the cytoskeleton structure degradation and density decrease of epithelial tumor cells (Suresh 2007). In several cancers, the higher is the metastatic potential of a cell, the lower is the stiffness of the cytoskeleton and, hence, the cell as a whole (Nematbakhsh et al. 2017; Cross et al. 2007; Liu et al. 2015b). Other diseases which change the cytoskeleton mechanics are neurodegeneration, liver cirrhosis, pulmonary fibrosis, and blistering skin diseases (Ramaekers and Bosman 2004). Some tissue-specific disorders are linked to the alteration of the mechanical properties of the nucleus due to the mutations in genes encoding lamins and associated nuclear envelope proteins: Emery-Dreifuss muscular dystrophy, dilated cardiomyopathy, familial partial lipodystrophy, as well as over 20 other diseases (Rowat et al. 2006; Rynearson and Sussman 2011).

The great variety of cell types as well as diseases altering cell mechanics led to the development of numerous experimental techniques for investigating cell mechanical properties. These techniques can be split into two groups. The methods in the first group, such as atomic force microscopy (AFM) and micropipette aspiration, operate with individual cells to estimate their viscoelastic properties. A classical review of methods of this type can be found elsewhere (Suresh 2007). The second group of experimental techniques with higher throughput includes microfluidic experiments. There is a great variety of microfluidic devices and principles they employ to operate. A good review of the current status of the microfluidics with application to cells can be found elsewhere (Chaudhuri et al. 2016).

2   Modeling Approaches

Despite a significant progress in the experimental techniques in cell biophysics, the computational modeling remains an irreplaceable complementing tool. Firstly, it improves our understanding of the subcellular component mechanics, which is valuable for many applications. In particular, it can influence the drug discovery by offering a rich target space for possible alteration of cell properties. Secondly, numerical simulations allow to predict cell behavior in a complex environment, and thus, their use may improve and accelerate the design of various microfluidic devices (Rossinelli et al. 2015). In particular, numerical simulations can help in the acquisition of information related to fluid microrheology, including often unexpected flow properties in complex domains resulting in various cell phenomena (merging, lysis, adhesion, damage, signaling) (Nan et al. 2014; Yung et al. 2009; Pinho et al. 2013).

Recent advances in computational methods, as well as hardware development, catalyzed the creation of a number of eukaryotic cell models in the last 5 years. The variety of models can be explained by the complexity of the problem. The first source of this complexity is the diversity of the cell types along with processes they are involved in. The second one is the lack of understanding of the role that each cell component plays during mechanical deformations. Finally, there is no full understanding of the active cell response. These factors explain the fact that the

current state-of-the-art models are phenomenological, and each one is tailored to describe a very particular cell line and related processes.

The computational methods employed in eukaryotic cell modeling can be split into three groups: mesh-based, particle-based, and combined. Among mesh-based methods, the most popular one is finite element method (FEM). Typically, models exploiting FEM describe only cell mechanics without considering surrounding fluid. Dissipative particle dynamics (DPD) is a particle-based method, which is also used in many models. DPD allows parametrizing the fluid and fluid-structure interactions, while the cell itself is modeled with the help of various many-body potentials. Among combined methods, a well-developed approach is to model the fluid with the Lattice-Boltzmann (LB) method, cells with finite element method (FE), and cell-fluid interactions using immersed boundary (IB) method. We will call this approach LB/IB.

The focus of this review is mechanical eukaryotic cell models. We do not consider cell models used for the tissue modeling; a review of these models can be found elsewhere (Fletcher et al. 2017). The models for cells without nucleus such as red blood cells (RBCs) are also beyond the scope of this review. Overview of recent developments in this field can be found in other chapters of this book. Further, we split the existing eukaryotic cell models into two groups. The first group contains models for the suspended cells which are usually employed to study fluid flow problems. The second group is dedicated to the modeling of the cells in the adherent state.

2.1   Suspended Cell Models

Suspended cells can be of different origin: most of them are blood cells (RBCs, platelets, white blood cells (WBCs)) and mesenchymal stem cells; others are cells of epithelial origin traveling with the blood stream such as circulating tumor cells (CTCs). CTCs are emitted by the primary cancer tumor and can travel long distances to create secondary tumors. Both WBCs and CTCs are bigger and stiffer than RBCs. This leads to the so-called margination, which is a process whereby these cells are displaced to the vessel wall (Nematbakhsh and Lim 2015). An important problem is the detection of CTCs in the blood sample, and numerous microfluidic devices have been developed for this purpose. Devices employing margination, cell rolling, hydrodynamics-induced cell sorting are all studied using models considered in this section.

Perhaps one of the simplest approaches to modeling of the eukaryotic cell is to ignore its internal structure and to apply one of the existing RBC models with altered parameters. The reasoning is that RBC and eukaryotic cell membranes are similar in many aspects and there are several well-established and validated RBC models (Ye et al. 2016). These models usually take into account volume and area conservation, bending resistance, and membrane viscoelastic forces. Such models, coupled with various fluid solvers, capture the effects of cell size and shape and, hence, are often used in simulations of microfluidic devices employing deterministic

lateral displacement principle or cell-free layer formation. The main strength of these models is low computational cost and resolved hydrodynamic interactions in a flow.

In a series of papers, Takeishi et al. modeled the margination of WBCs and CTCs and, in a later work, adhesion of CTCs to vessel walls (Takeishi et al. 2014, 2015) (see Fig. 1a). This development allowed the authors to numerically investigate the effect of the cell size on the flow mode and cell velocity, as well as to identify similarities and differences between WBCs and CTCs. Further, it was found that the bullet motion enables firm adhesion of a cell to the capillary wall (Takeishi et al. 2016). These results suggest that even under the interaction between proteins responsible for WBC rolling, a cell can show firm adhesion in a small capillary. Modeling-wise, Takeishi et al. employed LB/IB method and empty shell cell representation. For validation of the model, simulation results for the deformation of a spherical cell in shear flow were compared with previously published works.

Fedosov et al. also studied WBCs margination, but they, additionally, took into account the effect of RBCs (Fedosov and Gompper 2014) (see Fig. 1b). It was found that WBC margination occurs mainly within a region of intermediate hematocrits and for relatively low flow rates. Moreover, simulations showed that RBC aggregation slightly enhances WBC margination, particularly at the high hematocrit values. DPD method was used to describe fluid and cell particles.

Later, Rossinelli et al. employed a similar DPD-based empty shell model to describe both WBCs and CTCs in two microfluidic devices (Rossinelli et al. 2015). Snapshot of cells squeezing between obstacles mimicking the geometry of the microfluidic devices is shown in Fig. 1c. They demonstrated the applicability of this model for the predictive studies of the performance of microfluidic devices. In particular, the simulation results reproduced the experimentally observed effect of deterministic lateral displacement of bigger cells for the whole blood passing through the array of cylindrical obstacles.

The work by Xiao et al. is dedicated to the mutual effect of rolling CTC in the capillary on the blood flow dynamics (Xiao and Fu 2016a) (see Fig. 1d). By the help of computational modeling of suspended CTC and RBCs, it was demonstrated that, in the microvessel of $15\,\mu m$ diameter, the CTC has an increased probability of adhesion due to a growing wall-directed force. However, with the increase in microvessel size, an enhanced lift force at higher hematocrit detaches the adherent CTC quickly. An increased blood flow resistance in the presence of CTC was also found. Moreover, the significant deformation induced by high flow rate and the presence of aggregation promote the adhesion of CTC. In another work, the same group of authors applied a very similar cell model to investigate an individual cell passing through a narrow slit (Xiao and Fu 2016b). Specifically, they studied the effect of cell size, nucleus, and cell membrane shape on the transmigration through a slit. This model represents the cell and nucleus as an empty shell.

Zhang et al. studied CTC and WBC passing through micro-filtering channels of different shapes (Zhang et al. 2014). In particular, they considered channels with circular, rectangular, and triangular cross sections and inspected pressure signatures
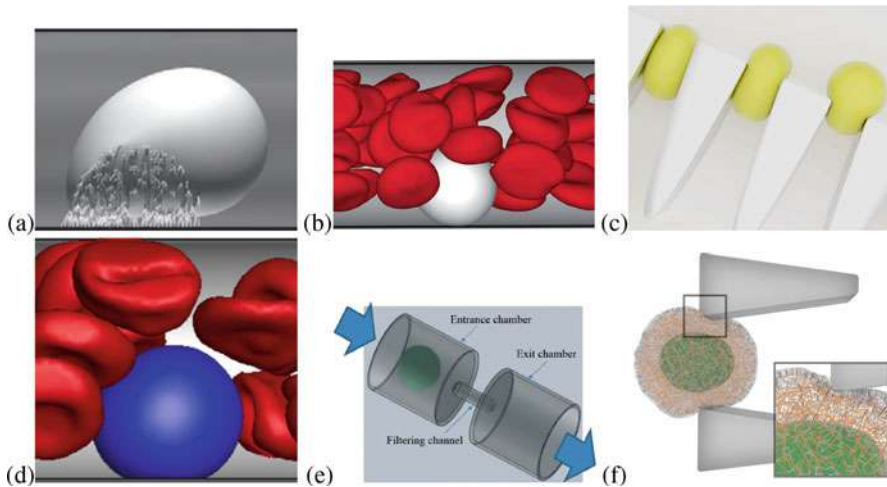
Fig. 1 (**a**) Simulation of a cell adherent to the tube wall (Takeishi et al. 2016). (**b**) Simulation demonstrating WBC margination in the RBC flow (Fedosov and Gompper 2014). (**c**) Cell models squeezing between obstacles in microfluidic device (Rossinelli et al. 2015). (**d**) Snapshot from the simulation of the cell adherent to the wall (blue) together with RBCs (Xiao and Fu 2016a). (**e**) Sketch of a cell passing through a micro-filtering circular channel (Zhang et al. 2014). (**f**) Multicomponent model of eukaryotic cell passing between two constrictions of microfluidic device (Lykov et al. 2017). Nucleus shown in green, cytoskeleton filaments in orange

for CTCs and WBCs passing through this channel. They demonstrated that circular channels feature the highest critical pressure and, thus, are more suitable for cell-separation microfluidic devices (see Fig. 1e). The developed model exploits volume of fluid method, and no subcellular components were explicitly modeled.

Finally, in recent combined experimental and modeling work, Lykov et al. investigated cell mechanics in micropipette aspiration and microfluidic experiments (Lykov et al. 2017) (see Fig. 1f). The mesoscale particle-based model of the eukaryotic cell consisted of a membrane, cytoskeleton, and nucleus. The fluid was discretized using DPD method. The calibration of the model was done using micropipette aspiration experimental data. The model was validated using microfluidic experiments. Authors studied the role of subcellular components and observed that cell cytoskeleton filament density and cross-link concentration could both significantly affect cell resistance to stress. Other results include quantification of the nucleus components contribution and effect of cell viscosity during cell traversal in the microfluidic devices.

Although considered models of suspended cells use different discretization methods, most of them employ a simple empty shell cell representation. This approach can be applied to study flow phenomena in which cell size and shape play a key role. To study a broader class of phenomena, which include severe flow-induced cell deformations, internal structure disruption, and effect of changes in the cytoskeleton due to drug treatment, models that explicitly describe internal cell

components are necessary. The drawback of such models is that they require more complicated parametrization procedure. Increase in computational cost, however, is modest, since it is usually dominated by the cost of flow discretization.

2.2   Adherent Cell Models

The interest in studying processes, which imply significant deformation of the adherent cell and, thus, involve mechanical response of the internal cell components, led to the development of approaches which explicitly model the nucleus and cytoskeleton. Typically, these models use the same hollow sphere model for the cell membrane and, sometimes, the nucleus. The cytoskeleton is modeled by various approaches. These models opened the route to the numerical investigation of such processes as needle microinjection, AFM indentation, cell migration, and deformation in microfluidic devices.

Kardas et al. proposed a computational approach to model the structure of bone cells (Kardas et al. 2013). The model takes into account integrins, nucleus, centrosome, and cytoskeletal proteins (see Fig. 2a). It was shown that the load acting on the nucleus is rising with increasing deformation applied to the integrins. The numerical simulations demonstrated that the nucleus is more affected by stress if the distributions of intermediate filaments and microtubules are random than if they are regular. Computational-wise, FEM was employed to discretize governing equations.

Ujihara et al. presented a cell model to study tensile and compression tests (Ujihara et al. 2010, 2012). The model incorporates a cell membrane, a nuclear envelope, and actin filaments (see Fig. 2b). During the tensile test, it was observed that the total elastic energy of the model is dominated by actin fibers contribution. The compression test revealed that the alignment of bundles of actin filaments significantly affects cell stiffness. In addition, the passive reorientation of actin filament bundles perpendicular to the direction of compression induced an increase in the resistance to the elongation of a cell and, thereby, increased the cell stiffness. Particle-based model was used in simulations utilizing minimum energy concept. It was validated by comparing load-deformation curve with experimental data.

Dowling et al. presented a study of the role of the active remodeling and contractility of the actin cytoskeleton in the response of chondrocytes (cartilage cells) to shear (Dowling et al. 2012) (see Fig. 2c). The key feature of the model is that it incorporates both passive viscoelastic component and active, describing cytoskeleton remodeling. By the help of numerical simulations, the authors showed that a purely passive cell model is incapable of predicting the response of normal chondrocytes to the stress, while adding the cytoskeleton remodeling gives results close to the in vitro study. Interestingly, the passive model can predict drug-treated cells with the disrupted cytoskeleton, which might be considered as evidence that actin cytoskeletal network is the main contributor to the stress resistance during discussed processes. In another paper, where a similar model was used, authors predicted the increased compressive resistance of spread cells compared with round
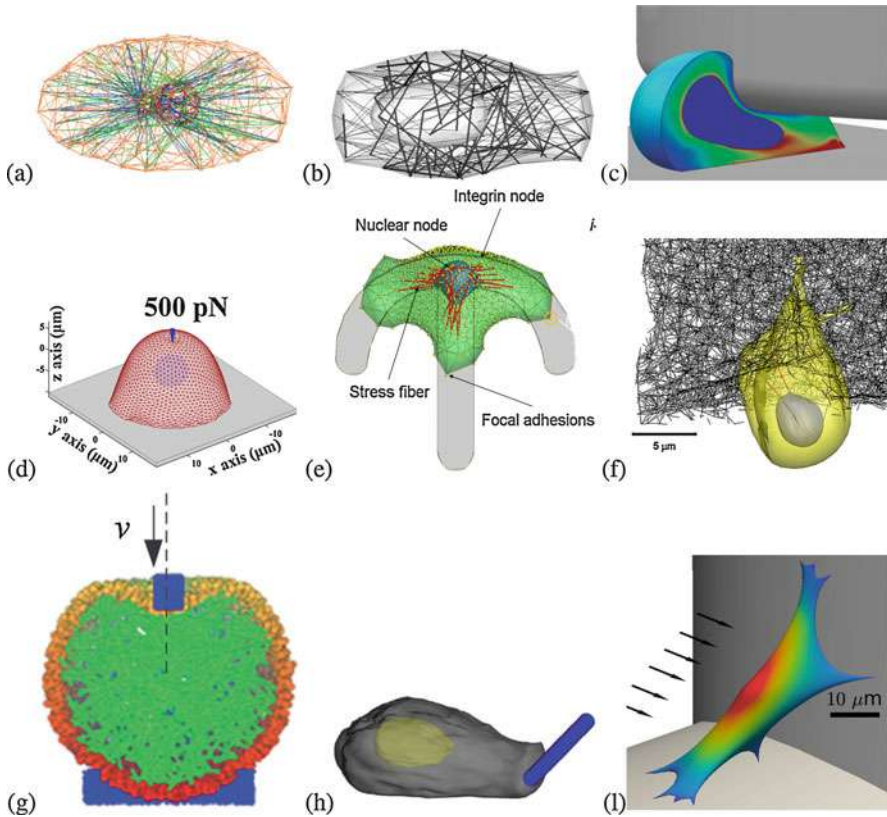
Fig. 2 (**a**) Cell model taking into account integrins, nucleus, centrosome, and cytoskeletal protein (Kardas et al. 2013). (**b**) Snapshot from simulation of the tensile test (Ujihara et al. 2010). (**c**) Simulation of the probe indentation process; color map depicts average stress fiber activation levels (Dowling et al. 2012). (**d**) Modeling of the AFM nanoindentation process (Fang and Lai 2016). (**e**) Integrated cell migration model consisting of the cytoskeleton, the nucleus, and cytoskeleton components (Kim et al. 2013). (**f**) Simulated cell invasion into ECM fiber network model (Kim et al. 2015). (**g**) Simulated cell microinjection: green particles represent cytoskeleton, blue rectangular is needle tip, and yellow-orange particles represent cell membrane. (**h**) Deformation of the plasma and nuclear membranes induced by micropipette pulling (Zeng et al. 2012). (**d**) Snapshot from AFM experiment simulation for adherent cell (Fang and Lai 2016). (**i**) Snapshot from simulation of the cell attached to the walls of the bioreactor; the flow direction is shown with arrows; the shear stress distribution is depicted with color (Guyot et al. 2016)

cells (Ronan et al. 2012). The nucleus and membrane were represented in the model as a passive hyperelastic material, while ODE described the cytoskeleton remodeling. The equations were discretized with the help of FEM.

Fang and Lai studied changes in the cell mechanics as cell shifts from suspended to the adherent state (Fang and Lai 2016). The model consists of cell membrane, nucleus envelope, and three internal networks representing microtubules, F-actin, and intermediate filaments. In addition, it takes into account the movement of

adhesion molecules which allows simulating cell spreading. In order to estimate the cell model elastic modulus, simulation of AFM experiment was employed (see Fig. 2d). The force-indentation relationship was used to determine the mechanical changes in cells during state shift. The explicit modeling of the subcellular components allowed to investigate the impact of different cell components on the resistance to the external stress and to examine the effect of nucleus presence on the AFM results. The model was validated by comparing the AFM simulation results with the experimental data.

Kim et al. developed a model to predict cell migration behavior on 2D and 3D curved surfaces (Kim et al. 2012). The model takes into account focal adhesion dynamics, actin motor activity, as well as cytoskeleton remodeling and nucleus (see Fig. 2e). The simulations revealed that cell migration speed depends on the cross-sectional area of the microfluidic channel. The relationship between migration speed and the channel width agreed with the experimental data. This model was also applied to study cell migration on 2D micropatterned geometries (Kim et al. 2013). The further development of the model allowed prediction of cell invasion into a 3D extracellular matrix (ECM) in response to different extracellular biochemical cues (Kim et al. 2015) (see Fig. 2f). The new model takes into account filopodia penetration dynamics. The average filopodia speed was predicted, and cell membrane advance velocity agreed with experiments of 3D HUVEC migration for diverse ECMs with different pore sizes and stiffness.

By the help of numerical modeling, Liu et al. predicted the cell damage induced by the needle during the microinjection procedure (Liu et al. 2015a). In particular, authors quantified the effects of the size, shape of the microinjector tip, and injection velocity on cell damage. The proposed model is based on DPD method. It explicitly describes cell membrane as well as the cytoskeleton and motor activity. To validate the model, authors measured the mechanical properties of the model using the particle-tracking microrheology. The cell model showed power law behavior in terms of mean square displacement and lag time. The mechanical moduli obtained from the simulations were in agreement with the experimental data.

Zeng et al. applied cell model to study the role of the actin cytoskeleton network in mechanotransduction and nucleus deformation (Zeng et al. 2012). In the experiment simulated in this work, a micropipette is pushed into the cytoplasm of an endothelial cell and then pulled away at a constant rate. Authors showed that the stress propagation through the random cytoskeletal network could be a mechanism to effect nucleus deformation, without invoking any biochemical signaling activity. It was reported that nucleus strain varies in a sigmoidal manner with actin filament concentration, while there exists an optimal concentration of actin-binding proteins that maximize nucleus displacement. In addition, a theoretical analysis for these nonlinearities in terms of the connectivity of the random cytoskeletal network was provided. Particle-based model was employed, which explicitly describes cell membrane, nucleus envelope, and cytoskeletal actin network. The simulation results were validated using experimental data.

Guyot et al. modeled an adherent cell in the bioreactor subjected to the shear stress due to the fluid flow (Guyot et al. 2016). The simulations were used to estimate

shear stress experienced by cells as a function of the bioreactor inlet flow velocity. The authors showed that the shear stress values predicted by their model are different from shear stress obtained with the help of empty scaffolds model, which is often used as a reference. The proposed model can be applied to optimize the bioreactor geometry and flow conditions as well as to provide insight into the cell deformation during these processes. With the help of LB/IB method, the cell was simulated as an empty shell with the elastic nucleus inside. To estimate elastic properties of the cell, micropipette aspiration experiments were used.

The most important common feature of the models considered in this section is that they take into account, besides of nucleus envelope and cell membrane, mechanics of the cytoskeleton. Due to that, such models can describe more accurately changes in the cell during different mechanical tests as well as cytoskeleton-related processes. Additionally, these models give an insight into the impact of subcellular components during particular experiments.

3   Conclusion

The range of applications of eukaryotic cell models is expanding rapidly. Suspended cell models are now widely used in studies of behavior and deformations of cells in various microfluidic devices and blood circulation. In turn, adherent cell models provide an insight into cell mechanics when cells are subjected to mechanical stress or during migration in the extracellular matrix. Majority of suspended cell modeling approaches employ empty shell representation of the cell, which narrows the applicability of these models to problems with moderate cell deformations. The drawback of the existing adherent cell models is their specialization on a very particular phenomenon for a specific cell type.

All of the models considered in this review follow the top-down or phenomenological approach (Kollmannsberger and Fabry 2011). The mechanical properties of the cell model components are not obtained by considering the properties of the material at the microscopic or molecular level. Instead, the choice of model and parametrization for the subcellular components is based on general principles and macroscopic properties of the cell. An alternative bottom-up or reductionist method assumes gradual development of the model starting from the very accurate models of all constituents. At the very moment, there is a lack of experimental and theoretical knowledge to follow the latter approach. Future developments will allow construction of eukaryotic cell models by rigorously linking mechanical behavior of the entire cell with the properties of its molecular constituents, connecting cell chemistry and mechanobiology.

References

Agrawal H, Zelisko M, Liu L, Sharma P (2016) Rigid proteins and softening of biological membranes with application to HIV-induced cell membrane softening. Sci Rep 6:25412. https://doi.org/10.1038/srep25412, http://www.nature.com/articles/srep25412

Brangwynne C, MacKintosh F, Kumar S, Geisse N, Talbot J, Mahadevan L, Parker K, Ingber D, Weitz D (2006) Microtubules can bear enhanced compressive loads in living cells because of lateral reinforcement. J Cell Biol (173):733–741

Chaudhuri PK, Ebrahimi Warkiani M, Jing T, Kenry, Lim CT (2016) Microfluidics for research and applications in oncology. Analyst 141:504–524. https://doi.org/10.1039/C5AN00382B

Cross SE, Yu-Sheng J, Jianyu R, Gimzewski JK (2007) Nanomechanical analysis of cells from cancer patients. Nat Nanotechnol 2(12):780–783. https://doi.org/10.1038/nnano.2007.388, http://search.ebscohost.com/login.aspx?direct=true%26db=a9h%25%26AN=32097033%26site=ehost-live

Denais C, Lammerding J (2014) Nuclear mechanics in cancer. Adv Exp Med Biol 773:435–470. https://doi.org/10.1007/978-1-4899-8032-8_20, http://link.springer.com/10.1007/978-1-4899-8032-8

Dowling EP, Ronan W, Ofek G, Deshpande VS, McMeeking RM, Athanasiou Ka, McGarry JP (2012) The effect of remodelling and contractility of the actin cytoskeleton on the shear resistance of single cells: a computational and experimental investigation. J R Soc Interface 9:3469–3479. https://doi.org/10.1098/rsif.2012.0428

Fang Y, Lai KWC (2016) Modeling the mechanics of cells in the cell-spreading process driven by traction forces. Phys Rev E 93:1–15. https://doi.org/10.1103/PhysRevE.93.042404

Fedosov Da, Gompper G (2014) White blood cell margination in microcirculation. Soft matter 10:2961–70. https://doi.org/10.1039/c3sm52860j, http://www.ncbi.nlm.nih.gov/pubmed/24695813

Fletcher AG, Cooper F, Baker RE, Fletcher AG (2017) Mechanocellular models of epithelial morphogenesis. Phil Trans R Soc B (372). https://doi.org/10.1098/rstb.2015.0519

Friedl P, Wolf K, Lammerding J (2011) Nuclear mechanics during cell migration. Curr Opin Cell Biol 23(1):55–64. https://doi.org/10.1016/j.ceb.2010.10.015, http://www.sciencedirect.com/science/article/pii/S0955067410001869, cell structure and dynamics

Gerlitz G, Bustin M (2012) The role of chromatin structure in cell migration. Trends Cell Biol 21(1):6–11. https://doi.org/10.1016/j.tcb.2010.09.002

Guilak F, Tedrow JR, Burgkart R (2000) Viscoelastic properties of the cell nucleus. Biochem Biophys Res Commun 269:781–786. https://doi.org/10.1006/bbrc.2000.2360

Guyot Y, Smeets B, Odenthal T, Subramani R, Luyten FP, Ramon H, Papantoniou I, Geris L (2016) Immersed boundary models for quantifying flow-induced mechanical stimuli on stem cells seeded on 3D scaffolds in perfusion bioreactors. PLoS Comput Biol 12(9):1–21. https://doi.org/10.1371/journal.pcbi.1005108

Kardas D, Nackenhorst U, Balzani D (2013) Computational model for the cell-mechanical response of the osteocyte cytoskeleton based on self-stabilizing tensegrity structures. Biomech Model Mechanobiol 12(1):167–183. https://doi.org/10.1007/s10237-012-0390-y

Kim MC, Kim C, Wood L, Neal D, Kamm RD, Asada HH (2012) Integrating focal adhesion dynamics, cytoskeleton remodeling, and actin motor activity for predicting cell migration on 3D curved surfaces of the extracellular matrix. Integr Biol 4(11):1386. https://doi.org/10.1039/c2ib20159c

Kim MC, Neal DM, Kamm RD, Asada HH (2013) Dynamic modeling of cell migration and spreading behaviors on fibronectin coated planar substrates and micropatterned geometries. PLoS Comput Biol 9(2). https://doi.org/10.1371/journal.pcbi.1002926

Kim MC, Whisler J, Silberberg YR, Kamm RD, Asada HH (2015) Cell invasion dynamics into a three dimensional extracellular matrix fibre network. PLoS Comput Biol 11(10):1–29. https://doi.org/10.1371/journal.pcbi.1004535, 0504028

Kollmannsberger P, Fabry B (2011) Linear and nonlinear rheology of living cells. Ann Rev Mater Res 41:75–97. https://doi.org/10.1146/annurev-matsci-062910-100351

Liu F, Wu D, Wu X, Chen K (2015a) Analyses of the cell mechanical damage during microinjection. Soft Matter 11(7):1434–1442. https://doi.org/10.1039/C4SM02773F, http://xlink.rsc.org/?DOI=C4SM02773F

Liu Z, Lee Y, Jang JH, Li Y, Han X, Yokoi K, Ferrari M, Zhou L, Qin L (2015b) Microfluidic cytometric analysis of cancer cell transportability and invasiveness. Sci Rep 5:14272. https://doi.org/10.1038/srep14272, http://www.nature.com/srep/2015/150925/srep14272/full/srep14272.html

Lykov K, Nematbakhsh Y, Shang M, Lim CT, Pivkin IV (2017) Probing eukaryotic cell mechanics via mesoscopic simulations. PLoS Comput Biol 13(9):1–26. https://doi.org/10.1371/journal.pcbi.1005726, e1005726

Nan L, Jiang Z, Wei X (2014) Emerging microfluidic devices for cell lysis: a review. Lab Chip 14:1060–1073. https://doi.org/10.1039/C3LC51133B

Nematbakhsh Y, Lim C (2015) Cell biomechanics and its applications in human disease diagnosis. Acta Mech Sin 31(2):268–273. https://doi.org/10.1007/s10409-015-0412-y

Nematbakhsh Y, Pang KT, Lim CT (2017) Correlating the viscoelasticity of breast cancer cells with their malignancy. Converg Sci Phys Oncol 3(3):034003. http://stacks.iop.org/2057-1739/3/i=3/a=034003

Pinho D, Yaginuma T, Lima R (2013) A microfluidic device for partial cell separation and deformability assessment. Biochip J 7:367–374. https://doi.org/10.1007/s13206-013-7408-0

Ramaekers FCS, Bosman FT (2004) The cytoskeleton and disease. J Pathol 204(4):351–354. https://doi.org/10.1002/path.1665

Ronan W, Deshpande VS, McMeeking RM, McGarry JP (2012) Numerical investigation of the active role of the actin cytoskeleton in the compression resistance of cells. J Mech Behav Biomed Mater 14:143–157. https://doi.org/10.1016/j.jmbbm.2012.05.016

Rossinelli D, Yu-Hang T, Lykov K, Alexeev D, Bernaschi M, Hadjidoukas P, Bisson M, Joubert W, Conti C, Karniadakis G, Fatica M, Pivkin I, Koumoutsakos P (2015) The in-silico lab-on-a-chip: petascale and high-throughput simulations of microfluidics at cell resolution. In: Proceedings of 2015 international conference for high performance computing, networking, storage and analysis, SC'15. ACM, New York

Rowat A, Lammerding J, Ipsen J (2006) Mechanical properties of the cell nucleus and the effect of emerin deficiency. Biophys J 91(12):4649–4664. https://doi.org/10.1529/biophysj.106.086454, http://www.sciencedirect.com/science/article/pii/S0006349506721774

Rynearson AL, Sussman CR (2011) Nuclear structure, organization, and oncogenesis. J Gastrointest Cancer 42(2):112–117. https://doi.org/10.1007/s12029-011-9253-5

Stephens AD, Banigan EJ, Adam SA, Goldman RD, Marko JF (2017) Chromatin and lamin a determine two different mechanical response regimes of the cell nucleus. Mol Biol Cell 28(2):1–29. https://doi.org/10.1091/mbc.E16-09-0653

Suresh S (2007) Biomechanics and biophysics of cancer cells. Acta Biomater 3(4):413–438. https://doi.org/10.1016/j.actbio.2007.04.002, http://www.sciencedirect.com/science/article/pii/S174270610700061X

Takeishi N, Imai Y, Nakaaki K, Yamaguchi T, Ishikawa T (2014) Leukocyte margination at arteriole shear rate. Physiol Rep 2(6):1–8. https://doi.org/10.14814/phy2.12037, http://www.ncbi.nlm.nih.gov/pubmed/24907300

Takeishi N, Imai Y, Yamaguchi T, Ishikawa T (2015) Flow of a circulating tumor cell and red blood cells in microvessels. Phys Rev E Stat Nonlinear Soft Matter Phys 92(6):1–6. https://doi.org/10.1103/PhysRevE.92.063011

Takeishi N, Imai Y, Ishida S, Omori T, Kamm RD, Ishikawa T (2016) Cell adhesion during bullet motion in capillaries. Am J Physiol Heart Circ Physiol 311(2):H395–H403. https://doi.org/10.1152/ajpheart.00241.2016, http://ajpheart.physiology.org/content/early/2016/06/01/ajpheart.00241.2016

Ujihara Y, Nakamura M, Miyazaki H, Wada S (2010) Proposed spring network cell model based on a minimum energy concept. Ann Biomed Eng 38(4):1530–1538. https://doi.org/10.1007/s10439-010-9930-8

Ujihara Y, Nakamura M, Miyazaki H, Wada S (2012) Contribution of actin filaments to the global compressive properties of fibroblasts. J Mech Behav Biomed Mater 14:192–198. https://doi.org/10.1016/j.jmbbm.2012.05.006

Xiao LL, Fu YLSCBM (2016a) Effects of flowing RBCs on adhesion of a circulating tumor cell in microvessels. Biomech Model Mechanobiol 16:597–610. https://doi.org/10.1007/s10237-016-0839-5

Xiao LL, Fu YLSCBM (2016b) Numerical simulation of a single cell passing through a narrow slit. Biomech Model Mechanobiol 15:1655–1667. https://doi.org/10.1007/s10237-016-0789-y

Ye T, Phan-Thien N, Lim CT (2016) Particle-based simulations of red blood cells - A review. J Biomechan 49(11):2255–2266. https://doi.org/10.1016/j.jbiomech.2015.11.050

Yung C, Fiering J, Mueller A, Ingber D (2009) Micromagnetic-microfluidic blood cleansing device. Lab Chip 9:1171–1177. https://doi.org/10.1039/B816986A

Zeng Y, Kia A, Teo YSk (2012) A three-dimensional random network model of the cytoskeleton and its role in mechanotransduction and nucleus deformation. Biomech ModelMechanobiol (11):49–59. https://doi.org/10.1007/s10237-011-0292-4

Zhang Z, Xu J, Hong B, Chen X (2014) The effects of 3D channel geometry on CTC passing pressure – towards deformability-based cancer cell separation. Lab Chip 14:2576–2584. https://doi.org/10.1039/c4lc00301b, http://pubs.rsc.org/en/content/articlehtml/2014/lc/c4lc00301b

Multiscale Modeling of Malaria-Infected Red Blood Cells

109

Anil K. Dasanna, Ulrich S. Schwarz, Gerhard Gompper, and Dmitry A. Fedosov

Contents

A. K. Dasanna · U. S. Schwarz
BioQuant-Center for Quantitative Biology and Institute of Theoretical Physics, Heidelberg University, Heidelberg, Germany
e-mail: anil.dasanna@bioquant.uni-heidelberg.de; schwarz@thphys.uni-heidelberg.de

G. Gompper · D. A. Fedosov (✉)
Institute of Complex Systems and Institute for Advanced Simulation, Forschungszentrum Jülich, Jülich, Germany
e-mail: g.gompper@fz-juelich.de; d.fedosov@fz-juelich.de

Abstract

Malaria is a parasitic disease which takes approximately half a million lives every year. The unicellular parasites are transmitted by mosquitoes and mainly affect vascular blood flow by invading red blood cells (RBCs). The pathogenicity of malaria primarily results from substantial changes in the stiffness of infected RBCs and their ability to adhere to endothelial cells and other circulating blood cells, leading to a substantial disruption of normal blood circulation and inflammation of the vascular endothelium. Multiscale modeling of malaria has proved to contribute significantly to the understanding of this devastating disease. In particular, modeling on the level of single infected RBCs allows quantification of their mechanics, cytoadherence, and individual as well as collective behavior in blood flow. Recent modeling advances in this direction are discussed. We show how computational models in malaria are validated and used for the interpretation of experimental observations or the establishment of new physical hypotheses. Such computational models have a strong potential to elucidate a number of physical mechanisms relevant for malaria and to aid in the development of novel diagnostic tools and treatment strategies.

1 Introduction

Malaria is a mosquito-borne disease caused by *Plasmodium* parasites (Miller et al. 2002). These are unicellular eukaryotic cells whose continuous battle with humans has left a larger imprint on our genome than any other infectious diseases. There exist several *Plasmodium* species, including *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*, which cause very different disease severity. *P. falciparum* (Pf) is considered to be the most dangerous form of the disease with several hundred thousand deaths per year worldwide. The life cycle of a parasite is schematically illustrated in Fig. 1. Motile sporozoites are injected into the human skin during a mosquito blood meal. After entering the bloodstream, the sporozoites travel to the liver and infect hepatocytes. Within about 2 weeks, tens of thousands of merozoites form and are then released into the bloodstream, where they invade red blood cells (RBCs). Following RBC invasion, merozoites multiply inside infected RBCs (iRBCs) for about 48 h and finally rupture the RBC membrane (Abkarian et al. 2011), releasing around 20 new merozoites into the bloodstream. At the same time, some merozoites develop into the sexual form called gametocytes, which can be taken up from the blood by mosquitoes. These gametocytes infect female mosquitoes, where they first develop in the gut and later move to the salivary glands. This completes the life cycle of malaria parasites.

Infection of RBCs by malaria parasites leads to significant stiffening of the iRBC membrane for the case of Pf parasites (Cranston et al. 1984; Diez-Silva et al. 2010). Membrane stiffening can be moderate for other malaria species. The intraerythrocytic development of Pf parasites includes three major stages, called ring
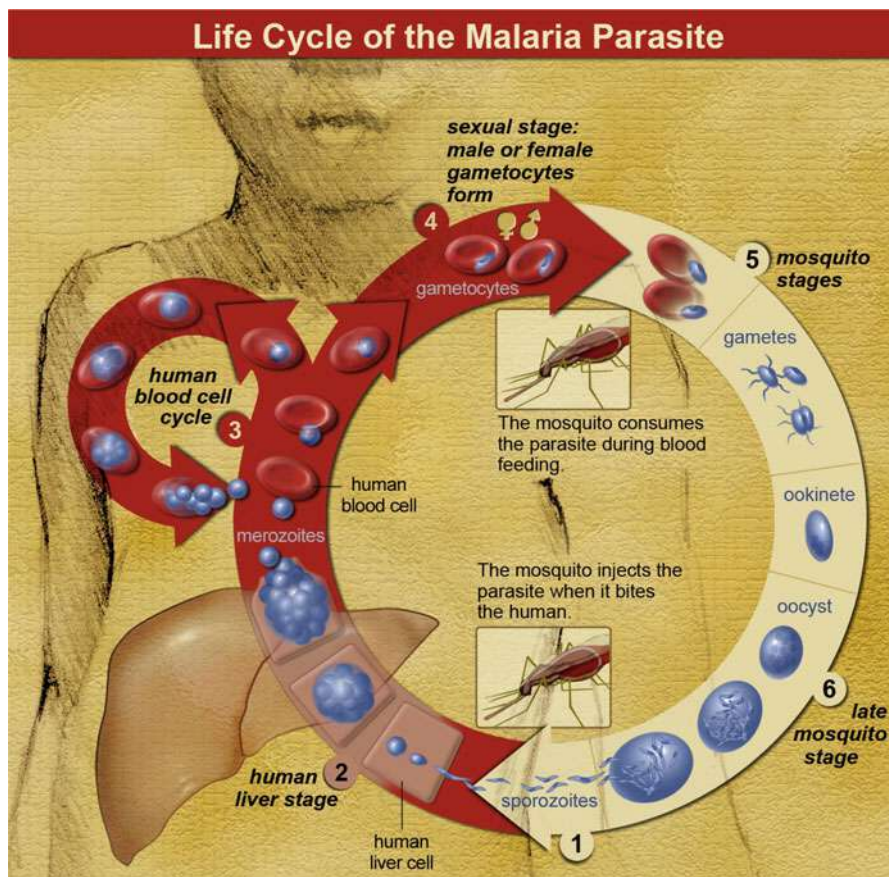
Fig. 1 The life cycle of malaria parasites. A mosquito infects a human host during a blood meal (stage 1). Then, malaria parasites in the form of sporozoites use blood vessels to reach the liver. They infect hepatocytes, develop into merozoites, and finally rupture the liver cells and enter the bloodstream again (stage 2). The merozoites invade RBCs, where they progress through the asexual intraerythrocytic stages, including ring, trophozoite, and schizont forms. Finally, RBCs are ruptured and new merozoites are released into the bloodstream, closing the asexual cycle in the blood (stage 3). Sexual forms are also produced in the blood (stage 4), which are taken up by a mosquito (stage 5), resulting in its infection and development in the gut (stage 6). They again become sporozoites and move to the salivary glands (stage 1), thus closing the malaria life cycle. From National Institute of Allergy and Infectious Diseases (NIAID)

(0–24 h), trophozoite (24–36 h), and schizont (40–48 h). During this development, the RBC membrane can stiffen by up to a factor of ten in comparison with healthy RBCs (Diez-Silva et al. 2010; Fedosov et al. 2011b). In addition, at the schizont stage iRBCs attain a near-spherical shape, since the parasites change the osmotic pressure inside the iRBC and also because they fill more and more of its volume.

The stiffening and shape change of iRBCs impair their ability to deform and may lead to capillary occlusions (Cranston et al. 1984; Shelby et al. 2003).

Inside iRBCs, merozoites remain virtually invisible to the immune system. However, the membrane stiffening can be detected in the spleen, where RBCs have to squeeze through narrow slits as a part of the RBC quality control system (Engwerda et al. 2005; Pivkin et al. 2016). In order to survive at the late stages (trophozoite and schizont), parasites position adhesive proteins at the RBC membrane surface, which mediate cell adhesion to the endothelium (Brown et al. 1999; Miller et al. 2002). This mechanism delays or even prevents the passage of iRBCs through the spleen, allowing enough time for parasites to develop and rupture the cell membrane. Interestingly, the adhesive system induced by the parasite is similar to the way white blood cells (WBCs) adhere in the vasculature as a prerequisite for extravasating into the surrounding tissue, e.g., in the context of an infection (Helms et al. 2016). While WBCs use hundreds of microvilli with adhesive tips, iRBCs use thousands of adhesive knobs. In both cases, the protrusion from the surface seems essential to make contact with the endothelium (Korn and Schwarz 2006). Cytoadhesion is also the main reason for the absence of the late stages of iRBCs in patient blood samples, because adhesive cells remain immobile in the microvasculature. While cytoadhesion facilitates further progression of malaria, it has a strong potential for severe disruption of normal blood flow. Infected cells at the late stages are able to adhere not only to the vascular endothelium but also to other infected and healthy RBCs. Adherence of iRBCs is likely to be the main cause of bleeding complications in cerebral malaria due to blockages of small vessels in the brain (Adams et al. 2002). Cytoadherence also leads to inflammation of the vascular endothelium, which is an important part of the symptoms of the disease (Miller et al. 2002).

In this chapter, we focus on modeling various aspects of the malaria disease, using a multiscale computational framework. At the current stage, the multiscale approach to malaria already spans all relevant levels: from molecular aspects of RBC remodeling by the parasite to cell-level changes in mechanics and cytoadherence of iRBCs, and to the flow of many RBCs in complex geometries. Computational models are validated by comparison with a growing range of single-cell experiments under healthy and diseased conditions. Here, we discuss all levels of the multiscale problem of the blood stage of malaria. In particular, we discuss how the RBC is remodeled by the parasite. We then address modeling of RBC mechanics and the changes it experiences over the course of intraerythrocytic parasite development. Simulations of parasite invasion of RBCs and the cytoadherence of iRBCs at the late stages are described. Finally, we also review numerical efforts to understand blood rheology and flow in malaria and illustrate several microfluidic examples which can be utilized as novel devices for malaria detection and diagnosis. The computational modeling has proved to be a valuable tool in elucidating involved physical and biological mechanisms in malaria, and we discuss possible future steps, which can bring realistic simulations closer toward medical applications.

2   Methods and Models

2.1   Structure of Healthy and Infected Red Blood Cells

Healthy human RBCs have a biconcave shape with approximately 7.5–8.7 μm in diameter and 1.7–2.2 μm in thickness (Fung 1993). The RBC envelope consists of a phospholipid bilayer and a network of spectrin proteins (cytoskeleton) attached at the inner side of the bilayer via transmembrane proteins; see Fig. 2a. The spectrin network supplies shear elasticity to a RBC membrane, while the lipid bilayer serves as a barrier for exchange of solutes and provides resistance to bending and viscous damping when sheared. Human RBCs neither have organelles nor bulk cytoskeleton and are filled with a highly concentrated hemoglobin solution. Viscosity of the cytosol is about 6×10^{-3} Pa·s, while the plasma viscosity is approximately 1.2×10^{-3} Pa·s at a physiological temperature of 37 °C (Wells and Schmid-Schönbein 1969). Thus, the cytosol viscosity is about five times larger than that of the plasma under physiological conditions.

In comparison to healthy cells, iRBCs contain a growing parasitic body inside them. Furthermore, after the end of the ring stage or about halfway through the asexual cycle, the iRBC membrane starts developing adhesive protrusions on its surface, known as knobs, which are responsible for cytoadhesion. Knobs have been visualized using scanning electron microscopy (SEM) (Gruenberg et al. 1983) and atomic force microscopy (AFM) (Nagao et al. 2000; Quadt et al. 2012) during the different stages of the life cycle. These experiments have shown that the formation of knobs is strain-dependent and variable, with a density of about 10–30 knobs/μm² at the trophozoite stage and around 40–60 knobs/μm² at the schizont stage. The
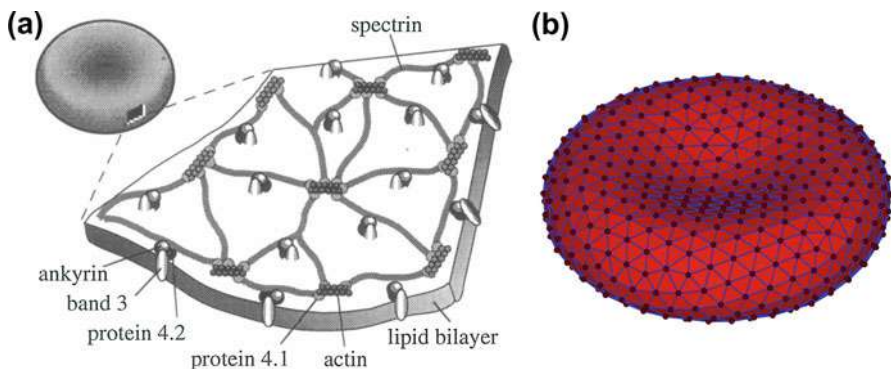
Fig. 2 (**a**) A schematic of the RBC membrane structure, showing the spectrin cytoskeleton being anchored to the lipid bilayer by transmembrane proteins (Reproduced with permission from Hansen et al. 1997). (**b**) Mesoscopic representation of a RBC membrane by a triangular network of mechanical bonds. For iRBCs, adhesive bonds are anchored at the vertices of this network (Reproduced with permission from Fedosov et al. 2014)

height of knobs remains nearly constant around 10–20 nm, while the knob diameter decreases from 160 to 100 nm (Gruenberg et al. 1983), when iRBCs progress from the trophozoite to the schizont stage.

2.2   Overview of Hydrodynamic Methods

Modeling blood flow in malaria requires mathematical formulations for fluid flow on the one side and blood cell shape and mechanics on the other side. Both blood plasma and RBC cytosol can be considered to be viscous Newtonian fluids. For microcirculatory blood flow, we deal with hydrodynamics at low Reynolds numbers, and therefore the Stokes equation has to be solved to describe fluid flow. For WBCs and iRBCs in the schizont stage, a spherical cell shape can be assumed and analytical solutions are available for the Stokes equation (Cichocki and Jones 1998). For the standard case of non-spherical RBCs, however, numerical approaches have to be employed. Fluid flow can be simulated by a variety of methods including continuum approaches based on the Navier-Stokes equations (Wendt 2009) and particle-based simulation methods, such as smoothed particle hydrodynamics (Monaghan 2005), dissipative particle dynamics (Pivkin et al. 2011), and multi-particle collision dynamics (Gompper et al. 2009). Even though continuum techniques are often faster and more robust than particle-based methods, certain features such as thermal fluctuations and the inclusion of suspended complex structures (e.g., molecules and cells) are better suited for particle-based methods. Therefore, particle-based methods are very popular for the modeling of the dynamics of complex fluids such as blood.

2.3   Adhesive Dynamics of Spherical Cells

Modeling the motion of spherical cells, such as WBCs and iRBCs at the schizont stage (Helms et al. 2016; Dasanna et al. 2017), is mathematically much simpler than the motion of cells with a non-spherical shape. Adhesive dynamics of a spherical cell in shear flow can be described by the Langevin equation (Hammer and Apte 1992; Korn and Schwarz 2007) as

$$\partial_t X(t) = u^\infty + M\{\mathbf{F}_S + \mathbf{F}_D\} + k_B T \nabla M + \xi(t), \tag{1}$$

where $X(t)$ is a six-dimensional vector describing sphere translation and rotation, M is the mobility matrix, u^∞ is the imposed flow, and $\mathbf{F}_S$ and $\mathbf{F}_D$ denote shear (i.e., from fluid flow) and direct (e.g., gravity, adhesion) forces and torques, respectively. The term with ∇M is a non-zero spurious drift, since the mobility matrix depends on the sphere position. The last term $\xi(t)$ corresponds to a random force which represents thermal fluctuations and obeys the fluctuation-dissipation theorem with an equilibrium temperature T.

For cell adhesion, stochastic bond dynamics is modeled between receptors on the cell surface and ligands at the substrate. Bond association and dissociation are governed by on- and off-rates, k_{on} and k_{off}, respectively. Usually k_{on} is taken to be constant, and a bond with the rate k_{on} may form between a receptor and a ligand if the distance between them is less than a specified encounter distance r_0. For an existing bond, Bell's equation $k_{off} = k_{off}^0 \exp(F/F_d)$ (Bell 1978) is used to model that most bonds dissociate faster under force F, with an internal force scale F_d. Mechanically, a bond is assumed to be a harmonic spring with a spring constant k_s and equilibrium bond length ℓ_0. The probability for bond association and dissociation is calculated as $P = 1 - \exp(-k\Delta t)$, where Δt is the time step employed in simulations.

2.4   Modeling Cell Deformation

In the general case of a non-spherical cell shape, hydrodynamic flow has to be combined with a deformable cell model. RBCs are modeled by a flexible network of springs with triangular elements (Seung and Nelson 1988; Noguchi and Gompper 2005; Fedosov et al. 2010), as shown in Fig. 2b. The free energy of the membrane is defined as

$$U_{cell} = U_s + U_b + U_{a+v}, \tag{2}$$

where U_s is the spring's potential energy to impose membrane shear elasticity, U_b is the bending energy to represent bending rigidity of a membrane, and U_{a+v} stands for the area and volume conservation constraints, which mimic area incompressibility of the lipid bilayer and incompressibility of the cytosol, respectively. Linear properties of a regular hexagonal network can be derived analytically, providing a relation between the model parameters and the membrane macroscopic properties (Seung and Nelson 1988; Fedosov et al. 2010). Similar to the adhesive dynamics for spherical cells, receptors can be placed at the spring network and on- and off-rates can be used to model adhesive interactions with the substrate (Fedosov et al. 2011a, b). More details on blood cell and flow modeling can be found in recent reviews (Li et al. 2013; Fedosov et al. 2014; Freund 2014).

3   Results

3.1   RBC Shapes and Mechanics

The biconcave shape of RBCs is controlled by the relative ratio between their membrane area and cell volume and by the membrane elastic properties. The volume of a RBC is about 35–40% smaller than the volume of a sphere with the same surface area. Thus, the reduced volume of a RBC is normally in the range between 0.6 and 0.65. Lipid vesicles with this reduced volume attain a biconcave shape which

closely resembles that of RBCs (Seifert et al. 1991). In comparison to vesicles, RBC membrane also possesses shear elasticity supplied by the spectrin network. This mechanical component is important to prevent budding of vesicles from the RBC membrane and membrane rupture under very large deformations. Therefore, the equilibrium shape of RBCs in general is determined by a minimum energy of both bending and shear-elastic contributions (Lim et al. 2002).

For both contributions to the Hamiltonian, it is not clear what the appropriate reference state should be. In case of the bending energy, it is important to consider whether a membrane has spontaneous curvature, especially because the RBC membrane is known to have very different compositions in the inner and outer leaflets. Most existing RBC models (Noguchi and Gompper 2005; Fedosov et al. 2010) assume the spontaneous curvature to be zero. A large enough spontaneous curvature can lead to a variety of different shapes (e.g., stomatocyte, echinocyte) and change the importance of bending and shear-elastic contributions (Lim et al. 2002). However, questions whether a human RBC membrane possesses a non-zero spontaneous curvature and whether this curvature would be isotropically distributed on the RBC surface still remain unanswered.

The reference state for the shear-elastic energy is generally referred to as stress-free shape of a RBC (Peng et al. 2014). It appears plausible to assume the biconcave shape of RBCs as the stress-free shape, because transmembrane proteins, which connect the spectrin network and the lipid bilayer, are able to diffuse within the bilayer, leading to a potential relaxation of existing elastic stresses. However, recent simulations (Peng et al. 2014) show that the assumption of biconcave stress-free shape leads to an incorrect prediction of the transition from tumbling dynamics to tank-treading dynamics of a RBC in shear flow. On the other hand, the assumption of a spherical stress-free shape results in a stomatocytic shape of the membrane for physiological values of RBC bending rigidity and shear elasticity (Li et al. 2005). An assumption of a spheroidal stress-free shape close to a sphere seems to be able to resolve these issues (Peng et al. 2014).

The mechanical properties of RBC membranes have been measured by a number of single-cell experimental techniques, including micropipette aspiration (Waugh and Evans 1979; Discher et al. 1994), deformation by optical tweezers (Henon et al. 1999; Suresh et al. 2005), optical magnetic twisting cytometry (Puig-de-Morales-Marinkovic et al. 2007), and three-dimensional measurement of membrane thermal fluctuations (Popescu et al. 2007; Park et al. 2008). Optical tweezers and micropipette aspiration methods apply a strong global deformation to the RBC membrane, while optical magnetic twisting cytometry and measurements of membrane thermal fluctuations correspond to local measurements. The two former techniques allow a measurement of the macroscopic shear modulus of a RBC membrane. Figure 3 presents a comparison between the deformation of a healthy RBC by optical tweezers (Suresh et al. 2005) and the corresponding numerical simulations (Fedosov et al. 2010). The numerical simulations mimic experimental conditions and allow the quantification of observed RBC deformation. A best fit (black line) between simulations and experiments is achieved for the two-dimensional (2D) Young's modulus of $Y = 18.9\,\mu$N/m and 2D shear modulus
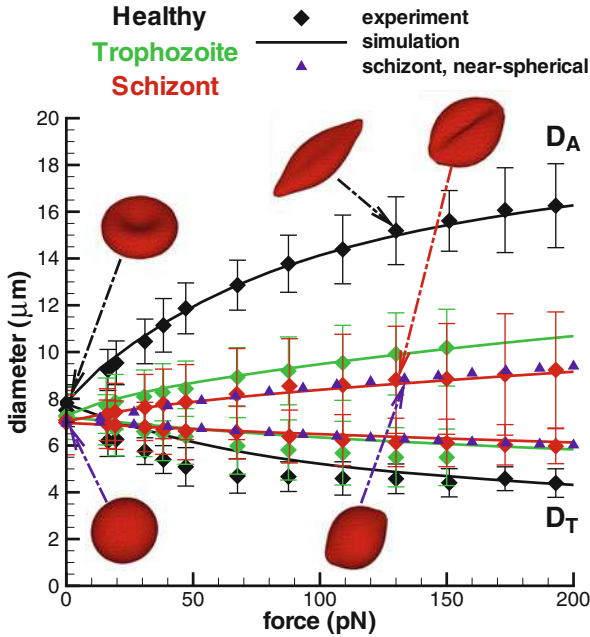
Fig. 3 Stretching response of a healthy RBC and iRBCs for different stages compared with the optical tweezers experiments (Suresh et al. 2005). D_A and D_T refer to the axial and transverse diameters (Reused with permission from Fedosov et al. 2011b)

of $\mu = 4.73\,\mu\text{N/m}$. Note that this value of the shear modulus is in 2D such that $\mu \approx Y/4$, while Fedosov et al. (2010) provide a 3D value with $\mu \approx Y/3$.

Optical magnetic twisting cytometry (Puig-de-Morales-Marinkovic et al. 2007) allows the measurement of local rheological properties (e.g., the complex modulus) of a RBC membrane. Complex modulus consists of storage and loss moduli, which are determined by elastic and viscous properties of the membrane, respectively. Therefore, both bending rigidity and shear elasticity affect the storage modulus, while the loss modulus is associated with the viscosity of a membrane, cytosol, and suspending medium (Fedosov et al. 2010). Simulations performed to quantify the experimental data (Puig-de-Morales-Marinkovic et al. 2007) have yielded a value of $\kappa = 3 \times 10^{-19}$ J for the bending rigidity (about $70 k_B T$) and $\eta_m = 2.2 \times 10^{-8}$ Pa·s·m for the membrane viscosity (Fedosov et al. 2010). Measurements of membrane thermal fluctuations (Popescu et al. 2007; Park et al. 2008) are directly associated with the membrane characteristics; however, their reliable interpretation for RBCs still remains difficult. There exist large discrepancies between different studies, which are likely to originate from the approximations used in analytical models derived for planar and near-spherical membranes (Strey et al. 1995; Betz and Sykes 2012). Hence, a quantitative interpretation of fluctuation measurements requires reliable and accurate simulation models of a RBC.

A number of experiments with RBCs provide sufficient evidence for a complex membrane mechanical response including its unique viscoelastic properties. In addition, it has been recognized that RBCs possess a metabolic activity through the consumption of adenosine triphosphate (ATP), which contributes to measured membrane flickering. The studies with ATP depletion (Betz et al. 2009; Park et al. 2010) have shown that membrane fluctuations decrease; however, during the ATP depletion process, there is no guarantee that RBCs are not subject to changes in membrane elasticity. In contrast, another investigation (Evans et al. 2008) has questioned the effect of ATP on measured flickering. A recent experimental and simulation study (Turlier et al. 2016) has provided compelling evidence for cell activity by testing directly the fluctuation-dissipation relation, which is valid for any system in equilibrium. A violation of the fluctuation-dissipation relation has been shown, and a contribution of active processes to the observed membrane flickering has been demonstrated (Turlier et al. 2016).

3.2   Mechanics of RBC Invasion by a Parasite

The first important step in the blood cycle of malaria is the invasion of RBCs by merozoites. Conceptually, this process can be divided into discrete steps, as shown schematically in Fig. 4a. Invasion commences with a low-affinity,
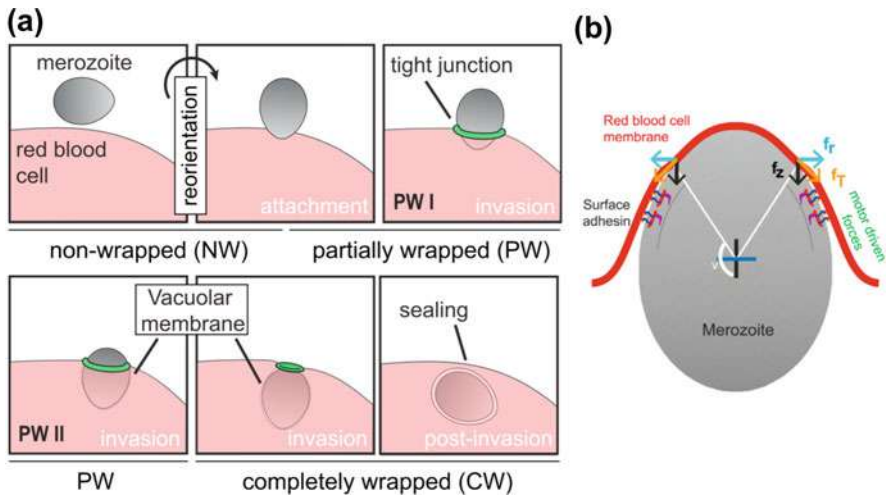
Fig. 4 (**a**) The stages of merozoite invasion. Schematic representation depicting different wrapping phases of the merozoite from reorientation and attachment through two partially wrapped states (with different wrapping fractions) to post-invasion. (**b**) Actomyosin force supports merozoite invasion. f_T is the force acting tangentially along the membrane surface as it wraps along the particle, whereas f_z is the component of this tangential force along the z axis whose role is to inject the particle into the membrane, while the component f_r is balanced by an equal force at the other side of the membrane (Reprinted with permission from Dasgupta et al. 2014b)

long-range (12–40 nm), and nondirectional binding of a RBC by the merozoite, which then reorients such that its apex directly contacts the RBC. Then, formation of a close-range interaction (4 nm or less) leads to the establishment of a RBC-merozoite tight junction. This critical structure is the organizing nexus around which invasion events are orchestrated. It acts as the aperture through which the merozoite passes during invasion and segregates RBC membrane from an emerging vacuolar membrane (likely derived in part from the parasite membrane), which fuses to form the parasitophorous vacuole surrounding the parasite after invasion (Cowman et al. 2012).

The merozoite has a shape, very similar to a chicken egg, with a width-to-length ratio of 0.7, but with a characteristic length of only 2 μm (Dasgupta et al. 2014b). The adhesion and wrapping of spherical (Deserno 2004; Bahrami et al. 2016) and non-spherical (Dasgupta et al. 2013, 2014a) nano- and microparticles by membranes have been investigated quite intensively in recent years (Dasgupta et al. 2017). These theoretical and numerical approaches provide the basis for the modeling of merozoite invasion. Particles are expected to adhere to a membrane with its least-curved side, because this configuration provides the largest adhesion area with the lowest cost in membrane deformation energy. Therefore, the reorientation of the merozoite for apical invasion requires a nonuniform adhesion energy, with an adhesion maximum at the apex. This could be facilitated by a transient gradient of adhesive proteins from apex to base on the merozoite surface (Cowman et al. 2012). Once the tight junction has been formed, the wrapping of the merozoite is controlled by four energetic contributions: (i) the adhesion strength, (ii) the membrane bending rigidity and its spontaneous curvature, (iii) the lateral tension of the RBC membrane, and (vi) the effective line tension of the tight junction. Different invasion scenarios can be discussed on the basis of such a model (Dasgupta et al. 2014b). Obviously, if the adhesion strength is very large (or small) and the membrane tension is very low (or high), complete wrapping will always (or never) occur. Large adhesion strengths allow immediate complete wrapping and RBC entry, but might also be associated with unspecific binding to other membranes and problems associated with membrane surface-coat shedding. Thus, the most interesting scenario is where these various contributions nearly balance each other. Also, this scenario for merozoite invasion is plausible, because merozoites can successfully enter RBCs, but not with certainty. As an example, phase diagrams for parasite wrapping from Dasgupta et al. (2014b) show that for a fixed line tension, there is a non-wrapped state for small, reduced adhesive strength, two partial-wrapped states for intermediate adhesion, and a complete-wrapped state for large adhesive strength. The parasite can navigate in the phase diagram to move from the non-wrapped to the complete-wrapped state. For instance, it can move between the partial-wrapped states (shallow wrapped and deep wrapped) by reducing the membrane tension through secretion of unstructured lipids. There is indeed experimental evidence for such a process, the discharge of lipids from the rhoptry organelles (Hanssen et al. 2013; Bannister et al. 1986). Similarly, a change of the spontaneous curvature to a value more favorable for wrapping (e.g., by detachment of the spectrin cytoskeleton from the bilayer or its reorganization

Kabaso et al. 2010) can induce a transition from the deep-wrapped to complete-wrapped state.

It is important to note that the transitions from shallow- to deep-wrapped states and from deep- to complete-wrapped states are associated with large energy barriers. Even though the complete-wrapped state is energetically favorable, it might be difficult to reach it from the deep-wrapped state. Thus, it is conjectured that the barrier crossing is facilitated by active processes. Indeed, there is a long-standing experimental evidence that activity of the parasite actomyosin motor governs successful host-cell entry (Angrisano et al. 2006). The current model for the source of parasite active motor force (see Fig. 4b) posits that an anchored myosin motor inside the parasite (tethered to the RBC cytoskeleton) transmits force directly through a short polymerized actin filament, which is linked to the surface-bound adhesins (Baum et al. 2006). The height of energy barrier obtained from the wrapping model described above now allows an estimate of the magnitude of the required motor activity, which indicates that a few 10s to a few 100s of motor proteins should suffice. Thus, even when membrane-wrapped states are stable, the essential role of the motor lies in overcoming energy barriers between the partial-wrapped and complete-wrapped states.

3.3 RBC Remodeling During Infection

The shapes of iRBCs can be measured, for example, by fluorescence confocal microscopy, and from these data surface area and volume can be deduced. An early study with restricted time resolution has reported relatively constant values (Esposito et al. 2010), in contrast to a colloid-osmotic model, which predicts the total volume of iRBCs to increase dramatically in the later stages (Lew et al. 2003; Mauritz et al. 2009). To resolve this issue, Waldecker et al. (2017) have reconstructed iRBC shapes over the whole intracellular cycle with a time interval of 4 h. Figure 5a shows typical shapes of iRBCs at different times, including the shape of the parasite inside. At the ring stage, the parasite is located at the rim of iRBC, causing a little bump along the edge of the cell. Starting from the trophozoite stage, parasite begins to multiply and the overall parasitic volume increases. At the end of the life cycle, iRBC reaches a nearly spherical shape with around 20 parasites inside the cell. The corresponding changes in the iRBC and parasite volume are quantified in Fig. 5b. The experimental measurements agree well with the predictions of the colloid-osmotic model, which are shown by solid lines (Mauritz et al. 2009; Lew et al. 2003). While the volume increases by 60%, the surface area remains nearly constant throughout the life cycle.

Starting around the midpoint of development, iRBCs induce adhesive protrusions at the membrane, which significantly modify the structure of the iRBC cytoskeleton. Shi et al. (2013) performed AFM imaging of inverted iRBCs to study the changes in the spectrin network during infection, as shown in Fig. 6a. The average length of spectrin filaments is found to be 43 ± 5 nm for healthy RBCs, whereas for the ring stage, the spectrin-filament length becomes about 48 ± 7 nm. For the
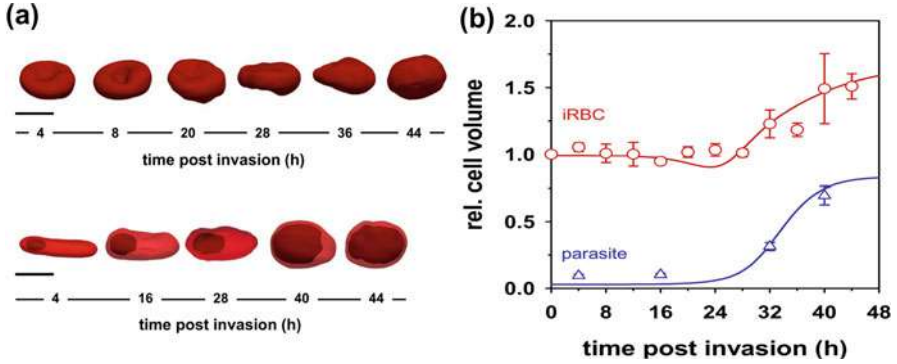
Fig. 5 (**a**) Reconstructed shapes from confocal images of iRBCs at different times of the parasite development. The bottom row shows the shapes of a growing parasite inside the iRBC at various times. Scale bar is $5\,\mu m$. (**b**) Relative cell volume and parasite volume as a function of post-infection time. The solid lines are predictions by the osmotic colloid model (Reused with permission from Waldecker et al. 2017)
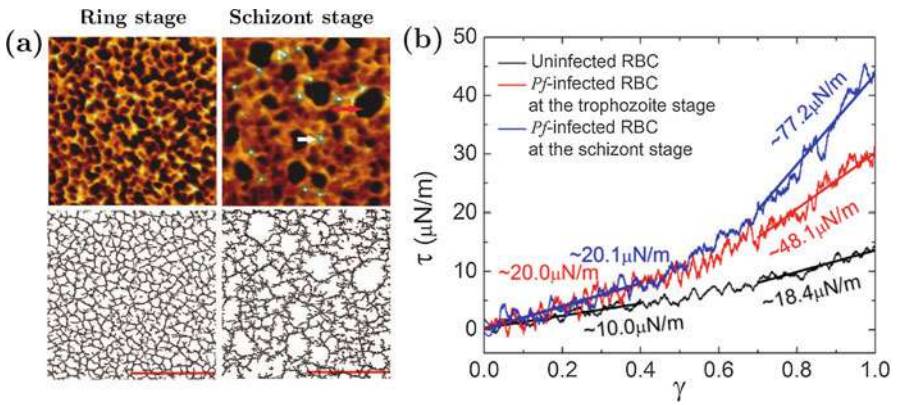
Fig. 6 (**a**) AFM images of spectrin cytoskeleton and image-processed representations for ring and schizont stages. Scale bar is 500 nm. The red arrow in schizont AFM image indicates a hole region, while the white arrow points to a knob Reused with permission from Shi et al. 2013). (**b**) Shear stress-strain response curves for uninfected and infected RBCs from mesoscopic model of RBC membrane (Reused with permission from Zhang et al. 2015)

middle trophozoite stage, the length is $64 \pm 9\,nm$, and for the schizont stage, it is about $75 \pm 11\,nm$, which is roughly twice the spectrin length of healthy RBCs. In addition, spectrin filaments are found to accumulate at the knob areas, but they become sparser in non-knobby areas. Zhang et al. (2015) have studied the effect of knobs and enhanced coupling between plasma membrane and spectrin network using coarse-grained simulations based on these observations. Knobs were modeled as stiff regions of plasma membrane, and coupling between the plasma membrane and spectrin network was increased at the knobby areas. Simulation

of the strain response of this composite membrane to shear stress results in the estimation of shear modulus for uninfected and infected RBCs. Figure 6b shows shear stress-strain response curves for uninfected, trophozoite, and schizont stages with the 2D shear moduli (at large shear strains) $\mu \simeq 18\,\mu\text{N/m}$, $\mu \simeq 48\,\mu\text{N/m}$, and $\mu \simeq 78\,\mu\text{N/m}$, respectively, which are close to values reported in experiments (Suresh et al. 2005). In general, the simulations show strong strain-hardening effects, presumably due to the polymer nature of the spectrin.

3.4 RBC Mechanics During Infection

Progressive stiffening of iRBCs has been measured by optical tweezers (Suresh et al. 2005) and by diffraction phase microscopy through monitoring thermal fluctuations (Park et al. 2008). Figure 3 shows a comparison of RBC stretching between simulation results for healthy and infected RBCs at different stages and experimental results (Suresh et al. 2005). Fitting of RBC stretching in simulations to the experimental data leads to 2D shear moduli $\mu = 14.5\,\mu\text{N/m}$ for the ring stage, $\mu = 29\,\mu\text{N/m}$ for the trophozoite, and $\mu = 40\,\mu\text{N/m}$ for the schizont (Fedosov et al. 2011b). Note that the geometry of an iRBC at the schizont stage is taken to be of ellipsoidal shape with the axes $a_x = a_y = 1.2a_z$, which is not very far from a sphere. Taking a biconcave shape for the schizont stage in simulations results into a prediction with a larger shear modulus than that for the near-spherical shape. In the case of the deflated biconcave shape, the form of an iRBC is altered first in response to stretching and followed by membrane deformation, while for the near-spherical shape, the fluid restricts shape deformation and membrane stretching has to occur right away. In conclusion, the cell geometry plays an important role in the quantification of experimental data.

3.5 Adhesion of Infected Cells

Adhesion of iRBCs lies at the heart of parasite survival, as it prevents elimination of iRBCs in the spleen. Adhesive receptors on iRBC surface (PfEMP-1 encoded by the family of var genes) are localized to the knobs and can bind multiple ligands (mainly ICAM-1 and CD36) on vascular endothelial cells. The precise roles of different ligands in iRBC adhesion are quite complex, as the knob receptors bind to the various ligands with distinct adhesion strength (Yipp et al. 2000). For instance, experiments with flow assays hint at an increase of ICAM-1 adherence of iRBCs with shear stress (i.e., displaying characteristics of a catch bond), while CD36 does not show this behavior (Nash et al. 1992). Enhancement of adhesion under flow is not uncommon for cells in shear flow, including WBCs and bacteria, and has been demonstrated for iRBCs adhering to supported bilayers (Rieger et al. 2015). Experiments using single-bond force spectroscopy (Lim et al. 2017) have shown that iRBCs form slip bonds with CD36 and catch bonds with ICAM-1, as suggested before by flow chamber experiments. The characteristic rupture force for

both slip and catch bonds is close to 10 pN, which is also comparable with the force for WBC bonds in the range of 10–50 pN (Marshall et al. 2003; Hanley et al. 2004). On the modeling side, the slip bond can be simulated by Bell's model (Bell 1978), while the catch bond is often described by Dembo's model (Dembo et al. 1988).

The standard model for the adhesion of single cells under flow has been established with adhesive dynamics for round cells. Originally developed for WBCs (Hammer and Apte 1992; Korn and Schwarz 2007), this approach has recently been applied to iRBCs at the schizont stage (Helms et al. 2016; Dasanna et al. 2017). A phase diagram with different dynamic states (free motion, transient adhesion, rolling adhesion, and firm adhesion) has been simulated for the slip-bond model as a function of molecular on- and off-rates (Korn and Schwarz 2008; Helms et al. 2016), which can serve as a reference case to interpret experimental data. In the rolling adhesion regime, rolling velocity as a function of shear rate can be used to estimate molecular data. Hence, it has been estimated that ICAM-1 receptor distance should be between 100 and 400 nm to yield the observed rolling velocity around 100 μm/s at a shear rate of about $100\,s^{-1}$ (Dasanna et al. 2017).

An interesting aspect of iRBC adhesive motion is the nature of single trajectories, which often displays an oscillatory pattern in velocity and has been studied both numerically and in experiments (Fedosov et al. 2011a; Dasanna et al. 2017). Unlike WBCs, iRBCs have a non-spherical shape for the majority of parasite's life cycle, except the schizont stage. Mesoscopic modeling of iRBCs (Fedosov et al. 2011a) has demonstrated that cells at the trophozoite stage flip rather than roll on a substrate because of their discoid shape. For a Young's modulus similar to that of healthy RBCs, crawling motion is observed, while a large enough Young's modulus (more than three times that of healthy RBCs) leads to a rather regular flipping motion. The flipping may become less regular with the presence of parasite inside the cytosol
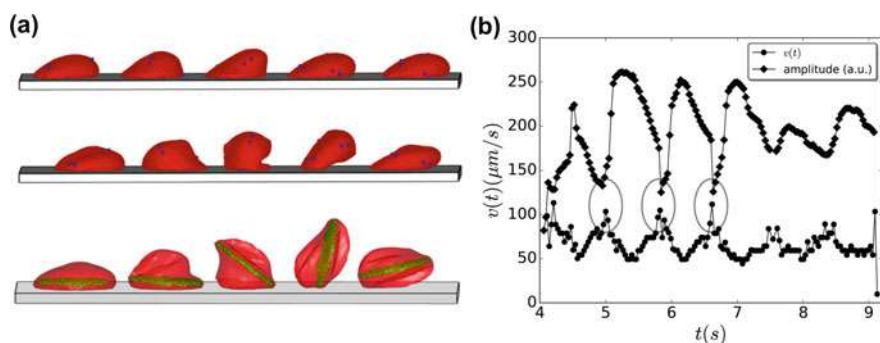
Fig. 7 (**a**) Snapshots of modeled flipping iRBCs with a small membrane Young's modulus (top), a large Young's modulus (middle), and, when it contains a rigid body, imitating the presence of parasite (bottom) (Reused with permission from Fedosov et al. 2011a). (**b**) Representative trajectory of a rolling trophozoite cell on endothelial layer, showing cell velocity and fluorescence intensity of the parasite inside the iRBC (Reused with permission from Dasanna et al. 2017)

(Fedosov et al. 2011a). Figure 7a shows snapshots of RBC motion for the three cases: (top) a crawling cell with a Young's modulus of $Y = 18.9\,\mu\text{N/m}$, (middle) a flipping iRBC with a Young's modulus of $Y = 168\,\mu\text{N/m}$, and (bottom) a flipping cell with a parasite body inside. Dasanna et al. (2017) have studied rolling adhesion of iRBCs at the trophozoite and schizont stages on endothelial cells in flow chamber experiments. The effect of the parasite on the iRBC rolling dynamics has been investigated, and therefore both the iRBC and fluorescently labeled parasite were tracked. Figure 7b presents translational velocity and fluorescence amplitude of the parasite inside the iRBC for a rolling trophozoite. The elliptic markings of the peaks in velocity and fluorescence amplitude indicate the anticorrelation between them, which is significant for flipping. Thus, when the iRBC starts a flip, the parasite moves away from the substrate (fluorescence amplitude thereby decreases), and when it comes down, the fluorescent amplitude increases. It has been also shown that this anticorrelation disappears and the fluorescent amplitude does not oscillate for iRBCs at the schizont stage, as they are round and rigid, so that the flipping transforms to rolling. Rolling velocities of schizont-stage iRBCs are found to be smaller than those of trophozoites, presumably because of the larger number of knobs (Dasanna et al. 2017).

3.6 Blood Rheology in Malaria

Bulk viscosity of blood is a macroscopic characteristic that depends on single-cell properties and their collective interactions (Fedosov et al. 2011d; Lanotte et al. 2016). The bulk viscosity of blood in malaria increases with the parasitemia level (or fraction of iRBCs) because of the increased stiffness of iRBCs (Raventos-Suarez et al. 1985) in comparison to healthy RBCs. A significant viscosity increase leads to an increased blood flow resistance and a reduced blood perfusion. Numerical simulations (Fedosov et al. 2011b) predict an increase of blood flow resistance in microvessels in malaria up to 50% for high parasitemia levels.

Bulk viscosity of infected blood in malaria for different parasitemia levels, hematocrit 30%, and shear rate $230\,\text{s}^{-1}$ has been measured experimentally (Raventos-Suarez et al. 1985) and estimated in simulations (Fedosov et al. 2011c), as shown in Fig. 8. In the simulations, infected blood is modeled as a suspension of healthy and infected RBCs at the schizont stage with $\mu = 40\,\mu\text{N/m}$. The simulated viscosity as a function of the parasitemia is in excellent agreement with the corresponding experimental data, which show a roughly linear dependence of the infected blood viscosity on parasitemia level.

iRBCs adhere not only to the endothelium but also to other healthy and infected RBCs. Simulations with attractive interactions between infected and healthy RBCs for the parasitemia of 25% show that the blood viscosity does not increase significantly (see Fig. 8). This is likely due to the fact that the simulated shear rate is high enough to disperse RBCs within the suspension and to diminish aggregation effects on the bulk viscosity. Thus, a much stronger effect of the aggregation

Fig. 8 Viscosity of the malaria-infected blood at hematocrit 30% for different parasitemias in comparison to experimental data (Raventos-Suarez et al. 1985). The diamond symbol corresponds to a simulation with aggregation interactions among RBCs (Reused with permission from Fedosov et al. 2011c)
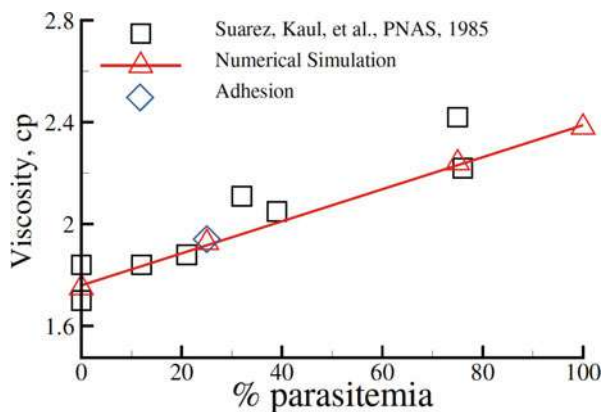
interactions should be expected at low shear rates. Furthermore, iRBCs at later stages form specific bonds with other cells, which should generally be stronger than the attractive interactions modeled by Fedosov et al. (2011c).

3.7   Blood Flow in Malaria

It is intuitive that blood flow resistance in malaria should increase due to an elevated stiffness of iRBCs and their cytoadherence. A simple model for the estimation of blood flow resistance is Poiseuille flow of blood in tubes, which mimics blood flow in microvessels. Blood flow in malaria has been simulated as a suspension of healthy and infected RBCs at the trophozoite stage ($\mu = 29 \, \mu\text{N/m}$) and hematocrit 45%. Figure 9 shows the relative apparent viscosity in malaria for several parasitemia levels from 25% to 100% and microvessels with diameters $10 \, \mu\text{m}$ and $20 \, \mu\text{m}$. The relative apparent viscosity is computed as $\eta_{\text{rel}} = \eta_{\text{app}}/\eta_s$, where η_s is the plasma viscosity and $\eta_{\text{app}} = \pi \Delta P D^4/(128 Q L)$ is the apparent viscosity. Here, ΔP is the pressure difference, Q is the flow rate, and L is the length of the tube. The relative apparent viscosity is a measure of flow resistance, as it compares apparent viscosity of blood with the viscosity of blood plasma.

The inset of Fig. 9 shows a snapshot of RBCs flowing in a tube of diameter $20 \, \mu\text{m}$ at a parasitemia level of 25%. The effect of parasitemia level on the flow resistance in Fig. 9 appears to be more prominent for small diameters and high hematocrit values. Thus, at $H_t = 45\%$ blood flow resistance in malaria may increase up to 50% in vessels of diameters around $10 \, \mu\text{m}$ and up to about 40% for vessel diameters around $20 \, \mu\text{m}$. Note that these increases do not include any contributions from the adhesive interactions between iRBCs and other cells. Therefore, the inclusion of cytoadhesion would result in a further increase in blood flow resistance in malaria.
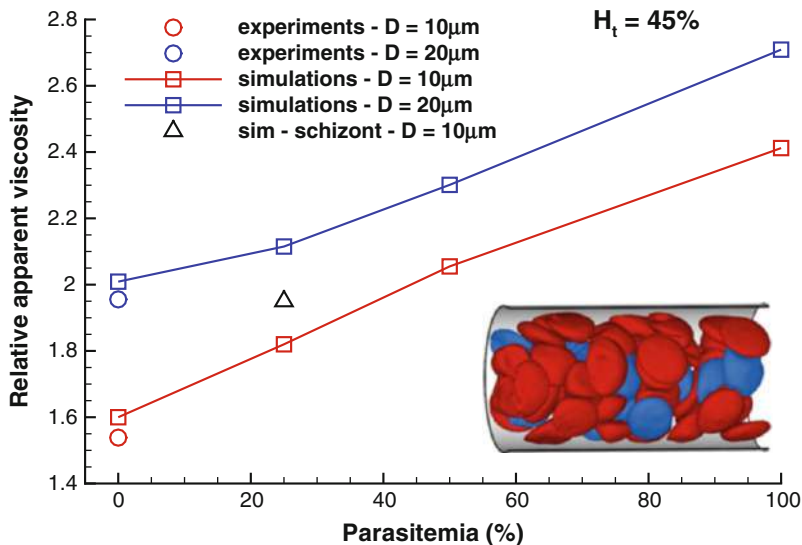
Fig. 9 Flow resistance in malaria. Healthy (red) and infected RBCs (blue) in tube flow with a diameter $D = 20\,\mu m$, hematocrit 45%, and parasitemia 25%. The relative apparent viscosity of blood in malaria is plotted for various parasitemia levels and tube diameters. Triangle symbol corresponds to the schizont stage with a near-spherical shape. Experimental data from the empirical fit are from Pries et al. (1992)

3.8 Malaria and Microfluidics

Microfluidic devices offer unique opportunities for the detection and manipulation of different suspended particles and cells. One of the first observations of iRBCs in microfluidics is a study of the passage of iRBCs through small constrictions driven by fluid flow (Shelby et al. 2003). The experiments have been performed for different channel sizes and stages of parasite development and have shown that the late stages, such as trophozoite and schizont, may result in blockage of microvessels with diameters smaller than about $6\,\mu m$. Corresponding simulation studies (Imai et al. 2010; Wu and Feng 2013) have also come to a similar conclusion, suggesting that the smallest vessel in the microvasculature can be blocked by single iRBCs in the trophozoite and schizont stages.

The filtration concept has been also employed in another microfluidic device with many obstacles (Bow et al. 2011), where both healthy and infected RBCs are able to pass constriction geometries (no blockage is observed). However, iRBCs exhibit lower average velocities in comparison to healthy RBCs. This speed difference between healthy and infected RBCs is attributed to an increased membrane stiffness and viscosity of the iRBCs in comparison to healthy cells. Deformability-based sorting has been also proposed in the context of deterministic lateral displacement (DLD) microfluidic devices (Henry et al. 2016; Krüger et al. 2014). DLDs utilize micropost arrays, which have been originally designed to continuously sort spherical
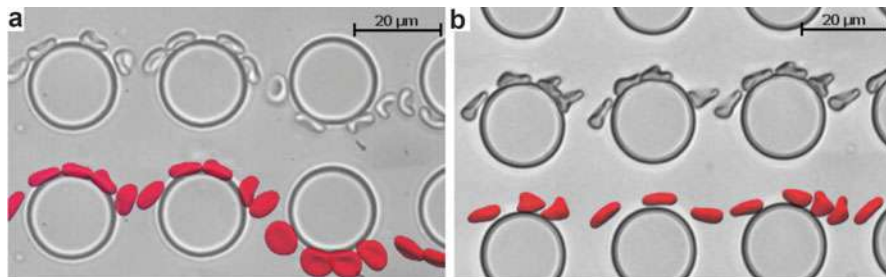
Fig. 10 Stroboscopic images of RBC deformation in a DLD device, taken from simulations and experiments. The viscosity contrast C between the cell's cytosol and suspending medium, which is (**a**) $C = 5$ and (**b**) $C = 1$, affects RBC dynamics and determines its traversal through the device (Reused with permission from Henry et al. 2016)

colloidal particles according to their size (Davis et al. 2006; Holm et al. 2011). However, DLDs do not directly follow the filtration concept, because the gaps between the posts are generally larger than suspended cells, as shown in Fig. 10. Sorting in DLDs is achieved by invoking differences in cell deformation and dynamics, which determine the trajectories of the cells within DLD devices (Henry et al. 2016; Krüger et al. 2014). For example, in Fig. 10, the differences in RBC trajectories within the DLD device are governed by different viscosity contrasts between the RBC cytosol and suspending medium. Hence, a proper tuning of DLD device geometry and flow conditions can turn them into accurate and precise deformability-based sensors useful in the context of malaria.

4   Discussion and Outlook

The development of quantitative models for malaria infections is very important, because they allow the explanation of experimental results, the testing of various hypotheses, and the proposition of new biophysical mechanisms in disease development and progression. Due to the long evolutionary history of malaria and its complicated life cycle, this is not an easy task and it requires close cooperation between theoreticians and experimentalists. In particular, future studies have to further incorporate molecular information, for example, the molecular details of how the spectrin network is remodeled by the parasite, and the exact nature of the adhesion receptors localized to the knobs and their ligand counterparts in the vascular endothelium.

From the conceptual point of view, the most important aspect of malaria modeling is the multiscale nature of blood flow, which involves a wide range of spatiotemporal scales starting from single molecules for parasite-RBC interactions and the adhesion of iRBCs, to the deformation and dynamics of single cells, and, finally, to multicellular flow in large microvascular networks. Current models have proven to be sophisticated enough to be used successfully for the quantification of

experiments with a few blood cells. Computational models of blood flow rapidly move to multicellular problems and already attempt to go beyond experimental predictions by generating and testing new physical and biological hypotheses. Thus, such simulations can be used to guide and optimize experimental settings for biophysical investigations and disease diagnosis (e.g., microfluidic devices).

Modeling of the malaria disease is very challenging, and many open questions remain, even in the context of the advances achieved over recent years, ranging from molecular mechanisms to cell deformation and adhesion under realistic blood flow conditions in the microcirculation. For instance, it is not fully understood how parasite can efficiently reorient itself after the initial random adhesion and then invade RBCs. Furthermore, exact changes and modifications of RBC membrane stiffness and cytoadherence by the parasite during its development need to be determined. In addition, it is unclear how the developing parasite, which can be thought of as a rigid-like body inside a RBC, affects the cell behavior in flow. Another important direction for future research concerns the characterization of the adhesive interactions between iRBCs and healthy RBCs, other blood cells (e.g., WBCs), and the endothelium, as this input is needed for realistic malaria modeling. Answers to these open questions for a single cell will allow the progression toward realistic modeling of blood flow in malaria, where a mixture of healthy and infected RBCs at different stages and complex flow geometries have to be generally considered.

Even though the impact of numerical modeling on malaria diagnosis and treatment is still rather limited, the development of new quantitative models and their application in practice acquire more and more momentum. Computational models have a great potential to provide guidance on how to improve disease detection and treatment and to optimize existing therapeutic tools, for instance, by introducing novel microfluidic approaches. Another substantial advancement needed for computational models is the development of predictive simulation approaches, which enable the modeling of long-time disease progression. This development requires the improvement of current models to much longer time scales including the ability of modeling various dynamic processes (e.g., parasite invasion, intra-RBC parasite development, dynamic adhesion changes).

References

Abkarian M, Massiera G, Berry L, Roques M, Braun-Breton C (2011) A novel mechanism for egress of malarial parasites from red blood cells. Blood 117:4118–4124

Adams S, Brown H, Turner G (2002) Breaking down the blood-brain barrier: signaling a path to cerebral malaria? Trends Parasitol 18:360–366

Angrisano F, Riglar DT, Sturm A, Volz JC, Delves MJ, Zuccala ES, Turnbull L, Dekiwadia C, Olshina MA, Marapana DS, Wong W, Mollard V, Bradin CH, Tonkin CJ, Gunning PW, Ralph SA, Whitchurch CB, Sinden RE, Cowman AF, McFadden GI, Baum J (2006) Spatial localization of actin filaments across developmental stages of the malaria parasite. PLoS ONE 7:e32188

Bahrami AH, Lipowsky R, Weikl TR (2016) The role of membrane curvature for the wrapping of nanoparticles. Soft Matter 12:581–587

Bannister LH, Mitchell GH, Butcher GA, Dennis ED (1986) Lamellar membranes associated with rhoptries in erythrocytic merozoites of Plasmodium knowlesi: a clue to the mechanism of invasion. Parasitology 92:291–303

Baum J, Papenfuss AT, Baum B, Speed TP, Cowman AF (2006) Regulation of apicomplexan actin-based motility. Nat Rev Microbiol 4:621–628

Bell GI (1978) Models for the specific adhesion of cells to cells. Science 200:618–627

Betz T, Sykes C (2012) Time resolved membrane fluctuation spectroscopy. Soft Matter 8:5317–5326

Betz T, Lenz M, Joanny JF, Sykes C (2009) ATP-dependent mechanics of red blood cells. Proc Natl Acad Sci USA 106:15320–15325

Bow H, Pivkin IV, Diez-Silva M, Goldfless SJ, Dao M, Niles JC, Suresh S, Han J (2011) A microfabricated deformability-based flow cytometer with application to malaria. Lab Chip 11:1065–1073

Brown H, Hien TT, Day N, Mai NTH, Chuong LV, Chau TTH, Loc PP, Phu NH, Bethell D, Farrar J, Gatter K, White N, Turner G (1999) Evidence of blood-brain barrier dysfunction in human cerebral malaria. Neuropathol Appl Neurobiol 25:331–340

Cichocki B, Jones RB (1998) Image representation of a spherical particle near a hard wall. Phys A 258:273–302

Cowman AF, Berry D, Baum J (2012) The cellular and molecular basis for malaria parasite invasion of the human red blood cell. J Cell Biol 198:961–971

Cranston HA, Boylan CW, Carroll GL, Sutera SP, Williamson JR, Gluzman IY, Krogstad DJ (1984) Plasmodium falciparum maturation abolishes physiologic red cell deformability. Science 223:400–403

Dasanna AK, Lansche C, Lanzer M, Schwarz US (2017) Rolling adhesion of schizont stage malaria-infected red blood cells in shear flow. Biophys J 112:1908–1919

Dasgupta S, Auth T, Gompper G (2013) Wrapping of ellipsoidal nano-particles by fluid membranes. Soft Matter 9:5473–5482

Dasgupta S, Auth T, Gompper G (2014a) Shape and orientation matter for the cellular uptake of nonspherical particles. Nano Lett 14:687–693

Dasgupta S, Auth T, Gov N, Satchwell TJ, Hanssen E, Zuccala ES, Riglar DT, Toye AM, Betz T, Baum J, Gompper G (2014b) Membrane-wrapping contributions to malaria parasite invasion of the human erythrocyte. Biophys J 107:43–54

Dasgupta S, Auth T, Gompper G (2017) Nano- and microparticles at biological and fluid interfaces. J Phys Condens Matter 29:373003

Davis JA, Inglis DW, Morton KJ, Lawrence DA, Huang LR, Chou SY, Sturm JC, Austin RH (2006) Deterministic hydrodynamics: taking blood apart. Proc Nat Acad Sci USA 103:14779–14784

Dembo M, Torney DC, Saxman K, Hammer D (1988) The reaction-limited kinetics of membrane-to-surface adhesion and detachment. Proc R Soc Lond B 234:55–83

Deserno M (2004) Elastic deformation of a fluid membrane upon colloid binding. Phys Rev E 69:031903

Diez-Silva M, Dao M, Han J, Lim CT, Suresh S (2010) Shape and biomechanical characteristics of human red blood cells in health and disease. MRS Bull 35:382–388

Discher DE, Mohandas N, Evans EA (1994) Molecular maps of red cell deformation: hidden elasticity and in situ connectivity. Science 266:1032–1035

Engwerda CR, Beattie L, Amante FH (2005) The importance of the spleen in malaria. Trends Parasitol 21:75–80

Esposito A, Choimet JB, Skepper JN, Mauritz JMA, Lew VL, Kaminski CF, Tiffert T (2010) Quantitative imaging of human red blood cells infected with Plasmodium falciparum. Biophys J 99:953–960

Evans J, Gratzer W, Mohandas N, Parker K, Sleep J (2008) Fluctuations of the red blood cell membrane: relation to mechanical properties and lack of ATP dependence. Biophys J 94:4134–4144

Fedosov DA, Caswell B, Karniadakis GE (2010) A multiscale red blood cell model with accurate mechanics, rheology, and dynamics. Biophys J 98:2215–2225

Fedosov DA, Caswell B, Karniadakis GE (2011a) Wall shear stress-based model for adhesive dynamics of red blood cells in malaria. Biophys J 100:2084–2093

Fedosov DA, Caswell B, Suresh S, Karniadakis GE (2011b) Quantifying the biophysical characteristics of Plasmodium-falciparum-parasitized red blood cells in microcirculation. Proc Natl Acad Sci USA 108:35–39

Fedosov DA, Lei H, Caswell B, Suresh S, Karniadakis GE (2011c) Multiscale modeling of red blood cell mechanics and blood flow in malaria. PLoS Comput Biol 7:e1002270

Fedosov DA, Pan W, Caswell B, Gompper G, Karniadakis GE (2011d) Predicting human blood viscosity in silico. Proc Natl Acad Sci USA 108:11772–11777

Fedosov DA, Noguchi H, Gompper G (2014) Multiscale modeling of blood flow: from single cells to blood rheology. Biomech Model Mechanobiol 13:239–258

Freund JB (2014) Numerical simulation of flowing blood cells. Annu Rev Fluid Mech 46:67–95

Fung YC (1993) Biomechanics: mechanical properties of living tissues, 2nd edn. Springer, New York

Gompper G, Ihle T, Kroll DM, Winkler RG (2009) Multi-particle collision dynamics: a particle-based mesoscale simulation approach to the hydrodynamics of complex fluids. Adv Polym Sci 221:1–87

Gruenberg J, Allred DR, Sherman IW (1983) Scanning electron microscope-analysis of the protrusions (knobs) present on the surface of Plasmodium falciparum-infected erythrocytes. J Cell Biol 97:795–802

Hammer DA, Apte SM (1992) Simulation of cell rolling and adhesion on surfaces in shear flow: general results and analysis of selectin-mediated neutrophil adhesion. Biophys J 63:35–57

Hanley WD, Wirtz D, Konstantopoulos K (2004) Distinct kinetic and mechanical properties govern selectin-leukocyte interactions. J Cell Sci 117:2503–2511

Hansen JC, Skalak R, Chien S, Hoger A (1997) Influence of network topology on the elasticity of the red blood cell membrane skeleton. Biophys J 72:2369–2381

Hanssen E, Dekiwadia C, Riglar DT, Rug M, Lemgruber L, Cowman AF, Cyrklaff M, Kudryashev M, Frischknecht F, Baum J, Ralph SA (2013) Electron tomography of Plasmodium falciparum merozoites reveals core cellular events that underpin erythrocyte invasion. Cell Microbiol 15:1457–1472

Helms G, Dasanna AK, Schwarz US, Lanzer M (2016) Modeling cytoadhesion of Plasmodium falciparum-infected erythrocytes and leukocytes-common principles and distinctive features. FEBS Lett 590:1955–1971

Henon S, Lenormand G, Richert A, Gallet F (1999) A new determination of the shear modulus of the human erythrocyte membrane using optical tweezers. Biophys J 76:1145–1151

Henry E, Holm SH, Zhang Z, Beech JP, Tegenfeldt JO, Fedosov DA, Gompper G (2016) Sorting cells by their dynamical properties. Sci Rep 6:34375

Holm SH, Beech JP, Barrett MP, Tegenfeldt JO (2011) Separation of parasites from human blood using deterministic lateral displacement. Lab Chip 11:1326–1332

Imai Y, Kondo H, Ishikawa T, Lim CT, Yamaguchi T (2010) Modeling of hemodynamics arising from malaria infection. J Biomech 43:1386–1393

Kabaso D, Shlomovitz R, Auth T, Lew VL, Gov NS (2010) Curling and local shape changes of red blood cell membranes driven by cytoskeletal reorganization. Biophys J 99:808–816

Korn C, Schwarz US (2006) Efficiency of initiating cell adhesion in hydrodynamic flow. Phys Rev Lett 97:138103

Korn CB, Schwarz US (2007) Mean first passage times for bond formation for a Brownian particle in linear shear flow above a wall. J Chem Phys 126:095103

Korn CB, Schwarz US (2008) Dynamic states of cells adhering in shear flow: from slipping to rolling. Phys Rev E 77:041904

Krüger T, Holmes D, Coveney PV (2014) Deformability-based red blood cell separation in deterministic lateral displacement devices – a simulation study. Biomicrofluidics 8:054114

Lanotte L, Mauer J, Mendez S, Fedosov DA, Fromental JM, Claveria V, Nicoud F, Gompper G, Abkarian M (2016) Red cells' dynamic morphologies govern blood shear thinning under microcirculatory flow conditions. Proc Natl Acad Sci USA 113:13289–13294

Lew VL, Tiffert T, Ginsburg H (2003) Excess hemoglobin digestion and the osmotic stability of Plasmodium falciparum-infected red blood cells. Blood 101:4189–4194

Li J, Dao M, Lim CT, Suresh S (2005) Spectrin-level modeling of the cytoskeleton and optical tweezers stretching of the erythrocyte. Biophys J 88:3707–3719

Li X, Vlahovska PM, Karniadakis GE (2013) Continuum- and particle-based modeling of shapes and dynamics of red blood cells in health and disease. Soft Matter 9:28–37

Lim GHW, Wortis M, Mukhopadhyay R (2002) Stomatocyte-discocyte-echinocyte sequence of the human red blood cell: evidence for the bilayer-couple hypothesis from membrane mechanics. Proc Natl Acad Sci USA 99:16766–16769

Lim YB, Thingna J, Cao J, Lim CT (2017) Single molecule and multiple bond characterization of catch bond associated cytoadhesion in malaria. Sci Rep 7:4208

Marshall BT, Long M, Piper JW, Yago T, McEver RP, Zhu C (2003) Direct observation of catch bonds involving cell-adhesion molecules. Nature 423:190–193

Mauritz JMA, Esposito A, Ginsburg H, Kaminski CF, Tiffert T, Lew VL (2009) The homeostasis of Plasmodium falciparum-infected red blood cells. PLoS Comput Biol 5:e1000339

Miller LH, Baruch DI, Marsh K, Doumbo OK (2002) The pathogenic basis of malaria. Nature 415:673–679

Monaghan JJ (2005) Smoothed particle hydrodynamics. Rep Prog Phys 68:1703–1759

Nagao E, Kaneko O, Dvorak JA (2000) Plasmodium falciparum-infected erythrocytes: qualitative and quantitative analyses of parasite-induced knobs by atomic force microscopy. J Struct Biol 130:34–44

Nash GB, Cooke BM, Marsh K, Berendt A, Newbold C, Stuart J (1992) Rheological analysis of the adhesive interactions of red blood cells parasitized by Plasmodium falciparum. Blood 79:798–807

Noguchi H, Gompper G (2005) Shape transitions of fluid vesicles and red blood cells in capillary flows. Proc Natl Acad Sci USA 102:14159–14164

Park YK, Diez-Silva M, Popescu G, Lykotrafitis G, Choi W, Feld MS, Suresh S (2008) Refractive index maps and membrane dynamics of human red blood cells parasitized by Plasmodium falciparum. Proc Natl Acad Sci USA 105:13730–13735

Park YK, Best CA, Auth T, Gov NS, Safran SA, Popescu G, Suresh S, Feld MS (2010) Metabolic remodeling of the human red blood cell membrane. Proc Natl Acad Sci USA 107:1289–1294

Peng Z, Mashayekh A, Zhu Q (2014) Erythrocyte responses in low-shear-rate flows: effects of non-biconcave stress-free state in the cytoskeleton. J Fluid Mech 742:96–118

Pivkin IV, Caswell B, Karniadakis GE (2011) Dissipative particle dynamics. In: Lipkowitz KB (ed) Reviews in computational chemistry, vol 27. Wiley, Hoboken, pp 85–110

Pivkin IV, Peng Z, Karniadakis GE, Buffet PA, Dao M, Suresh S (2016) Biomechanics of red blood cells in human spleen and consequences for physiology and disease. Proc Natl Acad Sci USA 113:7804–7809

Popescu G, Park YK, Dasari RR, Badizadegan K, Feld MS (2007) Coherence properties of red blood cell membrane motions. Phys Rev E 76:031902

Pries AR, Neuhaus D, Gaehtgens P (1992) Blood viscosity in tube flow: dependence on diameter and hematocrit. Am J Physiol 263:H1770–H1778

Puig-de-Morales-Marinkovic M, Turner KT, Butler JP, Fredberg JJ, Suresh S (2007) Viscoelasticity of the human red blood cell. Am J Physiol 293:C597–C605

Quadt KA, Barfod L, Andersen D, Bruun J, Gyan B, Hassenkam T, Ofori MF, Hviid L (2012) The density of knobs on Plasmodium falciparum-infected erythrocytes depends on developmental age and varies among isolates. PLoS ONE 7:e45658

Raventos-Suarez C, Kaul DK, Macaluso F, Nagel RL (1985) Membrane knobs are required for the microcirculatory obstruction induced by Plasmodium falciparum-infected erythrocytes. Proc Natl Acad Sci USA 82:3829–3833

Rieger H, Yoshikawa HY, Quadt K, Nielsen MA, Sanchez CP, Salanti A, Tanaka M, Lanzer M (2015) Cytoadhesion of Plasmodium falciparum-infected erythrocytes to chondroitin-4-sulfate is cooperative and shear enhanced. Blood 125:383–391

Seifert U, Berndl K, Lipowsky R (1991) Shape transformations of vesicles: phase diagram for spontaneous curvature and bilayer-coupling models. Phys Rev A 44:1182–1202

Seung HS, Nelson DR (1988) Defects in flexible membranes with crystalline order. Phys Rev A 38:1005–1018

Shelby JP, White J, Ganesan K, Rathod PK, Chiu DT (2003) A microfluidic model for single-cell capillary obstruction by Plasmodium falciparum-infected erythrocytes. Proc Natl Acad Sci USA 100:14618–14622

Shi H, Liu Z, Li A, Yin J, Chong AGL, Tan KSW, Zhang Y, Lim CT (2013) Life cycle-dependent cytoskeletal modifications in Plasmodium falciparum infected erythrocytes. PLoS ONE 8:e61170

Strey H, Peterson M, Sackmann E (1995) Measurement of erythrocyte membrane elasticity by flicker eigenmode decomposition. Biophys J 69:478–488

Suresh S, Spatz J, Mills JP, Micoulet A, Dao M, Lim CT, Beil M, Seufferlein T (2005) Connections between single-cell biomechanics and human disease states: gastrointestinal cancer and malaria. Acta Biomater 1:15–30

Turlier H, Fedosov DA, Audoly BA, Auth T, Gov NS, Sykes C, Joanny JF, Gompper G, Betz T (2016) Equilibrium physics breakdown reveals the active nature of red blood cell membrane fluctuations. Nat Phys 12:513–519

Waldecker M, Dasanna AK, Lansche C, Linke M, Srismith S, Cyrklaff M, Sanchez CP, Schwarz US, Lanzer M (2017) Differential time-dependent volumetric and surface area changes and delayed induction of new permeation pathways in p. falciparum-infected hemoglobinopathic erythrocytes. Cell Microbiol 19:e12650

Waugh R, Evans EA (1979) Thermoelasticity of red blood cell membrane. Biophys J 26:115–131

Wells R, Schmid-Schönbein H (1969) Red cell deformation and fluidity of concentrated cell suspensions. J Appl Physiol 27:213–217

Wendt JF (ed) (2009) Computational fluid dynamics, 3rd edn. Springer, Berlin

Wu T, Feng JJ (2013) Simulation of malaria-infected red blood cells in microfluidic channels: passage and blockage. Biomicrofluidics 7:044115

Yipp BG, Anand S, Schollaardt T, Patel KD, Looareesuwan S, Ho M (2000) Synergism of multiple adhesion molecules in mediating cytoadherence of Plasmodium falciparum-infected erythrocytes to microvascular endothelial cells under flow. Blood 96:2292–2298

Zhang Y, Huang C, Kim S, Golkaram M, Dixon MWA, Tilley L, Li J, Zhang S, Suresh S (2015) Multiple stiffening effects of nanoscale knobs on human red blood cells infected with Plasmodium falciparum malaria parasite. Proc Natl Acad Sci USA 112:6068–6073

Multiscale Modeling of Sickle Cell Anemia

110

Xuejin Li and Huan Lei

Contents

Abstract

Sickle cell anemia is a genetic blood disorder exhibiting heterogeneous cell morphology, decreased deformability, and abnormal adhesion. The primary pathogenic event of the disease is the intracellular nucleation and polymerization of sickle cell hemoglobin followed by sickling of red blood cells that result in vaso-occlusive crises and other pathophysiological aspects of the disease. In this chapter, we present the morphological, rheological, and adhesive properties of sickle cells and their relevance to the vaso-occlusion under physiological conditions.

X. Li (✉)
Department of Engineering Mechanics, Zhejiang University, Zhejiang, People's Republic of China
Division of Applied Mathematics, Brown University, Providence, RI, USA
e-mail: xuejin_li@zju.edu.cn

H. Lei
Advanced Computing, Mathematics and Data, Pacific Northwest National Laboratory, Richland, WA, USA
e-mail: Huan.Lei@pnnl.gov

Fig. 1 Vaso-occlusive crisis in SCA. The polymerization of HbS molecules causes cell sickling and damage to the cell membrane. The abnormal, sickle-shaped RBCs obstruct microcapillaries and restrict blood flow leading to vaso-occlusion

1 Introduction

Sickle cell anemia (SCA) is a genetic disorder of the blood that draws its name from the abnormal sickled shape of the affected red blood cells (RBCs) (Pauling et al. 1949; Gravitz and Pincock 2014). It afflicts about 100,000 individuals in the United States and millions of people worldwide (Modell and Darlison 2008). In addition, SCA is one of the most common cause of stroke in children. Despite modern health care and technology, the average lifespan of an individual with SCA in the United States still only hits 40–50 years. Patients with SCA have significant morbidity and reduced quality of life as a result of stroke, frequent painful vaso-occlusive crisis events, and other serious clinical complications such as acute chest syndrome (Bunn 1997; Barabino et al. 2010).

The origin of SCA has been traced to the intracellular polymerization of sickle cell hemoglobin (HbS) under deoxygenation conditions (Hofrichter et al. 1974; Noguchi and Schechter 1981), causing a substantial increase in intracellular viscosity and elastic stiffness as well as a severe damage to RBC membrane (Samuel and Briehl 1990; Liu et al. 1991). The affected sickle RBCs (SS-RBCs) become stiff, sticky, and abnormally shaped, so they tend to get stuck in the microcirculation, thereby clogging blood flow (Kaul et al. 1989; Kaul and Fabry 2004) (Fig. 1). Here, we present experimental and computational studies related to morphological, rheological, and adhesive properties of RBCs in SCA.

2 Intracellular Polymerization of Sickle Hemoglobin

SCA is characterized as the first molecular disease (as early as 1947 by Linus Pauling) (Pauling et al. 1949), being linked to the mutation of a single nucleotide, from glutamate to valine, in the hemoglobin molecule. The HbS polymerization has been characterized by a double nucleation mechanism (Ferrone et al. 1985a, b; Vekilov et al. 2008). According to this mechanism, the HbS polymerization process is initiated by homogeneous nucleation in solution and followed by heterogeneous

nucleation on preexisting HbS polymer fibers (Ferrone et al. 1985a, b). A different two-step nucleation mechanism has also been proposed (ten Wolde and Frenkel 1997; Shiryayev and Gunton 2004; Lutsko and Nicolis 2006): first, formation of a dense liquid droplet and, then, formation of HbS fiber nuclei inside the liquid droplet. Consequently, HbS polymer fibers grow spontaneously and distort the RBCs into abnormal sickle shape.

Numerical models provide valuable insights into the molecular mechanism of HbS fiber nucleation and polymerization. Here, we briefly describe a few of the major achievements. A precursory work was presented by Turner and co-workers (2006), who proposed a theoretical model of HbS polymer fibers and demonstrated that twist plays an essential role in stabilizing the polymerized fibers. Yang et al. performed Metropolis Monte Carlo simulations of self-assembled filamentous bundles (Yang et al. 2010). They found that the molecular chirality and attractive interactions can control the assembly size of HbS polymer fibers. Li and Lykotrafitis employed a coarse-grained particle-based HbS model to investigate the mechanical properties of HbS polymer fibers (Li and Lykotrafitis 2011; Li et al. 2012). They demonstrate that HbS polymer fiber frustration and compression play a critical role in fiber zipping and unzipping dynamics.

To provide new details of how SCA manifests inside RBCs, we have developed a mesoscopic patchy particle HbS model to study the polymerization of HbS and the fiber growth dynamics (Lu et al. 2016). In the patchy particle model, an HbS molecule is modeled by a rigid spherical particle with three patches on its surface, representing the locations of double-stranded contact sites among HbS molecules (Fig. 2a). In addition, to model the helical twist effect, we consider an angle-bending potential of two adjacent HbS particles within the same strand (Fig. 2b). The corresponding potential is taken as

$$V_a = k_a(\theta - \theta_0)^2, \quad V_l = k_l(\beta - \beta_0)^2, \tag{1}$$

where k_a and k_l are the bending constant and $\theta = \cos^{-1}(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{p}}_j)$ and $\beta = \cos^{-1}(\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_j)$ are the instantaneous angles between the axial vectors ($\mathbf{p}_i$ and $\mathbf{p}_j$) and lateral vectors ($\mathbf{q}_i$ and $\mathbf{q}_j$) of these two adjacent HbS particles (Fig. 2b), respectively. This HbS model was then employed to simulate the HbS polymerization with pre-existing HbS nuclei (Fig. 2c), showing that the growth dynamics of HbS polymer fiber is achieved by the addition of HbS monomers.

The study of intracellular HbS polymerization and consequent RBC sickling provides an important basis for understanding the implications and treatment routes of SCA; however, modeling the integrated process from HbS fiber nucleation and polymerization to RBC sickling and then to vaso-occlusion is challenging due to the multiple temporal and spatial scales of the entire process. To make progress toward understanding the integrated processes, in a follow-up study, we proposed a hybrid HbS fiber model, by coupling the aforementioned microscopic patchy particle model with a mesoscopic HbS polymer fiber model (Lu et al. 2017). In this proposed multiscale hybrid model, the function of microscopic patchy particle model is to simulate the dynamic polymerization process of HbS molecules,
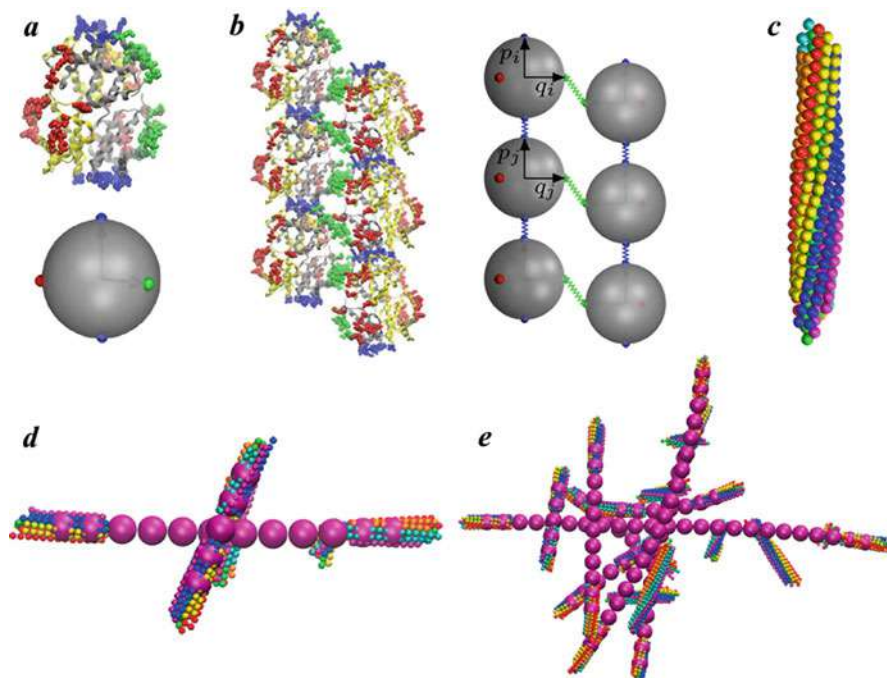
Fig. 2 Mesoscopic modeling of polymerization of HbS. (**a**) Molecular structure of sickle cell hemoglobin (*upper*) and the corresponding patchy particle model (*lower*). Green and blue mark the lateral and axial intra-strand contacts, while red marks the interstrand contacts. (**b**) Double-stranded structure of HbS filament (*left*) with its representation within patchy particle model. (**c**) HbS filament polymerized from HbS particles. (**d**) Interaction of two growing HbS fibers. (**e**) Modeling fiber bundle and domain formation. (Reproduced with permission from Lu et al. 2016, 2017)

whereas the mesoscopic fiber model represents the major body of HbS polymer fiber. In addition, the hybrid HbS fiber model undergoes a dynamics coarse-graining process during HbS polymerization through a mesoscopic adaptive resolution scheme (MARS). As a result, the hybrid method reduces the computational cost by reducing the number of degrees of freedom in the HbS polymer fiber model. Herein, it can be used to simulate the HbS polymer fiber growth and branching process. As an attempt, we applied the hybrid HbS fiber model to simulate the growth dynamics of two interacting HbS polymer fibers (Fig. 2d) and fiber-branching process of multiple HbS polymer fibers (Fig. 2e), which shows that the fiber-fiber orientation and interaction range play key roles in determining the structural and mechanical properties of HbS polymer fibers and filament bundles.

3   Biorheology of Sickle RBCs and Sickle Cell Blood

As shown in the previous section, SS-RBCs exhibit abnormal shapes due to the interaction between intracellular HbS polymer and the cell membrane. Furthermore, SS-RBC suspensions exhibit abnormal rheology and hemodynamics (Usami et al. 1975; Kaul et al. 1983; Kaul and Xue 1991; Kaul and Liu 1999) due to the heterogeneous cell morphology and elevated cell rigidity (Evans et al. 1984; Itoh et al. 1995). In this section, we employ the constructed SS-RBC model to investigate the abnormal rheology and hemodynamics of SS-RBC suspension under shear and tube flow.

Since the collective properties of the SS-RBC suspension is mainly determined by the cell morphology and rigidity, we ignore the modeling of intracellular HbS polymer on molecular level. Alternatively, we directly model the interaction between the intracellular polymer and the cell membrane as the stretching force applied on the cell membrane exerted by the growing HbS fibers. As shown in Fig. 3, the letters "A," "B," "C," and "D" represent four anchor points where the intracellular growing fibers can potentially approach the cell membrane. Depending on the MCHC, the intracellular HbS polymer may exhibit different configurations. Accordingly, as the growing HbS fibers may approach the cell membrane from multiple directions, various stretching forces are applied on the four anchor points. Detail simulation parameters are summarized in Table 1.

With the constructed SS-RBCs of various morphologies, we investigate the abnormal rheology of SS-RBC suspension in shear flow. We first consider the shear flow system with hematocrit Hct = 45% following the previous experiment (Usami et al. 1975). The viscosity of blood plasma is chosen to be $\eta_0 = 1.2$ cp. Figure 4
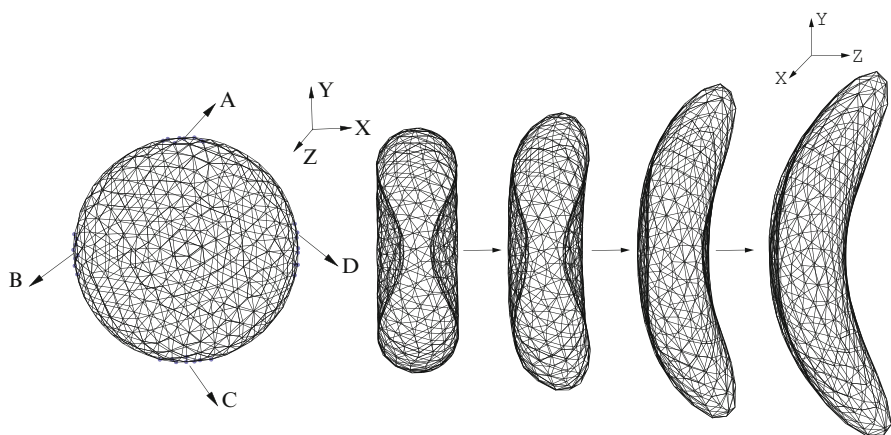
Fig. 3 Left: triangulated mesh of the RBC membrane. The label "A," "B," "C," and "D" represents the four anchor points where the stretching force is applied. Right: successive snapshots of a RBC during the morphological transition to the "sickle" shape (From Lei and Karniadakis 2012)

Table 1 Stretching force (pN) applied on the anchor points for each type of the cell morphology along x, y, and z direction. "A," "B," "C," "D" represent the anchor points sketched in Fig. 3

	A	B	C	D
S	(0, 55, 54)	(0, 0, 0)	(0, −55, 54)	(0, 0, 0)
G	(0, 23, 31)	(−23, 0, 31)	(0 − 23, 31)	(23, 0, 31)
E	(0, 55, 11)	(0, 0, 0)	(0, −55, −11)	(0, 0, 0)

Fig. 4 Shear viscosity of the healthy blood and SS-RBC suspensions with Hct = 45%. The dash lines represent the fitted curve to the simulation result by $\eta = be^{-a/\gamma^{0.5}} + c$, where γ is the shear rate. a, b, and c equal to $1.43\,s^{-1/2}$, $-6.04\,cp$, and $8.78\,cp$ for healthy blood and $1.08\,s^{-1/2}$, $-5.5\,cp$, and $23.9\,cp$ for deoxygenated SS-RBC suspension. The inset plot shows a snapshot of the "granular" SS-RBCs in shear flow (From Lei and Karniadakis 2012)
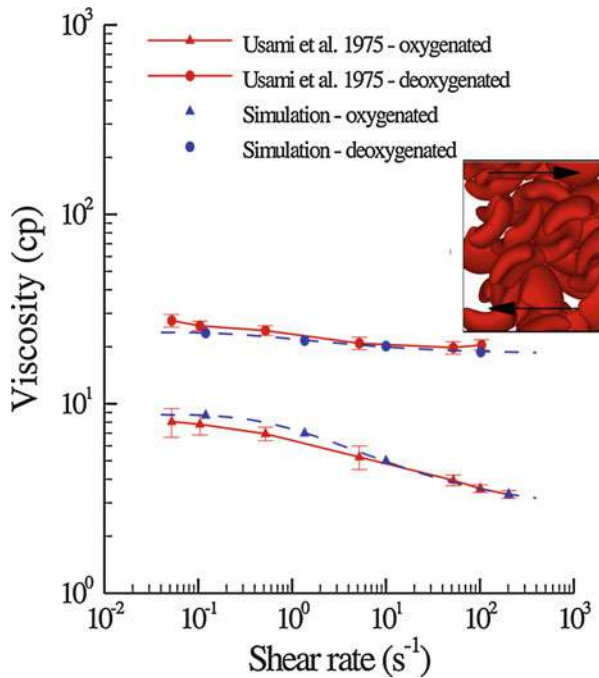
shows the viscosity computed for both healthy and diseased blood under different shear rates. For both types of blood, good agreement with the experimental results is obtained. It is well-known that healthy blood behaves as non-Newtonian fluid with shear-dependent viscosity (Fedosov et al. 2011). Under high shear rate, blood cells undergo large deformation with fluid properties pronounced; in contrast, blood cells behave as solid-like and less deformed under low shear rate conditions, resulting in relatively higher viscosity. Therefore, the viscosity of healthy blood increases as the shear rate decreases, as shown in Fig. 4. On the other hand, the deoxygenated SS-RBC suspension shows elevated viscosity values, which are nearly independent of the shear rate. This result is mainly due to the largely elevated cell rigidity; the deoxygenated SS-RBC cannot be deformed under the high shear rate employed in the experiment (Usami et al. 1975). Therefore, SS-RBCs exhibit "solid" behavior throughout the entire shear rate regime, leading to this transition from non-Newtonian to Newtonian flow.

Furthermore, experimental studies show that the rheology of SS-RBC suspension further depends on the morphology of the SS-RBCs. Kaul et al. investigated the shear viscosity of SS-RBC suspensions subjected to both fast and gradual deoxygenation procedures (Kaul and Xue 1991) that were measured. Two distinct stages were observed for SS-RBC suspensions subjected to the fast deoxygenation. The shear viscosity of the SS-RBC suspensions exhibits fast elevation within the first 7 min of deoxygenation accompanied with the cell morphology transition to granular shape. However, the shear viscosity gradually decreased during increased deoxygenation, accompanied with a large portion of cells transitioned into extremely elongated shape with the intracellular HbS fibers aligned in one direction. In contrast, SS-RBC suspensions subjected to gradual deoxygenation procedure showed monotonic elevation of shear viscosity and the formation of the sickle shape of blood cells over a period of 30 min until the full deoxygenated state was achieved. Following this experimental study, we computed the shear viscosity of SS-RBC suspensions with the three distinct types of sickle cell reported in the experiment (Hct = 40%). Figure 5 shows the shear viscosity values under shear rate from 25 to 75 s^{-1} with similar cell rigidity applied to all the three types. Similar to Fig. 4, the SS-RBC suspensions show shear-independent Newtonian behavior. Moreover, shear viscosity further depends on individual cell shapes, with the value 13.8, 13.1, and 9.6 cp for the granular, sickle, and elongated shape, respectively. The simulation results are consistent with the experimentally observed progressive decrease of the viscosity during the increased deoxygenation, since a large portion of granular
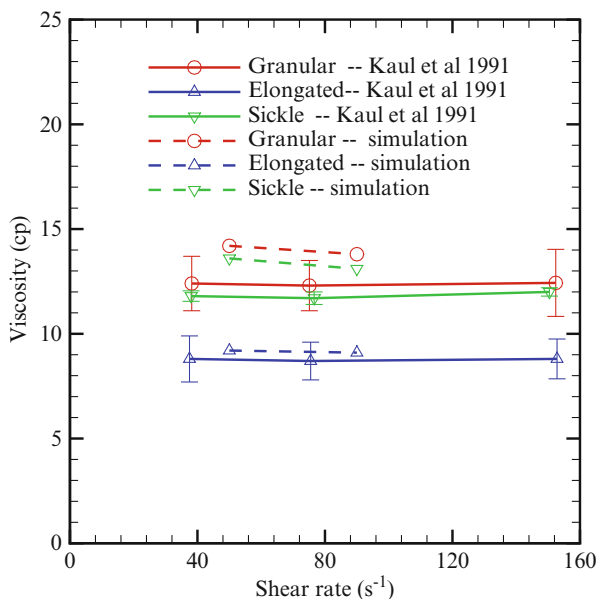
Fig. 5 Shear viscosity of the sickle blood flow with different cell morphologies reported in Kaul and Xue (1991), Hct = 40% (From Lei and Karniadakis 2012)

cell transforms into the elongated shape during the procedure. This result reveals the heterogeneous "effective volume" among the different types of SS-RBC under the shear flow system (Kaul and Xue 1991). Given the same Hct, heterogeneous cell morphologies may further affect the momentum transport between the cells, resulting in different shear viscosity values.

Next, we further examine the hemodynamics of SS-RBC suspensions in an isolated vasculature with different cell groups. While the oxygenated SS-RBC suspensions exhibit hemodynamics similar to healthy blood flow, the deoxygenated SS-RBC suspensions show heterogeneous hemodynamics among the different cell groups. Similar to the experimental study (Kaul et al. 1983), we simulate SS-RBC suspensions in a tube flow system with Hct = 30%. We set the diameter of the tube to $9.0\,\mu m$ as in capillary flow, since the detailed size and topology information of the microvasculature for the experiment are unknown. In this sense, we do not expect the apparent viscosity obtained from the simulation to match exactly with the experiment results. Instead, we focus on the effect of different SS-RBCs groups on the flow resistance in the microcirculation.

Deoxygenated blood flow is modeled as a suspension of SS-RBCs with sickle and granular shapes, where the same cell rigidity is applied to both cell groups. The viscosity of the cytosol is set to $4\eta_0$ and $50\eta_0$ for the healthy and deoxygenated blood flow, where $\eta_0 = 1.2\,cp$ is the viscosity of the blood plasma. The increase of the flow resistance for the sickle and granular shapes under different oxygen tensions is shown in Fig. 6a. While SS-RBC suspension of both cell groups shows further increased flow resistance at deoxygenated state, the granular type of blood flow shows a more pronounced elevation compared with the sickle shape. This result is also consistent with the experimental study of SS-RBCs transiting in a micro-fluidic channel conducted by Du et al. (2015). Figure 6b shows the transit velocities of individual SS-RBCs with different cell morphologies (discocyte, oval, sickle) through a periodic array of obstacles ($4\,\mu m$ between the two obstacles). The cell group of the "sickle" shape exhibits the largest cell velocity.

One possible explanation for the above result, as proposed by Kaul et al., is the different orientational distribution of SS-RBCs in the capillary: a cell of sickle shape tends to flow along the axis of the tube as also observed by La Celle et al. in experimental studies in LaCelle (1977). This is consistent with the inset snapshots of the sickle and granular cells in Fig. 6a. To quantify this phenomenon, we computed the *cell orientation angle distribution* for different cell groups in the tube flow, as shown in Fig. 7. The cell orientation is defined by the angle θ between the flow direction and the eigenvector of the gyration tensor V_1 defined by

$$G_{mn} = \frac{1}{N_v} \sum_i (r_m^i - r_m^C)(r_n^i - r_n^C), \tag{2}$$

where r^i is the coordinate of RBC vertex i, r^C is the center-of-mass coordinate, and m and n represent the x, y, or z direction. λ_1, λ_2, and λ_3 represent the three eigenvalues obtained from the gyration tensor, where $\lambda_1 < \lambda_2 < \lambda_3$. Compared with the granular cell, the orientation angle of the sickle cell shows a wider distribution
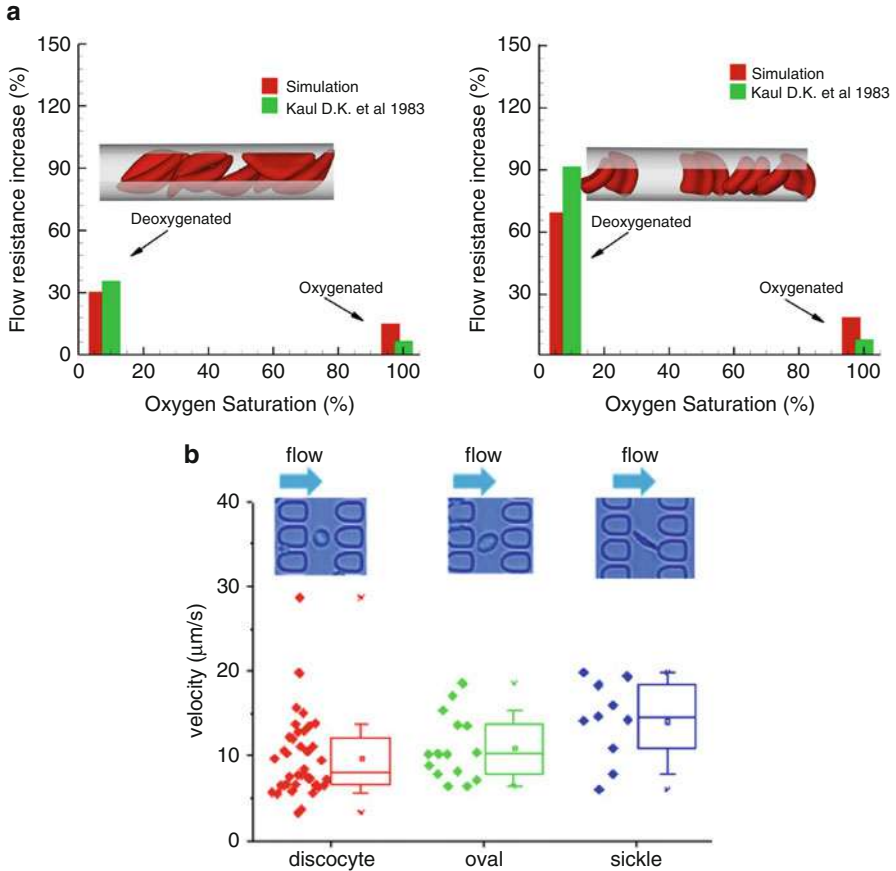
Fig. 6 Top: increase of the flow resistance induced by the sickle blood flow for both granular and sickle shapes. The inset plot shows a snapshot of the sickle cells in the tube flow. Bottom: experimental study on SS-RBC cell transition velocity for different cell morphologies through a periodic array of obstacles in the microchannel, provided by Du and Dao at MIT (From Lei and Karniadakis 2012)

for larger value, indicating that the sickle-shaped SS-RBCs are more likely to orient along the flow direction, which results in a lower flow resistance compared with the granular cells.

Finally, we note that the simulation of SS-RBC suspension in tube flow does not induce blood occlusion in the present study. We performed several sets of simulations with different combinations of cell rigidities and cytosol viscosity; however, no full occlusion was observed in any of these simulations. Moreover, the sickle-shaped SS-RBC, the widely believed precursor of the vaso-occlusion crisis, actually may result in lower flow resistance than other cell groups. We note that this is mainly due to the omission of the cell-endothelium adhesive interaction in the present study; we address this issue in the next section.

Fig. 7 Cell orientation angle distribution $f(\theta)$ for healthy, sickle, and granular cells in pipe flow. The cell orientation is defined by the angle θ between the flow direction (x) and the eigenvector $V1$ of the gyration tensor, as shown in the inset plot (From Lei and Karniadakis 2012)
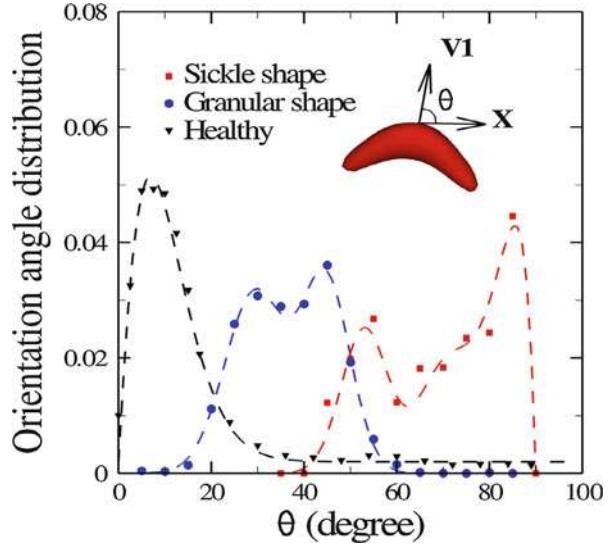
4 Adhesive Properties of Sickle RBCs

Vaso-occlusion crisis may trigger serious stroke or death for SCA patients. Early experimental studies suggested that this crisis is triggered by the blockage of single sickle/elongated SS-RBC in capillaries; later studies (Kaul et al. 1989, 1994; Kaul and Fabry 2004; Barabino et al. 2010) have shown that vaso-occlusion mainly occurs in post-capillaries (the microvasculatures following the capillaries). Moreover, some studies demonstrated that there exists no direct correlation between the percentage of the dense SS-RBCs and the disease severity (Ballas et al. 1988). Recent studies further revealed that vaso-occlusion is a complex process triggered by the interactions between multiple cell groups (Ballas and Mohandas 2004; Chiang and Frenette 2005; Turhan et al. 2002). In this section, we employ computational approach to explore the mechanism of the vaso-occlusion in SCA.

4.1 Adhesive Dynamics

In vitro studies by Barabino et al. (1987) show that SS-RBCs exhibit heterogeneous cell adhesivity among different density groups. The light density group shows the largest adhesion, while the densest irreversible sickle cells (ISCs) show the least adhesion. Ex vivo studies by Kaul et al. (1994) further examined the alteration of SS-RBC adhesivity after a dehydration/rehydration treatment on individual cell groups and found that the cell adhesivity of the deformable SS2 and the dense SS4 cells can be reversed after controlled treatment. They suspected that different cell groups have similar "adhesion potential," while the heterogeneous cell adhesivity is
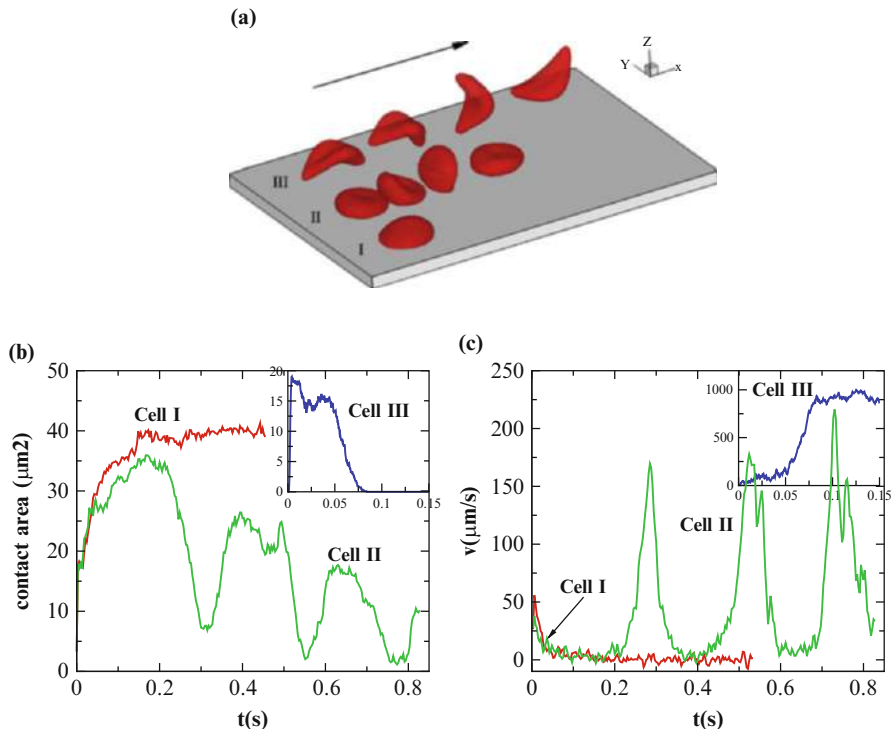
Fig. 8 Sickle cells in shear flow: (**a**) successive snapshots of SS-RBCs in shear flow. The labels I, II, and III correspond to a deformable SS2 cell, rigid SS3 cell, and ISC, respectively. The arrow indicates the flow direction. (**b–c**) Instantaneous contact area and velocity for SS-RBC in shear flow conditions (From Lei and Karniadakis 2013)

mainly attributed to the different cell morphological and biomechanical properties among the multiple cell density group. Accordingly, we investigate this hypothesis by simulating the adhesive dynamic of SS-RBCs with different cell rigidities and morphologies.

As shown in Fig. 8, we consider three different SS-RBCs under shear flow. Cell I represents a SS2 deformable discocyte. The cell rigidity is similar to the healthy RBC, and we set the shear modulus $\mu_0 = 6.8\,\mu N/m$ and bending rigidity $k_{c0} = 2.4 \times 10^{-19}$ J, i.e., similar to healthy cell rigidity (Hupert and Baumann 2003; Mills et al. 2004; Suresh et al. 2005). Cell III represents an ISC generated through the deoxygenation-reoxygenation process; we set the shear modulus $\mu = 10\,\mu_0$ following (Itoh et al. 1995; Clark et al. 1980). Modification to cell bending rigidity in the deoxygenation-reoxygenation process is unknown; here we set $k_c = 10k_{c0}$ and also conduct sensitivity studies. Cell II represents a rigid discocyte (Evans and Mohandas 1987) with medium MCHC value; we set $\mu = 3\,\mu_0$ and $k_c = 3k_{c0}$ for the comparative study.

Following Kaul et al. (1994), we assume that the three cells have similar "adhesive potential" and set identical adhesive parameters. With the same shear rate $\dot{\gamma} = 192\,\text{s}^{-1}$, the three cells exhibit substantially different adhesive dynamics as shown in Fig. 8a. Cell I exhibits firm adhesion to the lower plate with contact area around $40.5\,\mu\text{m}^2$. Cell II, however, shows weaker adhesivity than cell I. Although it also exhibits transient adhesion to the lower plate initially, it undergoes a periodic flip movement along the flow direction and eventually detaches from the plate after two to three flips, as characterized by the peak values of the instantaneous cell velocity at 0.30, 0.53, and 0.72 s. Accordingly, the contact area achieves minimum values at those times. Different from cells I and II, cell III does not show any adhesion to the plate; instead, it directly detaches from the lower plate and moves freely without adhesive bonds established thereafter. Given the same "adhesive potential," the present results validate the hypothesis that heterogeneous cell adhesive dynamics is mainly due to the different cell rigidities and peculiar cell morphologies. To further investigate this effect, we use free energy analysis to quantify the cell adhesion in static condition.

We compute the instantaneous contact area for each cell until steady state is reached, as shown in Fig. 9a. While the contact area for all of the cells increases sharply to $10\,\mu\text{m}^2$ within the initial stage, the contact area between the plate and cells I and II further increases to 30 and $21\,\mu\text{m}^2$ at the later stage. This inverse relationship is consistent with the different cell adhesive dynamics in shear flow system, and it can be understood by the free energy analysis during the adhesion process. If we define the cell and the ligand particles as a single system, the change of total free energy ΔE during the process can be written as

$$\Delta E = \Delta E_{\text{deform}} - \Delta E_{\text{adhesion}}, \tag{3}$$

where ΔE_{deform} represents the increase of the cell free energy due to the deviation of cell shape from the equilibrium state. $\Delta E_{\text{adhesion}}$ represents the absolute value of the energy decrease due to the adhesive bond formation between the cell and the ligand particles. The final state is determined by the balance between the two free energy terms. A deformable SS2 cell is prone to form larger contact area than the rigid cell; it can be understood as follows: for SS2 cell with smaller cell rigidity, the energy barrier induced by the cell deformation is relatively small, and free energy decrease induced by cell adhesion plays a dominant role, resulting in further cell deformation and larger cell contact area. In contrast, the rigid SS4 cell exhibits "solid"-like properties with a larger energy barrier for cell deformation. The adhesive interaction between the cell and plate is more like the attraction between two solid objects where the cell deformation plays a less important role. After the initial incubation stage, the adhesive interaction driven by the bond formation cannot overcome the free energy increase induced by the cell deformation, and therefore it prohibits further increase of the contact area.

Figure 9b shows the ΔE_{deform} as a function of the contact area ΔA for the three cells induced by different adhesive affinities. The numerical results are fitted by $\alpha\Delta A^4$, where α depends on the cell rigidity of individual cells. The adhesion free
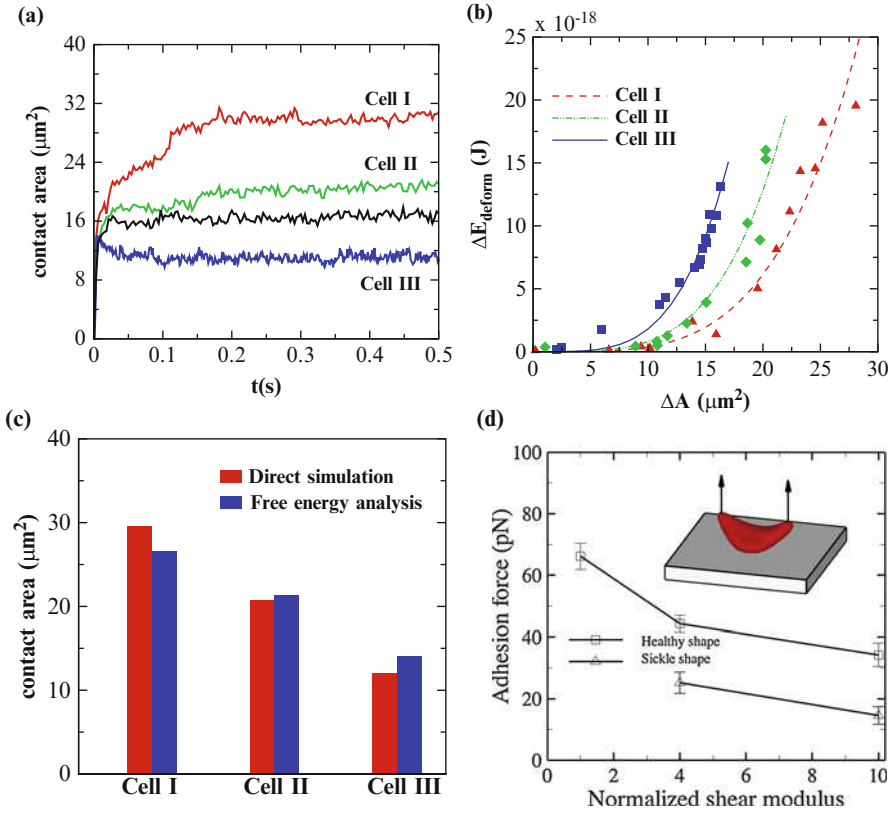
Fig. 9 Adhesion in static conditions: (**a**) instantaneous contact area between the SS-RBC and the plate coated with adhesive ligands. The black curve represents the contact area of a discocyte with cell rigidity similar to the ISC (cell III). (**b**) Increase of the cell free energy as a function of the contact area. (**c**) Cell contact area ΔA computed from direct simulation and free energy analysis by Eq. (4). (**d**) Adhesive force between the cell and the plate as a function of the membrane rigidities for cell morphologies. The error bar represents the adhesive force computed from four independent simulations. The inset plot shows a sketch of the simulation setup, where a uniform lift force is applied on the upper part of an ISC (From Lei and Karniadakis 2013)

energy $\Delta E_{\text{adhesion}}$ can be approximated by $\beta \Delta A$, where β is the energy coefficient determined by the adhesive affinity. In the present work, the adhesive interaction is modeled by the adhesive bond interaction between ligands coated on the substrate and the receptors on the cell membrane. While each receptor on the cell membrane can form multiple adhesive bonds with ligands, each ligand can only form one adhesive bond with the receptor. Therefore, β can be approximated by $k_s l_s^2 n_{\text{ligand}}$, where k_s is the adhesive spring bond constant, l_s is the cutoff length of the adhesive bond interaction and n_{ligand} is number density of the ligands on the substrate. The equilibrium state is determined as the state with minimum free energy term, e.g.,

$$\Delta A = \left(\frac{\beta}{4\alpha} \right)^{1/3}.$$ (4)

According to Fig. 9b, α is about 3.8×10^{25}, 8.0×10^{25}, and 1.8×10^{26} J/m^8 for cells I, II, and III. The contact area $A = A_0 + \Delta A$ yields approximately 26.5, 21.3, and 14.0 μm^2 for cells I, II, and III, where A_0 is the initial contact area without any adhesion, which is 3.5 μm^2 for cells I and II and 0.6 μm^2 for cell III. This result agrees well with the direct simulation results of 29.5, 20.7, and 12.0 μm^2, as shown in Fig. 9c. Compared with the shear flow system, the contact area of cell I computed in static condition is smaller. This discrepancy is mainly due to the increased cell deformation induced by the hydrodynamic interaction under shear flow condition. The extended cell membrane facilitates the bond formation and results in larger contact area. For cell III, we note that the contact area is smaller than the value for a discocyte with similar cell rigidity. This result indicates that the peculiar cell morphology may further affect the cell adhesivity.

To quantify the effects discussed above, we directly compute the adhesive force between the plate and SS-RBCs with different cell rigidities and morphologies. The adhesive force is determined as the lift force that detaches the cell from the plate in quasi-static process, as shown in Fig. 9d. Similar to the contact area, the adhesive force also exhibits an inverse relationship with the cell rigidity. Moreover, compared with the discocyte, the ISC exhibits smaller adhesive force given the similar cell rigidity, indicating less adhesivity induced by its peculiar cell morphology.

4.2   Vaso-Occlusion Crisis

We model blood circulation of SS-RBC suspensions in post-capillaries as in a tube with diameter $D = 10$ μm and hematocrit Hct = 30% similar to Kaul et al. (1989) and Kaul et al. (1994). As shown in Fig. 10, small green particles represent the adhesive ligands which can interact with the SS-RBCs. To quantify the distinct role of different cell groups, we infuse suspensions of different cell groups into the tube by applying a pressure gradient $\Delta P / \Delta x = 8.7 \times 10^4$ Pa/m.

First, we consider suspensions composed of SS2 cells (labeled by blue) and ISC cells (labeled by red) with the ratio 1 : 1. We prepare the steady flow by turning off the adhesive interaction. The mean flow velocity at the initial stage is around 150 μm/s. Starting from the steady state ($t = 0$ s), we turn on the adhesive interaction between the SS-RBCs and the ligand particles and compute the instantaneous flow velocity across the tube, as represented by the red curve in Fig. 10. Blood flow maintains steady state until one of the SS2 cells adheres to the vessel wall, triggering a sharp decrease of blood flow around $t = 0.5$ s. As a positive feedback, the decreased blood flow induces more SS2 cells adherent to the vessel wall, leading to a further decrease of flow rate at $t = 0.63$ and $t = 0.7$ s. Moreover, these adherent cells decrease the effective vessel lumen near the adherent region resulting in a secondary trapping of the ISC groups.
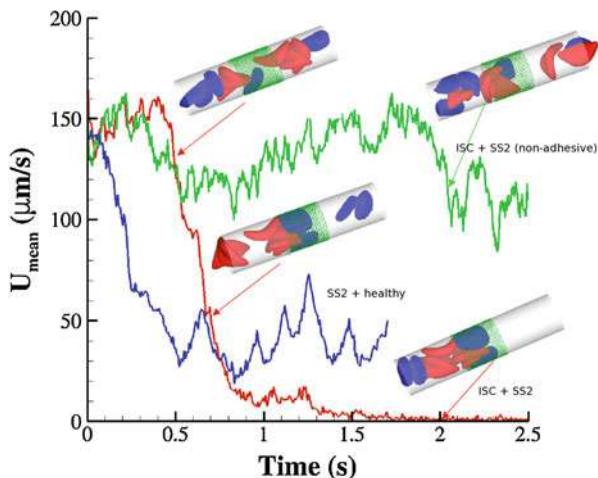
Fig. 10 Vaso-occlusion in post-capillaries: instantaneous mean velocity of blood flow in a cylindrical tube of $D = 10\,\mu m$ infused with different SS-RBC suspensions. The red curve represents the resultant velocity infused with SS2 and ISC cell groups. The inset plots represent the instantaneous snapshots where SS2 cells adhere to the vessel wall, consequently trapping the ISCs and resulting in cell blockage. The green curve represents the blood velocity infused with SS2 and ISC cell groups, where adhesive interaction is only applied to the ISC group. The time axis is scaled by 0.5 for better visualization. The inset plot represents a snapshot where transient adhesion is established between ISC and the tube wall. Steady flow is recovered as the cell detaches from the tube wall. The blue curve represents the instantaneous velocity of blood flow infused with SS2 and healthy cell groups. Blood flow exhibits a slowdown but not a full occlusion (From Lei and Karniadakis 2013)

The final occlusion state is reached around $t \approx 2\,s$ with cell patterns similar to the experimental observations (Kaul et al. 1989, 1994). This result demonstrates that under physiological conditions similar to microcirculation in post-capillaries, the *interplay* of deformable SS2 cells and ISCs can potentially trigger full blood occlusion. Although the same adhesive parameters are applied to the two cell groups, no adherent ISC is observed in the present simulations. Several independent sets of simulation were conducted with different initial conditions, and similar patterns of full occlusion were observed.

To identify the unique contribution of the deformable SS2 cell group, we performed a similar simulation for the suspension consisting of ISC and healthy RBCs. Starting from the steady flow at $t = 0\,s$, we compute the instantaneous velocity of the blood suspension for 6 s, as represented by the green line in Fig. 10. Blood flow shows decreased velocity due to the ISC-ligand interaction, where transient adhesive bonds can be formed. However, blood occlusion is not observed in this simulation since blood flow can recover the initial flow rate when the adherent ISCs detach from the tube wall, as consistently observed in ex vivo studies (Kaul et al. 1994), where the ISCs, when infused alone, did not result in microvascular blockage. This

result reveals the specific contribution of the SS2 cell group in the vaso-occlusion crisis and is consistent with clinical investigation results (Ballas et al. 1988).

In contrast, the ISC cell group contributes differently to the occlusion crisis. We investigate its unique contribution by simulating the blood suspension mixed with the deformable SS2 and healthy cells. Starting from the steady flow at $t = 0$ s, we compute the instantaneous velocity of the blood flow, as represented by the blue line in Fig. 10. Although blood flow slows down due to the cell adhesion at $t = 0.25$ and $t = 0.48$ s, full occlusion is not observed in the present simulation. In particular, we find that healthy RBCs can squeeze through the sievelike region formed by adherent SS2 cells due to its high deformability. The present result, in turn, identifies the specific contribution of the ISC group in the vaso-occlusion crisis. Although the ISC group may not participate in the cell adhesive interaction with the vascular wall, it serves as the particular cell group trapped by adherent cells in the post-capillaries.

Besides the cell interaction among the SS-RBCs, recent studies (Kaul and Hebbel 2000) have shown that SCA is often accompanied with an inflammatory endothelial phenotype with elevated leukocyte recruitment in blood circulation, where the inflammation-stimulated leukocytes may further interact with SS-RBC, leading to blood occlusion in venular flow. We refer to Turhan et al. (2002) and Lei and Karniadakis (2013) for details.

5   Summary

In this chapter, we overview the pathogenesis of vaso-occlusive crisis in sickle cell anemia (SCA) that involves several processes across multiple time and length scales. We cover the morphological, rheological, and adhesive abnormalities of sickle red blood cells and discuss individual and environmental factors underlying the process of SCA vaso-occlusion. The genetic basis and molecular causes of the disease have been understood for more than half a century; however, progress in developing treatments to prevent painful vaso-occlusion and the other myriad of associated symptoms has been slow. Therefore, the need to develop new therapeutic interventions and treatment strategies for SCA remains paramount.

References

Ballas SK, Mohandas N (2004) Sickle red cell microrheology and sickle blood rheology. Microcirculation 11:209–225

Ballas SK, Larner J, Smith ED, Surrey S, Schwartz E, Rappaport EF (1988) Rheologic predictors of the severity of the painful sickle cell crisis. Blood 72:1216–1223

Barabino GA, McIntire LV, Eskin SG, Sears DA, Udden M (1987) Endothelial cell interactions with sickle cells, sickle cell, sickle trait, mechanically injured, and normal erythrocytes under controlled flow. Blood 70:152–157

Barabino GA, Platt MO, Kaul DK (2010) Sickle cell biomechanics. Annu Rev Biomed Eng 12:345–367

Bunn HF (1997) Pathogenesis and treatment of sickle cell disease. N Engl J Med 337:762–769

Chiang EY, Frenette PS (2005) Sickle cell vaso-occlusion. Hematol Oncol Clin North Am 19: 771–784

Clark MR, Mohandas N, Shohet SB (1980) Deformability of oxygenated irreversibly sickled cells. J Clin Invest 65:189–196

Du E, Diez-Silva M, Kato GJ, Dao M, Suresh S (2015) Kinetics of sickle cell biorheology and implications for painful vasoocclusive crisis. Proc Natl Acad Sci USA 112:1422–1427

Evans E, Mohandas N (1987) Membrane-associated sickle hemoglobin: a major determinant of sickle erythrocyte rigidity. Blood 70:1443–1449

Evans E, Mohandas N, Leung A (1984) Static and dynamic rigidities of normal and sickle erythrocytes. major influence of cell hemoglobin concentration. J Clin Invest 73:477–488

Fedosov DA, Pan W, Caswell B, Gompper G, Karniadakis GE (2011) Predicting human blood viscosity in silico. Proc Natl Acad Sci USA 108(29):11772–11777

Ferrone FA, Hofrichter J, Eaton WA (1985a) Kinetics of sickle hemoglobin polymerization. I. Studies using temperature-jump and laser photolysis techniques. J Mol Biol 183:591–610

Ferrone FA, Hofrichter J, Eaton WA (1985b) Kinetics of sickle hemoglobin polymerization. II. A double nucleation mechanism. J Mol Biol 183:611–631

Gravitz L, Pincock S (2014) Sickle-cell disease. Nature 515:S1

Hofrichter H, Ross PD, Eaton WA (1974) Kinetics and mechanism of deoxyhemoglobin s gelation: a new approach to understanding sickle cell disease. Proc Natl Acad Sci USA 71:4864–4868

Hupert C, Baumann M (2003) Local membrane curvature affects spontaneous membrane fluctuation characteristics. Mol Membr Biol 20:155–162

Itoh T, Chien S, Usami S (1995) Effects of hemoglobin concentration on deformability of individual sickle cells after deoxygenation. Blood 85:2245–2253

Kaul DK, Fabry ME (2004) In vivo studies of sickle red blood cells. Microcirculation 11:153–165

Kaul DK, Hebbel RP (2000) Hypoxia/reoxygenation causes inflammatory response in transgenic sickle mice but not in normal mice. J Clin Invest 106:411–420

Kaul DK, Liu X (1999) Rate of deoxygenation modulates rheologic behavior of sickle red blood cells at a given mean corpuscular hemoglobin concentration. Clin Hemorheol Microcirc 21:125–135

Kaul DK, Xue H (1991) Rate of deoxygenation and rheologic behavior of blood in sickle cell anemia. Blood 77:1353–1361

Kaul DK, Fabry ME, Windisch P, Baez S, Nagel RL (1983) Erythrocytes in sickle-cell-anemia are heterogeneous in their rheological and hemodynamic characteristics. J Clin Invest 72:22–31

Kaul DK, Fabry ME, Nagel RL (1989) Microvascular sites and characteristics of sickle cell adhesion to vascular endothelium in shear flow conditions: pathophysiological implications. Proc Natl Acad Sci USA 86:3356–3360

Kaul DK, Chen D, Zhan J (1994) Adhesion of sickle cells to vascular endothelium is critically dependent on changes in density and shape of the cells. Blood 83:3006–3017

LaCelle PL (1977) Oxygen delivery to muscle cells during capillary vascular occlusion by sickle erythrocytes. Blood Cells 3:263–272

Lei H, Karniadakis GE (2012) Quantifying the rheological and hemodynamic characteristics of sickle cell anemia. Biophys J 102:185–194

Lei H, Karniadakis GE (2013) Probing vasoocclusion phenomena in sickle cell anemia via mesoscopic simulations. Proc Natl Acad Sci USA 110:11326–11330

Li H, Lykrafitis G (2011) A coarse-grain molecular dynamics model for sickle hemoglobin fibers. J Mech Behav Biomed Mater 4:162–173

Li H, Ha V, Lykrafitis G (2012) Modeling sickle hemoglobin fibers as one chain of coarse-grained particles. J Biomech 45:1947–1951

Liu SC, Derick LH, Zhai S, Palek J (1991) Uncoupling of the spectrin-based skeleton from the lipid bilayer in sickled red cells. Science 252:574–576

Lu L, Li H, Bian X, Li X, Karniadakis GE (2017) Mesoscopic adaptive resolution scheme toward understanding of interactions between sickle cell fibers. Biophys J 113:48–59

Lu X, Wood DK, Higgins JM (2016) Deoxygenation reduces sickle cell blood flow at arterial oxygen tension. Biophys J 110:2751–2758

Lutsko JF, Nicolis G (2006) Theoretical evidence for a dense fluid precursor to crystallization. Phys Rev Lett 96:46102

Mills JP, Qie L, Lim CT, Dao M, Suresh S (2004) Nonlinear elastic and viscoelastic deformation of the human red blood cell with optical tweezers. Mech Chem Biosyst 1(3):169–180

Modell B, Darlison M (2008) Epidemiology of haemoglobin disorders and derived service indicators. Bull World Health Organ 86:480–487

Noguchi CT, Schechter AN (1981) The intracellular polymerization of sickle hemoglobin and its relevance to sickle cell disease. Blood 58:1057–1069

Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia, a molecular disease. Science 110:543–548

Samuel RE, Briehl RW (1990) Nucleation and growth of fibres and gel formation in sickle cell hemoglobin. Nature 345:833–835

Shiryayev A, Gunton JD (2004) Crystal nucleation for a model of globular proteins. J Chem Phys 120:8318–8326

Suresh S, Spatz J, Mills JP, Micoulet A, Dao M, Lim CT, Beil M, Seufferlein T (2005) Connections between single-cell biomechanics and human disease states: gastrointestinal cancer and malaria. Acta Biomater 1:15–30

ten Wolde PR, Frenkel D (1997) Enhancement of protein crystal nucleation by critical density fluctuations. Science 277:1975–1978

Turhan A, Weiss LA, Mohanda N, Coller BS, Frenette PS (2002) Primary role for adherent leukocytes in sickle cell vascular occlusion: a new paradigm. Proc Natl Acad Sci USA 99:3047–3051

Turner M, Briehl R, Wang J, Ferrone F, Josephs R (2006) Anisotropy in sickle hemoglobin fibers from variations in bending and twist. J Mol Biol 357:1422–1427

Usami S, Chien S, Scholtz PM, Bertles JF (1975) Effects of deoxygenation on blood rheology in sickle cell disease. Microvasc Res 9:324–334

Vekilov P, Galkin O, Pettitt B, Choudhury N, Nagel R (2008) Determination of the transition-state entropy for aggregation suggests how the growth of sickle cell hemoglobin polymers can be slowed. J Mol Biol 377:882–888

Yang YS, Meyer RB, Hagan MF (2010) Self-limited self-assembly of chiral filaments. Phys Rev Lett 104:258102

Multiscale Modeling of Blood Flow-Mediated Platelet Thrombosis

111

Alireza Yazdani, Peng Zhang, Jawaad Sheriff, Marvin J. Slepian, Yuefan Deng, and Danny Bluestein

Contents

A. Yazdani
Division of Applied Mathematics, Brown University, Providence, RI, USA
e-mail: alireza_yazdani@brown.edu

P. Zhang (✉) · J. Sheriff · D. Bluestein
Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA
e-mail: peng.zhang@stonybrook.edu; pzhang99@gmail.com; jawaad.sheriff@stonybrook.edu; danny.bluestein@stonybrook.edu

Y. Deng
Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA
e-mail: yuefan.deng@stonybrook.edu

M. J. Slepian
Department of Medicine, Biomedical Engineering Department, University of Arizona, Tucson, AZ, USA
e-mail: slepian@email.arizona.edu

Abstract

The blood coagulation cascade that leads to thrombus formation may be ini-
tiated by flow-induced platelet activation, which prompts clot formation in
prosthetic cardiovascular devices and in arterial disease processes. Upon activa-
tion, platelets undergo complex morphological changes of filopodia formation
that play a major role in aggregation and attachment to surfaces. Numerical
simulations based on continuum approaches fail to capture such molecular-scale
mechano-transduction processes. Utilizing molecular dynamics (MD) to model
these complex processes across the scales is computationally prohibitive. We
describe multiscale numerical methodologies that integrate four key components
of blood clotting, namely, blood rheology, cell mechanics, coagulation kinetics
and transport of species, and platelet adhesive dynamics across a wide range
of spatiotemporal scales. Whereas mechanics of binding/unbinding for single-
molecule receptor-ligand complexes can be simulated by molecular dynamics
(MD), the mechanical structure of platelets in blood flow and their interaction
with flow-induced stresses that may lead to their activation can be efficiently
described by a model at coarser scales, using numerical approaches such as
coarse-grained molecular dynamics (CGMD). Additionally, CGMD provides
an excellent platform to inform other coarser-scale models in a bottom-up
approach in the multiscale hierarchy. The microenvironment of most biological
systems such as coagulation normally involves a large number of cells, e.g.,
blood cells suspended in plasma, limiting the utility of CGMD at the larger
transport scales of blood flow. However, dissipative particle dynamics (DPD),
along with its sub-models such as energy conserving and transport DPD,
provides a very flexible platform for scaling up these mesoscopic systems.
At the macroscopic top scales of the vasculature and cardiovascular devices,
simulating blood and tissues using continuum-based methods becomes viable
and efficient. However, the challenge of interfacing these larger transport scales
with the orders of magnitude smaller spatiotemporal scales that characterize
blood coagulation, and given the issue of the slow-dynamic timescales of bio-
logical processes, makes long-term simulations of such systems computationally
prohibitive. In this chapter, we describe various numerical remedies based
on these methodologies that facilitate overcoming this multiscale simulation
challenge.

1 Introduction

Hemostasis (thrombus formation) is the normal physiological response that prevents significant blood loss after vascular injury. The resulting clots can form under different flow conditions in the veins as well as the arteries. The excessive and undesirable formation of clots (i.e., thrombosis) in our circulatory system may lead to significant morbidity and mortality. Some of these pathologies are deep vein thrombosis and pulmonary embolism and atherothrombosis (thrombosis triggered by plaque rupture) in coronary arteries, to name a few. The process of clot formation and growth at a site on a blood vessel wall involves a number of simultaneous processes including multiple chemical reactions in the coagulation cascade, species transport, and platelet adhesion, all of which are strongly influenced by the hydrodynamic forces.

Platelets are fundamental to both hemostasis and thrombosis in many vascular diseases. Normal platelets do not interact with the healthy arterial wall. In cases of endothelial injury or exposure of extracellular matrix to blood flow, however, platelets can quickly activate and cover the injured area to stop bleeding. The initial adhesion of platelets on the thrombogenic area can be attributed to a variety of platelet membrane receptor-ligand interactions, such as glycoprotein Ib (GPIb)-V-IX with immobilized von Willebrand factor (vWF), GPIIb–IIIa ($\alpha_{\text{IIb}}\beta_3$) with vWF, GPVI with collagen, $\alpha_2\beta_1$ with collagen, $\alpha_{\text{IIb}}\beta_3$ with fibrinogen, and so on, depending on the nature of the lesion and the local shear rate of blood flow. At low shear rates ($\dot{\gamma} < 1000\,\text{s}^{-1}$), platelets adhere to the thrombogenic area through different pathways, relying on the exposed extracellular matrix (ECM) proteins. On the other hand, as shear rate increases, interactions between immobilized vWF and GPIb become exclusive in initializing platelet aggregation, while other interactions are broken down due to high bond failure rates. The reason that vWF-GPIb interactions persist at such high shear rates ($25{,}000\,\text{s}^{-1}$ shown in in vitro experiments) is that the vWF proteins, which are normally in a coiled state, tend to extend severalfold in high-shear environments. The conformational change of vWF exposes the repeating functional A1 domains in multimeric vWF, leading to enhanced adhesive interactions between GPIb and vWF (Schneider et al. 2007) (Fig. 1).

While platelet activation may be induced by biochemical agonists, shear stresses arising from pathological flow patterns enhance the propensity of platelets to activate and initiate coagulation cascade, causing thrombosis. Flow stresses acting on the platelets can be represented down to the order of microns by continuum models. However, processes such as the morphological changes in platelets upon activation happen at the molecular scales. Continuum approaches fail to capture such processes. Mesoscopic particle-based methods were developed to model blood cells and thrombosis in flowing blood, using dissipative particle dynamics (DPD) for modeling the blood, blood flows in microvessel, interactions of cells, and thrombosis growth. However, complex processes such as platelet activation require drastically
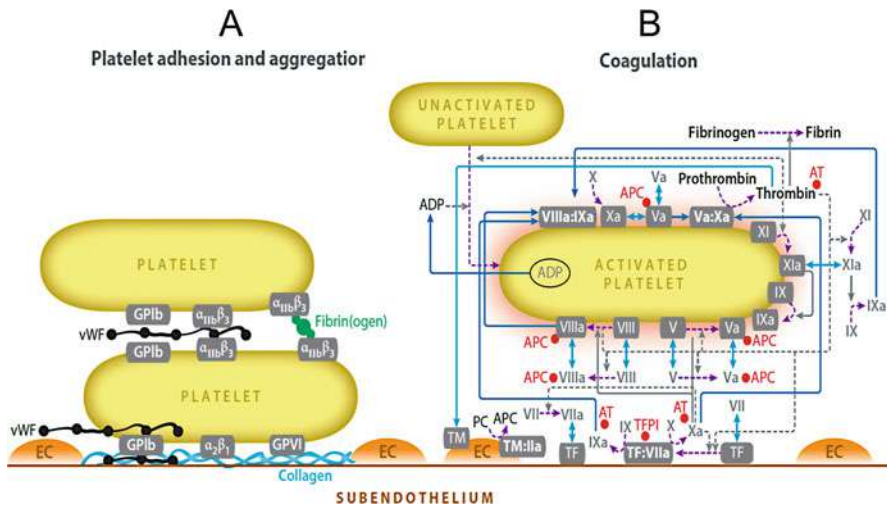
Fig. 1 *Bio-chemo-mechanics of platelet aggregation and coagulation.* (**a**) Platelet adhesion receptors and their ligands. Each platelet surface bears $\approx$25,000 GPIb receptors that bind to surface-bound von Willebrand factor; $\approx$50,000 integrin $\alpha_{IIb}\beta_3$ receptors that bind to fibrinogen and vWF; $\approx$4,000 GPVI receptors; and 1,000–4,000 integrin $\alpha_2\beta_1$ receptors that bind to several types of collagen. The integrins must be activated to form strong long-lived bonds. Collagen is a major constituent of the subendothelial (SE) matrix; fibrinogen is an abundant plasma protein; and vWF is adsorbed to SE collagen, circulates in plasma, and is secreted by endothelial cells. (**b**) Schematic illustration of coagulation reactions. Most coagulation proteins exist in an inactive (FVIII) and active form (FVIIIa). FIIa (also known as thrombin), FVIIa, FIXa, FXa, and FXIa are enzymes; their inactive precursors are called zymogens. FVa and FVIIIa are cofactors for the enzymes FXa and FIXa, respectively, and must be activated from their precursors FV and FVIII. Tissue factor (TF) is a cofactor for FVIIa. The coagulation enzyme-cofactor complexes form on SE and platelet surfaces and have enzymatic efficiencies 10^5-10^6-fold those of the enzymes alone. The activation of a coagulation protein is by proteolysis (i.e., cutting) of the precursor by another enzyme. Thrombomodulin (TM) on ECs is a cofactor for thrombin in producing the inhibitor activated protein C (APC). Other major inhibitors are antithrombin (AT) and tissue factor pathway inhibitor (TFPI). Surface-bound enzyme complexes TF:VIIa, VIIIa:IXa, Va:Xa, and TM:IIa and other surface-bound species are shown in boxes. Also shown are cell or chemical activation (purple lines), movement in fluid or along a surface (dark blue lines), enzyme action in a forward direction (solid gray lines), the feedback action of enzymes (dashed gray lines), binding to or unbinding from surface (light blue double-headed arrows), and chemical inhibitors (red circles). (Reprinted with permission from Fogelson and Neeves 2015)

different modeling approaches in order to achieve highly resolved microstructural mechanics and the mechano-transduction process of hemodynamic stresses that may induce platelet activation. Such approaches are needed for more accurately describing the platelets intracellular structure and the mechano-transduction processes of platelet activation by hemodynamic forces.

Incorporating all these processes that occur at different spatiotemporal scales, however, remains a rather challenging task. For example, platelet-wall and platelet-platelet interactions through receptor-ligand bindings occur at a subcellular

nanoscale, whereas the blood flow dynamics in the vessel around the developing thrombus is described as a macroscopic process from several micrometers to millimeters. Here, we describe a hierarchical multiscale numerical approach that may seamlessly unify and integrate different subprocesses within the clotting process.

1.1   Significance and Rationale for a Multiscale Model

Cardiovascular diseases account for over 31% of deaths globally and in the USA annually. Whether due to acute thrombosis associated with myocardial infarction or progressive intermittent atherothrombotic events, significant ventricular dysfunction may result, leading to heart failure. Presently over 6.5 million patients suffer from heart failure in USA, and their number is expected to grow by nearly 50% by 2030. Of those, a significant proportion will become candidates for mechanical circulatory support and prosthetic cardiovascular devices, also burdened with thromboembolic risk and complications, mandating antithrombotic medications for their recipients. Despite antithrombotic therapy, ventricular assist device patients have approximately a 12% stroke rate, of which 18% prove fatal. The alarming thrombosis rates, coupled with complex antithrombotic and anticoagulant management, necessitate a solution to save numerous lives and drastically reduce the ensuing healthcare costs.

Thrombosis in vascular disease is potentiated by the interaction of blood constituents with an injured vascular wall and the nonphysiologic flow patterns characterizing cardiovascular pathologies. Both initiate and enhance the hemostatic response through chronic platelet activation. Similarly, the thrombus and thromboemboli generated in devices are composed primarily of platelets. The elevated flow stresses they induce chronically activate platelets despite aggressive antithrombotic therapy. However, limitations in characterizing the mechanical stimuli, and the complex biochemical and morphological changes platelets undergo in response, still hamper our ability to successfully model such intricate phenomena. The coupling of the disparate spatiotemporal scales between molecular-level events and the macroscopic transport represents a major modeling and computational challenge, which requires a multidisciplinary integrated multiscale numerical approach. Continuum approaches are limited in their ability to cover the smaller molecular mechanisms such as filopodia formation during platelet activation. Utilizing molecular dynamics (MD) to cover the multiscales involved is computationally prohibitive. Innovative multiscale approaches are essential for elucidating this vexing problem (Bluestein et al. 2014; Zhang et al. 2014).

1.2   Flow-Induced Mechanisms of Platelet Activation, Aggregation, and Thrombosis

The "Virchow's triad" of altered blood, surface, and flow establishes the latter as the most complex component of the triad – the local flow patterns role in

platelet activation and aggregation, interaction with the vascular wall, and deposition onto foreign surfaces. It determines where a thrombus will form, its size, and composition and whether it will remain at its initiation nidus or embolize (Sorensen et al. 1999). Shear-induced platelet activation (SIPA) shows consistent "dose" and time response characteristics of biochemical agonists. It causes both aggregation and thrombin generation – the most potent platelet activator, with an established activation threshold (Kroll et al. 1996). Aggregation of platelets and deposition onto thrombogenic surfaces (Jackson 2007) have three distinct shear ranges:

(i) Low-intermediate shear in veins and larger arteries ($<10^3 \, \mathrm{s}^{-1}$). Aggregation is predominantly mediated by $\alpha_{IIb}\beta_3$ on free-flowing platelets which engage fibrinogen (Fg) adsorbed on platelet monolayers that compose the initial thrombi.
(ii) High shear in arterial microcirculation or moderate arterial stenosis ($10^3 - 10^4 \, \mathrm{s}^{-1}$). Aggregation is more vWF-dependent and involves the activity of both GPIba and $\alpha_{IIb}\beta_3$.
(iii) Pathological or "hyper" shear in atherothrombosis and cardiovascular devices ($>10^4 \, \mathrm{s}^{-1}$). Aggregation is mediated exclusively by vWF-GPIb bonds, not requiring platelet activation or adhesion via $\alpha_{IIb}\beta_3$.

1.3 Aggregation, Adhesion, and Wall Interaction

Shear-induced flow mechanisms are enhanced by eddies, flow vortices, and flow separation (Nesbitt et al. 2009), with prothrombotic agonists trapped in regions of recirculating flow. Shear-induced aggregation has two characteristic phases: (i) reversible phase between two discoid platelets and the presence of unstable platelet aggregates and (ii) stable phase that involves platelet shape change and granular release (Maxwell et al. 2007; Nesbitt et al. 2009). Unstable aggregates are formed with vWF alone, while stable aggregates require both vWF and fibrinogen, allowing for a shorter formation time and increased recruitment of platelets to aggregates (Maxwell et al. 2007). Initial aggregates typically involve clustering of platelets around a single activated platelet via membrane tethers, not pseudopodia, whereas stable aggregates require platelet sphering and filopodial projections. Increasing the shear rate from 1,800 to 5,000 s^{-1} increases the number of discoid platelet aggregates converting to stable aggregates, and $\alpha_{IIb}\beta_3$ is required for membrane tether formation between platelets in reversible aggregates (Maxwell et al. 2007). In stenotic regions, initial platelet recruitment occurs at the stenosis apex, and aggregates form at the downstream expansion zone where the flow decelerates. Larger decelerations result in larger aggregates embedded in the growing thrombus.

1.4   Continuum Approaches for Blood Flow at the Microscale

Several continuum approaches try to couple the timescales between molecular and macroscopic transport. In those the cellular level is represented by field equations and moving boundaries to allow the cell to change its shape with interfaces that have different timescales and properties (N'Dri et al. 2003; Shyy et al. 2001). Models of, e.g., adhesion of rolling leukocytes to endothelium are based on either equilibrium concepts (Bell 1978; Berk and Evans 1991; Evans 1985a, b) or kinetics approach (Dembo et al. 1988; Hammer and Lauffenburger 1987; Dong et al. 1999). Bond association and dissociation occur according to the forward and reverse reaction rates (N'Dri et al. 2003; Shyy et al. 2001). Initial models (Skalak et al. 1981a, 1989, 1981b; Weiss 1990) neglected cell rheological properties and were limited in their predictive capability (Bell 1978; Dembo et al. 1988). Flow-induced stresses acting on a cell may weaken the adhesion bonds or extract receptor molecules from the cell surface (Alon et al. 1997, 1998) with progress toward understanding the receptor-mediated cell adhesion process (Lauffenburger et al. 1987; Linderman and Lauffenburger 1986, 1988). Detailed review on such simulations using probabilistic and Monte Carlo approaches was published (Zhu 2000). Flow-mediated transport and initiation and inhibition of coagulation were modeled extensively by the Fogelson group (Fogelson and Tania 2005).

1.5   Coarse-Grained Particle Methods at the Mesoscale

CFD simulations are limited to continuum systems (Dwinzel et al. 2002; Groot and Warren 1997), representing a 10 nm to 100 μm gap between microscopic biochemistry and macroscopic scales. DPD models possess important properties of mesoscopic systems: DPD can easily model heterogeneous fluids, allowing simulations of processes which are very difficult to model by continuum approaches. Initial work focused on multiphase nature of these flows (Espanol 1998; Boek et al. 1996) to address limitations of continuum approaches that represent them consistently (Albert et al. 1998; Kechagia et al. 2002; Davis 2002). Blood flow was modeled with DPD (Dzwinel et al. 2003) with plasma represented by dissipative fluid particles, and red blood cells (RBCs) and capillary walls by coarse-grained (CG) particles ensemble (Boryczko et al. 2003, 2004), demonstrating aggregation and deformation. A triple-decker MD-DPD model interfaced atomistic-mesoscopic-continuum flow regimes (Fedosov and Karniadakis 2009). A flow-cell interaction model (Sweet et al. 2011) used subcellular elements to simulate cell motion, with the cell coupled to the plasma flow by the Langevin equation (Skeel and Izaguirre 2002). DPD was initially used in simple geometries (Fedosov and Karniadakis 2009; Soddemann et al. 2003; Espanol and Warren 1995) for modeling viscous flows with no-slip boundary conditions (Koplik and Banavar 1995a, b; Revenga et al. 1999; Pivkin and Karniadakis 2005), expanding to higher Reynolds numbers (Qian and Wang 2005), and more recently in a stenosis model (Gao et al. 2017).

2 Hierarchical Multiscale Modeling of Platelet Thrombosis

Numerical modeling of thrombus formation and growth is a challenging problem due to multiscale and multiphysics nature of clotting process, which involves fluid mechanics, cell mechanics, and biochemistry. Diverse studies have addressed this problem on different scales such as cellular, meso, and continuum levels, whereas attempts have been made to bridge these different scales to model the process at the initial phase of platelet activation and aggregation (cf. Xu et al. 2010, Zhang et al. 2014, and Grinberg et al. 2013). These studies may be broadly put in three distinct modeling strategies: cellular/subcellular modeling of adhesive dynamics, platelet transport, and aggregation in whole blood; continuum-based modeling of blood flow treating platelets as Lagrangian particles; and continuum-based modeling of thrombus formation and growth using empirical correlations for platelet deposition rates.

A major challenge for multiscale modeling of thrombogenesis is the coupling of the disparate spatiotemporal scales characterizing flow effects on deformable platelets and the resulting adhesion/aggregation bond scales. A schematic flowchart depicts flow-induced processes leading to clotting and thrombosis and the feedback between the scales (Fig. 2). A successful multiscale model requires molecular-level characterization of mechano-transduction processes activating platelets, the ensuing clotting reactions, and their interactions with the flow. Departing from traditional continuum approaches, a multiscale model is proposed by combining DPD interfaced with coarse-grained molecular dynamics (CGMD) – this integrated "DPD-CGMD" multiscale model that can bridge the gap across the vast range of the spatiotemporal scales. Starting with flow patterns with high propensity to activate and aggregate platelets, macro-to-mesoscale flow scales are modeled using DPD and micro-to-nanoscales using CGMD. An interactive DPD-CGMD dynamic coupling interface scheme will reflect any changes in platelets and thrombus morphology and feed back to affect the hemodynamics and vice versa, resulting in a multiscale model to bridge the gap between the macroscopic flow scales and the cellular/molecular scales.

2.1 Molecular Dynamics Modeling of Receptor-Ligand Interactions

Understanding how force affects receptor and ligand binding and unbinding is a long-standing effort in mechanobiology. Bond dissociation rates typically increase under mechanical stress; however, bond stability can be enhanced through specialized mechanisms induced by force, including *catch bonds* and switching to a *slip bond* with a slower off-rate (*flex bonds*) (Kim et al. 2015). At sites of vascular injury, hydrodynamic force in the bloodstream acting on vWF is pivotal in regulating the binding of the vWF A1 domain to GPIba on platelets and commencing the cross-linking of platelets by vWF to form a platelet plug. vWF circulates in the form of long, disulfide bonded concatemers, with tens to hundreds of monomers,
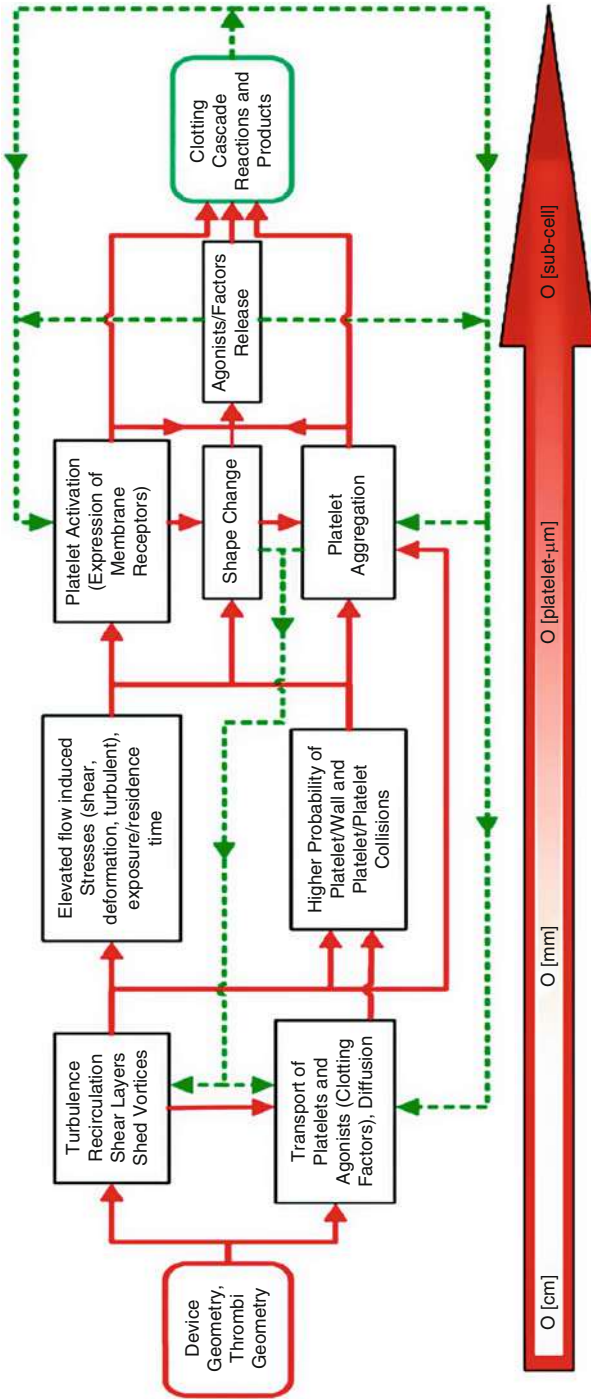
Fig. 2 Schematic flowchart of the multiscale flow-induced clotting, including mechanical and biochemical interactions

which mostly adopt a compact, irregularly coiled conformation during normal hemodynamics (see Fig. 3). At sites of hemorrhage, flow changes from shear to elongational. On the transition from low to high shear and from shear to elongational flow, irregularly coiled molecules extend to a threadlike shape, and elongational (tensile) force is exerted throughout their lengths. Molecular elongation exposes the multiple A1 binding sites in vWF concatemers for multivalent binding to GPIba (Springer 2014).

Atomic force microscope (AFM) experiments have been used extensively to probe the force required to rupture single-molecule complexes and have provided additional insights into their binding properties. By performing force spectroscopic measurements of binding/unbinding force distributions at different relaxation/pulling rates, these AFM experiments have been able to quantify the kinetics of ligand-receptor interactions.

To reveal the microscopic processes underlying the AFM observations, numerical simulations of binding/unbinding events for different complexes have been performed using steered molecular dynamics (SMD) technique. SMD allows us to explore these processes on timescales accessible to MD simulations. The basic idea behind any SMD simulation is to apply an external force to one or more atoms. In addition, you can keep another group of atoms fixed and study the behavior of the complex under various conditions. The only major difference between the AFM experiment and simulations concerns the value for the pulling velocity. Whereas the AFM experiment is carried out on a millisecond timescale, MD simulations

Fig. 3 *Conformation of vWF under elongation and structural form of its A1 domain bound to GPIba.* (**a**) Schematic organization of domains in vWF and head-to-head and tail-to-tail linkage of vWF monomers into concatemers. (**b**) The vWF A1-GPIba complex forms a super β-sheet at the interface between the A1 $\beta3$ and GPIba $\beta14$ strands. (Reprinted with permission from Springer 2014)
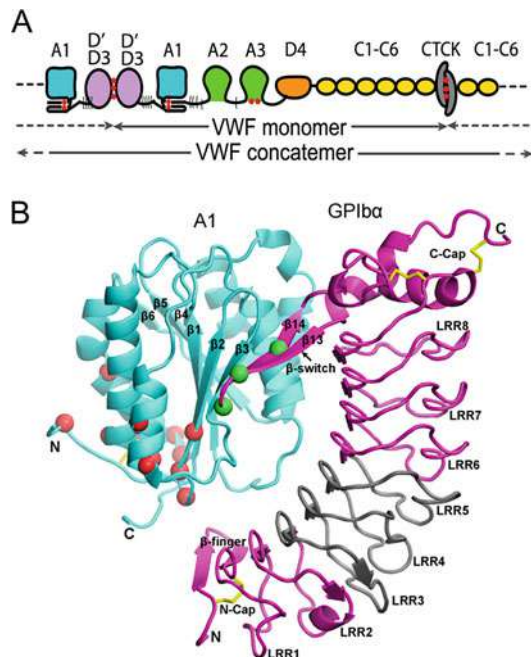
are limited to nanoseconds; therefore, one has to consider thermal fluctuations and dissipation. Because of these nonequilibrium phenomena, the rupture force should vary systematically with rupture speed, and thus the computed rupture forces should be able to be extrapolated to the experimental timescale (Grubmüller et al. 1996).

2.2   An Integrated DPD-CGMD Modeling Approach

The multiscale approach (Fig. 4) incorporates (i) top/macro blood flow scale using DPD fluid, with length scales down to $\mathcal{O}$(m), where multiple platelets each composed of an ensemble of attractive particles are suspended in a discrete particle medium with viscous fluid properties and interact with the blood vessel walls and each other, and (ii) bottom/cellular scale employing CGMD, $\mathcal{O}$ (nm) scales, in which platelets with multiple sub-components (actin filaments cytoskeleton, bilayer membrane, and cytoplasm) evolve during activation as pseudopodia grow and platelets lose their quiescent discoid shape. The fully interactive DPD-CGMD dynamic coupling interface passes the information across the scales as individual platelet activation evolves at the cellular bottom scale in response to viscous shear. This multiscale approach (Fig. 4) depicts the (a) scale exchange of shear from the DPD blood flow and (b) state exchange based on the coarse-graining of MD modeling of the individual platelets, and their shape change is transferred – interactively affecting the DPD fluid hemodynamics.
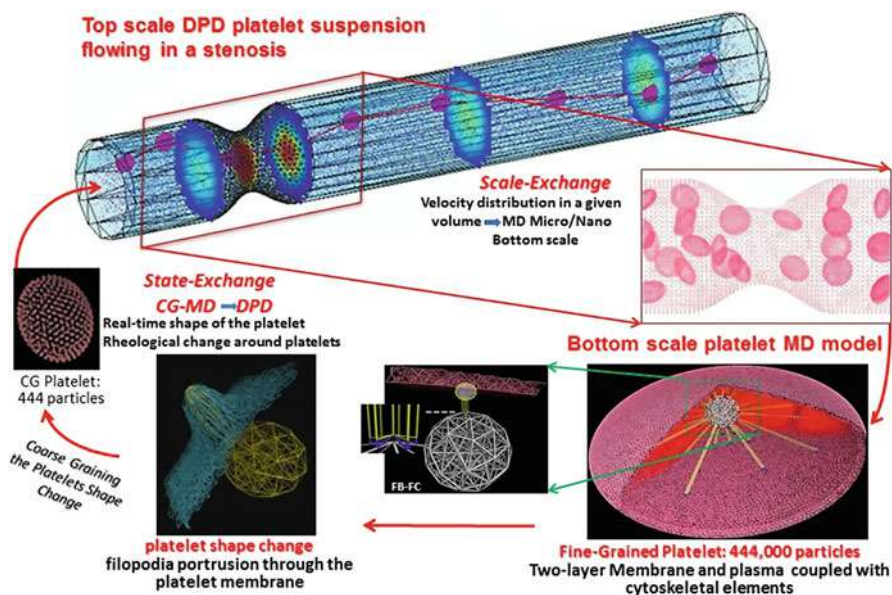
Fig. 4 Multiscale DPD-CGMD approach for modeling flow-induced platelet activation
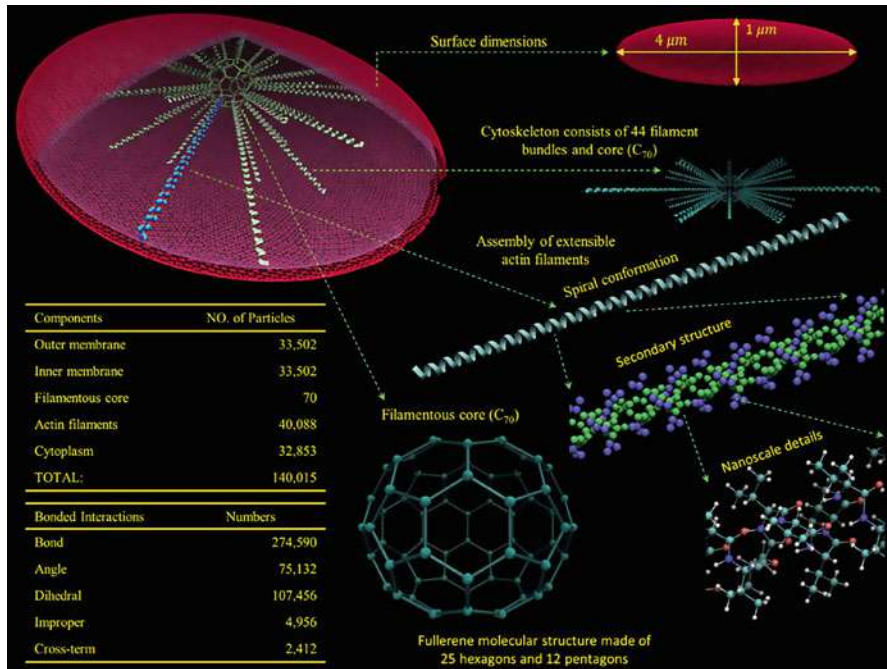
Fig. 5 *CGMD model of the intercellular constituents structure of human platelets.* Detailed connections of the filamentous core with actin filaments and the membrane particles are shown in the upper left corner. The rigid filamentous core made by carbon-70 structure at the center of platelet. One end of the actin bundles is attached to the core, and the other end is attached to the membrane. The cytoplasm fills the space between the membrane and the cytoskeleton. (Reprinted with permission from Zhang et al. 2017)

2.2.1 Coarse-Grained Molecular Dynamics Model of Platelet Cellular Structure

The cellular and subcellular structure of a resting platelet can be modeled using a CGMD approach in Fig. 5 (Zhang et al. 2017). A resting platelet has a discoid shape with a $2\,\mu m$ semimajor axis and $0.5\,\mu m$ semiminor axis. The peripheral zone is modeled as a homogeneous elastic material bilayer constructed by 2D triangulation method. A shell of $300\,AA$ thickness represents the phospholipid bilayer membrane ($100\,AA$) and an exterior coat (150–$200\,\text{Å}$). The membrane is allowed to deform under strain. The organelle zone, represented by the cytoplasm, is composed of homogeneous nonbonded particles filling the gap between the membrane and the cytoskeleton. At the fluid-platelet interface, the membrane prevents fluid particles from penetrating while maintaining the flowing platelet flipping dynamics.

The cytoplasm rheology is modeled using a Morse potential (Morse 1929). The cytoskeleton consists of two types of actin-based filaments: (a) a rigid filamentous core and (b) an assembly of radially spanning elastic actin filaments that mediates the contractility. A carbon-70 structure is used to generate the oval-shaped core.

Each actin filament is individually extensible and is tethered to the core. An α-helical structure mimics the spring-loaded molecular mechanism, with its spiral conformation continuously stretchable under external force.

A CGMD potential for the elastic membrane and filamentous core along with a Morse potential for the cytoplasm and the MD force fields for the actin filaments are adopted in this platelet model. The CGMD describes the lowest scale of the platelet intracellular constituents, including the bilayer membrane, the cytoplasm, and a cytoskeleton composed of extensible actin filaments of the pseudopods at the micro- to nanoscales. CGMD potential is written as

$$V_{CGMD}(r) = \sum_{bonds} k_b(r - r_0)^2 + \sum_{LJ} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^6\right], \tag{1}$$

and describes the deformability of the membrane (the first term) and the interaction between the membrane and intercellular particles (the second term). A MD potential without an electrostatic term for the actin-based filaments is

$$V_{MD} = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsion} k_\phi(1 + cos(n\phi - \sigma))$$

$$+ \sum_{LJ} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^6\right], \tag{2}$$

where k_b, k_θ, and k_ϕ are force constants for the α-helical structure, r_0 and θ_0 are the equilibrium distance and angle, and n is the symmetry of the rotor and δ is the phase. In the L-J term, ε_{ij} is the well depth of the L-J potential, σ_{ij} is the finite distance and r is the interparticle distance, and θ is the angle and ϕ is the torsional angle. Morse potential is used for the cytoplasm

$$V_{Morse}(r) = \varepsilon[e^{\alpha(1 - \frac{r}{R})} - 2e^{\frac{\alpha}{2}(1 - \frac{r}{R})}], \tag{3}$$

describing a pairwise nonbonded interaction for viscous flows at coarse-grained scales. R is the distance of minimum energy ε, and α is a parameter that measures the curvature of the potential around R, where r is the interparticle distance.

2.2.2 Modeling Blood Plasma Using DPD Fluid Model

In the standard DPD method, the pairwise forces consist of three components: (i) conservative force, $\mathbf{f}_{ij}^C = a_c(1 - r_{ij}/r_c)\hat{\mathbf{r}}_{ij}$; (ii) dissipative force, $\mathbf{f}_{ij}^D = \gamma\omega_d(r_{ij})(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{v}}_{ij})\hat{\mathbf{r}}_{ij}$; and (iii) random force, $\mathbf{f}_{ij}^R = \sigma\omega_r(r_{ij})(\zeta_{ij}/\sqrt{\Delta t})\hat{\mathbf{r}}_{ij}$. Hence, the total force on particle i is given by $\mathbf{f}_i = \sum_{i \neq j}(\mathbf{f}_{ij}^C + \mathbf{f}_{ij}^D + \mathbf{f}_{ij}^R)$, where the sum acts over all particles j within a cutoff radius r_c, and a_c, γ, and σ are the conservative, dissipative, and random coefficients, respectively, r_{ij} is the distance with the corresponding unit vector $\hat{\mathbf{r}}_{ij}$, $\hat{\mathbf{v}}_{ij}$ is the unit vector for the relative velocity, ζ_{ij} is a Gaussian random number with zero mean and unit variance, and Δt is the

simulation time step size. The parameters γ and σ and the weight functions are coupled through the fluctuation-dissipation theorem and are related by $\omega_d = \omega_r^2$ and $\sigma^2 = 2\gamma k_B T$, where k_B is the Boltzmann constant and T is the temperature of the system. The weight function $\omega_r(r_{ij})$ is given by

$$\omega_r\left(r_{ij}\right) = \begin{cases} (1 - \frac{r_{ij}}{r_c})^k & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c \end{cases} , \qquad (4)$$

with $k = 1$ in the original DPD form, whereas $k < 1$ is used to increase the plasma viscosity. A hybrid DPD-Morse model for blood flow simulations through severe stenotic microchannel is proposed by combining the Morse potential and classical DPD conservative force and is being used in order to reduce the fluid compressibility (Gao et al. 2017). The compression simulation presents a much smaller isothermal compressibility of the DPD-Morse fluid than that of the DPD fluid with $\kappa_T^{DM}/\kappa_T^{DPD} \approx 0.3$, which shows DPD-Morse fluid is more resistant to compression under the same condition.

2.2.3 DPD-CGMD Spatial Interfacing

Hybrid force field is developed for describing the dynamic fluid-platelet interaction (Zhang et al. 2014) and is written as

$$d\mathbf{v}_i = \frac{1}{m_i} \sum_{j \neq i}^{N} (\nabla U(r_{ij})dt + \mathbf{f}_{ij}^D dt + \mathbf{f}_{ij}^R \sqrt{dt}) , \qquad (5)$$

where $U(r_{ij})$ is the standard Lennard-Jones potential, with ε_p and σ_p the same as the energy and length characteristic parameters in CGMD, and $\mathbf{f}_{ij}^D$ and $\mathbf{f}_{ij}^R$ are the standard DPD dissipative and random forces, respectively. The L-J force term helps the cytoskeleton-confined shapes and the incompressibility of platelets against the applied shear stress of circumfluent plasma flow. The dissipative and random force terms maintain the flow local thermodynamic and mechanical properties and exchange momentum to express interactions between the platelet and the flow particles. A no-slip boundary condition was applied at the fluid-membrane surface interface. A dissipative or drag force was added to enforce the no-slip boundary condition at the fluid and membrane interface, so that the fluid particles are dragged by the dissipative forces of the membrane particles as they are getting closer to the membrane surface, mimicking boundary layer mechanism where one layer drags its adjacent layers. The hard-core L-J force simultaneously provides a bounce-back reflection of fluid particles on the membrane (to prevent fluid particles from penetrating through the platelet membrane) with the no-slip achieved by slowing down the fluid particles as the fluid particles are getting closer to the membrane surface.

Determination of the parameters for the spatial interface is conducted through numerical experiments for maintaining the Jeffery's orbits of spheroids in a shear

flow. The Jeffery's orbits solution (Jeffery 1922) describes the rotation of isolated ellipsoidal particles immersed in viscous shear flow and is widely used as a benchmark analytical solution. Zhang et al. (2014) presented the comparison of this multiscale fluid-platelet simulation results with Jeffery's orbit solution in Couette flows.

The shapes of the rigid and the deformable platelet models following the immersion are widely different: the rigid model maintains an ideal ellipsoidal shape, but the deformable model responds to the interactions between platelet and the hemodynamic flow stresses around it (Fig. 6(left)). The accumulation of shear stress-exposure time produced over time (Fig. 6(right)) indicated that the flow-induced stresses in the rigid platelet model are on average about 2.6 times larger than those of a deformable model. The hydrodynamic stress will be less on deformable objects, as the surface will move and deform absorbing part of the external force. Thus, the rigidity of the platelet model is not a negligible factor when estimating flow-induced stresses on the platelets. Neglecting the platelet deformability may overestimate the stress on the platelet membrane, which in turn may lead to erroneous predictions of the platelet activation under viscous shear flow conditions.

2.2.4   Modeling the Highly Resolved Mechano-Transduction Process of Hemodynamic Stresses in Platelets

The DPD-CGMD multiscale model described above enables an accurate study for the mechano-transduction process of the hemodynamic stresses acting onto the platelet membrane and transmitted to such intracellular constituents as cytoplasm and cytoskeleton. The platelets constituents continuously deform in response to the flow-induced stresses. Figure 7 shows the mapping of hemodynamic stresses on the membrane, cytoplasm, and cytoskeletal structure of the flowing and flipping platelet under a shear flow of 45 dyne/cm^2. Comparatively, the cytoplasm experiences a slightly lower stress level of 51.46 dyne/cm^2, implying that the deformable membrane absorbs part of the extracellular hemodynamic forces. The spread of the stress distribution on cytoplasm has a much larger standard deviation of 34.72 than the membrane's. This is expected, as the cytoplasm is more fluidic than the membrane, thereby affecting the spread of the stress distributions. The distribution of mechanical stresses on the actin filaments is 48.15 ± 11.33 dyne/cm^2. The actin-based filament structure experiences the smallest average stress level than the membrane and cytoplasm. The stress distribution on the actin filaments indicates that mostly the membrane-adjacent apical tips of the filaments that are attached to the membrane base experience higher stresses than the innermost parts. It is evident that simplifying the mechano-transduction process by using a completely rigid platelet model leads to erroneous estimation of the hemodynamic stresses acting on the platelet membrane and intracellular constituents.

2.2.5   Modeling the Activation of Platelets and Formation of Filopodia

Formation of filopodia observed during early-stage platelet activation is modeled in Pothapragada et al. (2015). Simulating the dynamics of varied filopodia formations
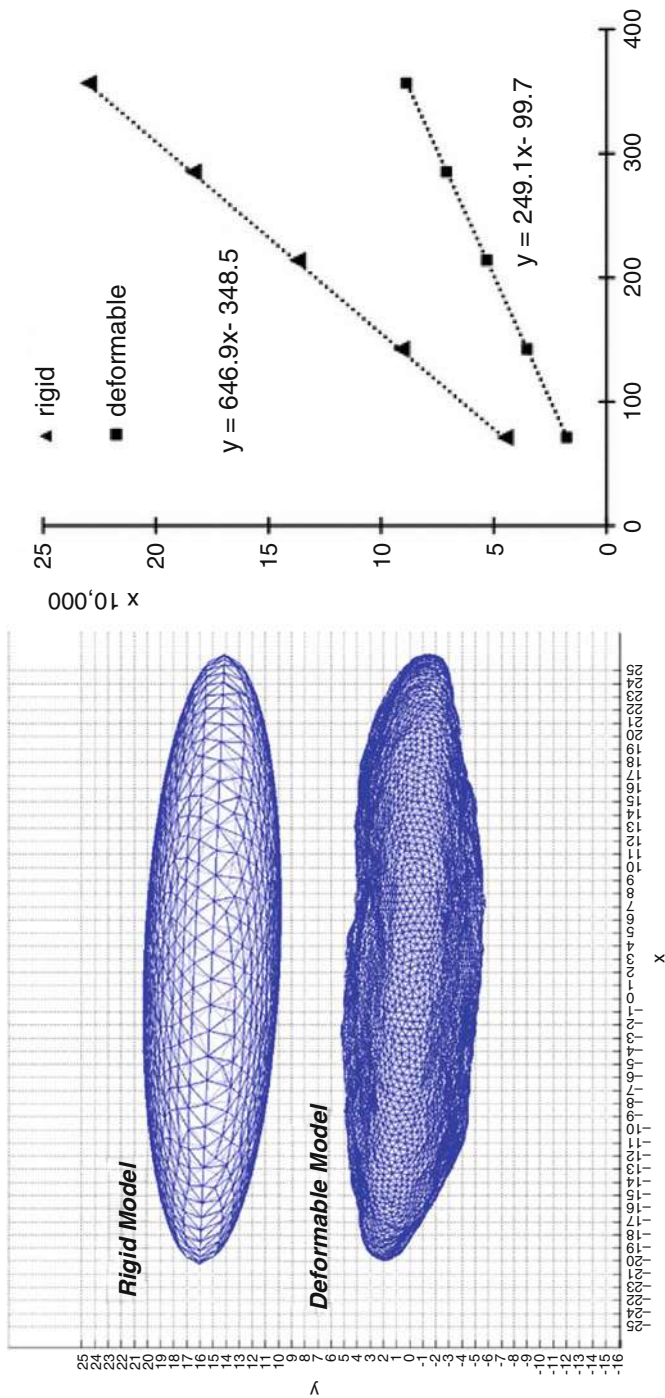
Fig. 6 *DPD-CGMD spatial interfacing for a platelet cell.* (left) Shapes of the rigid and the deformable platelet models following immersion in the plasma fluid. (right) Accumulation of shear stress-exposure time product (y-axis) on the rigid and deformable platelet models over time (x-axis in μs). (Reprinted with permission from Zhang et al. 2014)
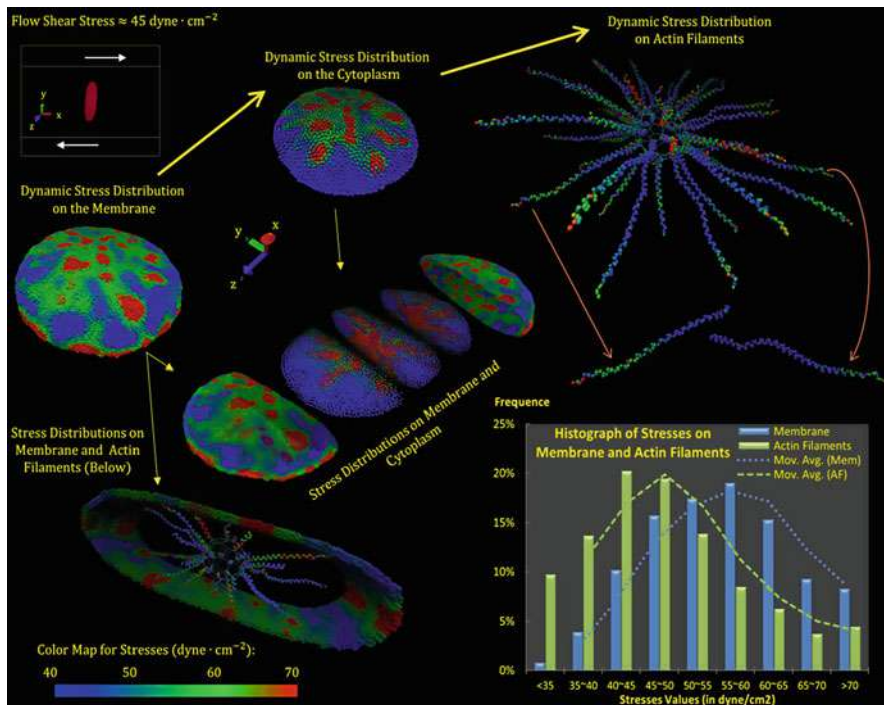
Fig. 7 *The shear flow-induced mechano-transduction process.* Hemodynamic stress distributions on the membrane, cytoplasm, and cytoskeleton (actin filaments). (Reprinted with permission from Zhang et al. 2017)

is enabled by exploring the parameter space of this CGMD platelet model. The dynamic simulation of the filopodia formation is achieved by incrementing the length and thickness parameters of filopod. The rest of the parameters remain constant during the simulation to preserve the structural integrity of the platelet during the formation of filopodia. The change of the filopod length and thickness is determined by correlating with in vitro experiments. In experiments, the platelets are exposed to constant shear stresses over a defined duration. Figure 8 shows (a) the scanning electron microscopy (SEM) images of filopodia formations under varied combinations of shear stress and exposure time and (b) corresponding simulated formation of filopods on the deformable platelet model.

Flow-mediated platelet shape change can be induced by exposing the platelets to a range of physiologic and pathologic shear stresses encountered during blood flow. Platelets isolated by gel filtration are diluted in Hepes-buffered modified Tyrodes solution with 0.1% bovine serum albumin, recalcified with a physiological concentration of $CaCl_2$, and exposed to shear stresses ranging from the physiologic 1 dyne/cm^2 to the pathologic 70 dyne/cm^2 for up to 4 mins in the hemodynamic shearing device (HSD), a programmable cone-plate-Couette viscometer that exposes cell suspensions to uniform dynamic shear stresses (Sheriff
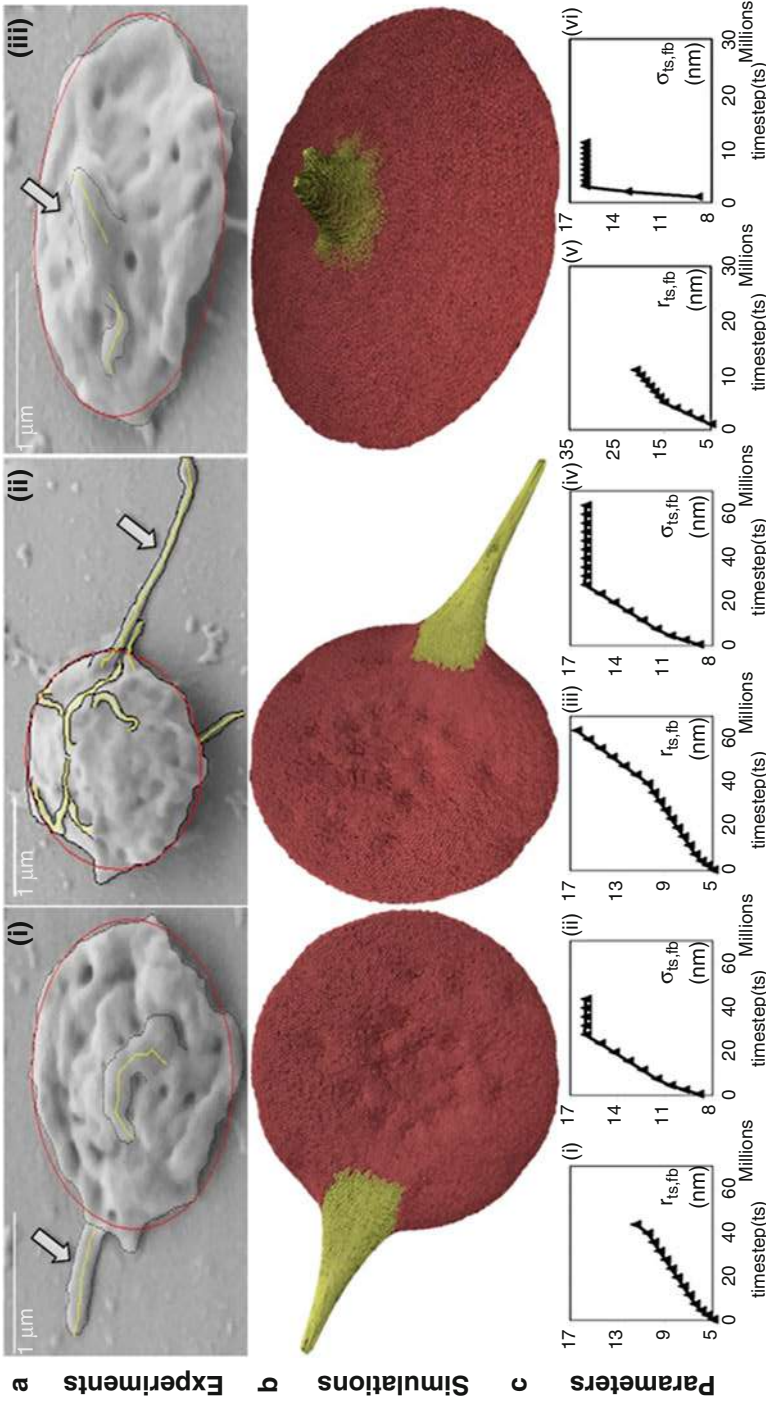
Fig. 8 *Visual comparison of experimental and simulated filopodia formation.* (**a**) SEM image after exposure to (i) 1 dyne/cm² for 4 min, (ii) 70 dyne/cm² for 4 min, and (iii) 70 dyne/cm² for 1 min. (**b**) Simulated filopodia formation. (Reprinted with permission from Pothapragada et al. 2015)

et al. 2013). During exposure, samples are withdrawn using a LabView-controlled syringe-capillary viscometer and immediately fixed in 1% paraformaldehyde to block additional shape change (Pothapragada et al. 2015). Fixed platelets are then placed on glass slides precoated with poly-L-lysine, dehydrated through a graded ethanol series, and sputter-coated with gold particles prior to imaging. Images are obtained at 30,000x magnification in a SEM and analyzed using a MATLAB image processing toolbox (Pothapragada et al. 2015). Platelet boundary parameters, such as the major and minor axis lengths, and circularity are measured by fitting an ellipse to the platelet perimeter, while filopod thickness is obtained by manually defining the intersection with the elliptical platelet body. Filopod lengths are then obtained by morphological erosion of the central filopod axis. These parameters are then recorded in a database and grouped by shear condition (i.e., shear stress and exposure time). The analyzed images are compared with snapshots of the simulation results in Fig. 8. The lengths of filopodia generated due to shear stresses in this study range from 0.24 to 2.74 μm and thicknesses range from 0.06 to 0.73 μm, and the simulations are able to achieve results within this range (Pothapragada et al. 2015).

2.2.6 Measuring Platelet Stiffness and Deformation with Dielectrophoresis-Mediated Electro-Deformation

A critical component of any biological model is validation through in vitro or in vivo experiments. Platelets behave as viscoelastic materials, storing energy as they are subject to shear deformation until their capacity to store and withstand this deformation is exceeded, resulting in shape change, fragmentation, and subsequent activation and thrombus formation. A variety of methods have been used to quantify single cell stiffness, including atomic force microscopy (AFM), cytoindenter, flow cytometry, magnetic tweezers, magnetic twisting cytometry, microfluidics, micropipette aspiration, microplate manipulation, molecular force spectroscopy, optical stretchers, and optical tweezers. Many of these methods require anchored, adhered, or shear-insensitive cells and are unsuitable for platelets, which are free-floating and sensitive to both shear forces and foreign surfaces. These limitations have been bypassed through the use of dielectrophoresis (DEP), in which neutral particles are exposed to a nonuniform electric field, leading to their translational motion. This method has been adapted as a means of cellular electro-deformation (EDF), in which mechanical properties are determined by trapping and deforming platelets exposed to an oscillating electric field without the need for surface contact, substratum attachment, or induction of membrane indentation or damage (Leung et al. 2015). This approach utilizes a microelectrode chip consisting of a patterned thin-film titanium-gold array on a glass substrate, covered by hollow silicone-coated polymer space for the fluid chamber and topped by a coverslip for microscopic observation. Isolated platelets placed in the fluid chamber are exposed to an electric current in a sine-wave pattern formed using a function generator coupled with a current amplifier, allowing modulation of the applied voltage. Platelet deformation is imaged via fluorescence microscopy, recorded with a CCD camera, and analyzed using NIH ImageJ software (Leung et al. 2015). Time-averaged electromagnetic force (F) on the platelet surface is calculated by integrating Maxwell's stress tensor

(**T**) over the cell body Ω, multiplied by the out-of-plane thickness d, with stress calculated over the platelet surface area (Chen et al. 2011):

$$F = d \int_{\partial\Omega} n \cdot T\, dS \,, \tag{6a}$$

$$\mathbf{T} = \begin{bmatrix} \varepsilon_0 \varepsilon_r E_x^2 - \frac{1}{2}\varepsilon_0 \varepsilon_r (E_x^2 + E_y^2) & \varepsilon_0 \varepsilon_r E_x E_y \\ \varepsilon_0 \varepsilon_r E_x E_y & \varepsilon_0 \varepsilon_r E_y^2 - \frac{1}{2}\varepsilon_0 \varepsilon_r (E_x^2 + E_y^2) \end{bmatrix} \,, \tag{6b}$$

where the electric field intensities in the x- and y-directions are given as E_x and E_y, respectively, and ε_r and ε_0 are the respective medium and vacuum relative permittivity. An increase in voltage yields an increase in deformation that is stiffness-dependent, as verified by comparing with platelets fixed with paraformaldehyde. The magnitude of DEP forces increases by the square of the electric field strength (15–90 V), yielding a Young's modulus of 3.5 ± 1.4 kPa for resting platelets, comparable to those measured using micropipette aspiration and AFM.

2.3 Seamless Multiscale Modeling of Thrombosis Using Dissipative Particle Dynamics

Dissipative particle dynamics, a mesoscopic particle-based hydrodynamics approach, has been effectively used to model plasma and suspending cells in blood in many applications. The advantage is the seamless integration of hydrodynamics and cell mechanics in a single framework. Further, a new submodel of DPD that is able to address chemical transport has been formulated (Li et al. 2015). The multiscale model of thrombosis includes four main processes: the hemodynamics, coagulation kinetics and transport of species, blood cell mechanics, and platelet adhesive dynamics. DPD can provide the correct hydrodynamic behavior of fluids at the mesoscale, and it has been successfully applied to study blood, where a coarse-grained representation of the cell membrane is used for red blood cells and platelets.

Transport of chemical species in the coagulation cascade is typically described by a set of continuum advection-diffusion-reaction (ADR) partial differential equations (cf. Eq. 12). In the context of Lagrangian particle-based methods such as DPD, these equations can be written as

$$\frac{d[C_i]}{dt} = \nabla(D_i \nabla[C_i]) + Q_i^S \,, \tag{7}$$

where $[C_i]$ denotes the concentration of ith reactant and D_i is the corresponding diffusion coefficient. This system can be modeled by transport DPD, which is developed as an extension of the classic DPD framework with extra variables for describ-

ing concentration fields. Therefore, equations of motion and concentration for each particle of mass m_i are written by ($d\mathbf{r}_i = \mathbf{v}_i\,dt$; $d\mathbf{v}_i = \mathbf{f}_i/m_i\,dt$; $dC_i = Q_i/m_i\,dt$) and integrated using a velocity-Verlet algorithm, where C_i represents the concentration of a specie per unit mass carried by a particle and $Q_i = \sum_{i\neq j}(Q_{ij}^D + Q_{ij}^R) + Q_i^S$ is the corresponding concentration flux. We note that C_i can be a vector $\mathbf{C}_i$ containing N components, i.e., $\{C_1, C_2, \ldots, C_N\}_i$ when N chemical species are considered. In the transport DPD model, the total concentration flux accounts for the Fickian flux Q_{ij}^D and random flux Q_{ij}^R, which are given by

$$Q_{ij}^D = -\kappa_{ij}\omega_{dc}(r_{ij})\left(C_i - C_j\right)\,, \tag{8a}$$

$$Q_{ij}^R = \varepsilon_{ij}\omega_{rc}(r_{ij})\zeta_{ij}\Delta t^{-1/2}\,, \tag{8b}$$

where κ_{ij} and ε_{ij} determine the strength of the Fickian and random fluxes. Q_i^S represents the source term for the concentration due to chemical reactions. Further, the fluctuation-dissipation theorem is employed to relate the random terms to the dissipative terms, i.e., $\omega_{dc} = \omega_{rc}^2$, and the contribution of the random flux Q_{ij}^R to the total flux is negligible.

To describe the cell mechanics, a coarse-grained representation of the cell membrane structure (lipid bilayer + cytoskeleton) can be adopted for RBCs and platelets. The membrane is defined as a set of N_v DPD particles with Cartesian coordinates $\mathbf{x}_i$, $i \in 1, \ldots, N_v$ in a two-dimensional triangulated network created by connecting the particles with wormlike chain (WLC) bonds. The free energy of the system is given by

$$V_t = V_s + V_b + V_a + V_v\,, \tag{9}$$

where V_s is the stored elastic energy associated with the bonds, V_b is the bending energy, and V_a and V_v are the energies due to cell surface area and volume constraints, respectively. RBCs are extremely deformable under shear, whereas platelets in their resting (i.e., unactivated) shape are nearly rigid. We parametrize the RBC membrane such that we can reproduce the optical tweezer stretching tests (see Fig. 9). In the case of platelets, we choose shear modulus and bending rigidity sufficiently large to ensure its rigid behavior. Platelets become more deformable upon activation through reordering the actin network in their membrane, which is accompanied by the release of their content to the plasma.

The platelet adhesive dynamics model describes the adhesive dynamics of receptors on the platelet membrane binding to their ligands. The cell adhesive dynamics model was first formulated in the work of Hammer and Apte (1992) and extended to platelets by others. This model utilizes a Monte Carlo method to determine each bond formation/dissociation event based on specific receptor-ligand binding kinetics. The probability of bond formation P_f and probability of bond rupture P_r are estimated using the following equations:
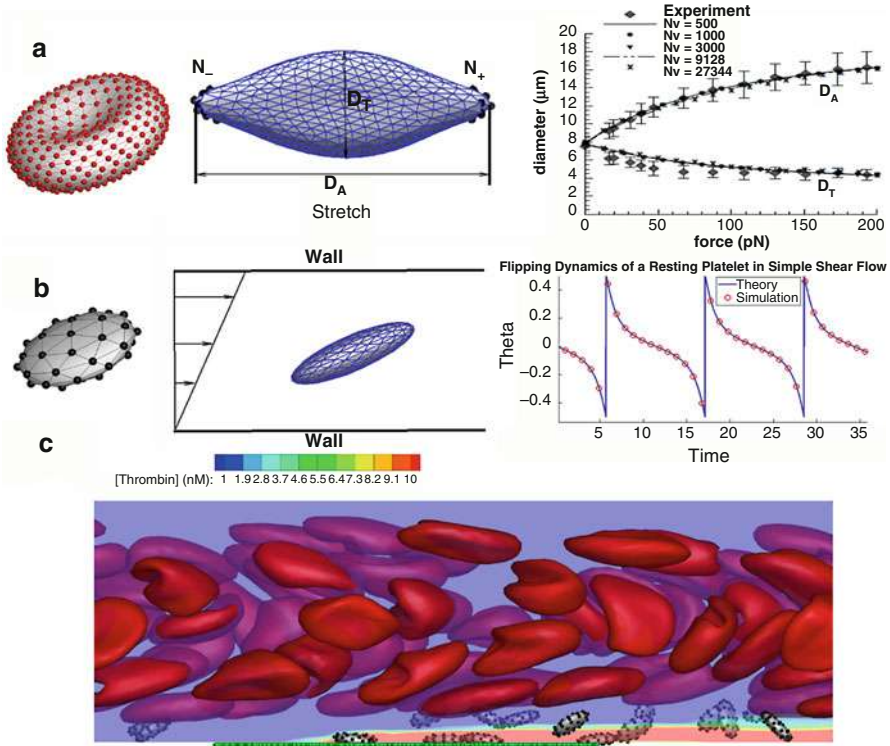
Fig. 9 *Particle representation of blood cell structure and their mechanics and simulation of clotting in whole blood.* (**a**) Coarse-grained DPD model of RBC membrane and the stretching response of stress-free RBC model for different coarse-graining levels compared with the optical tweezer stretching experiment. (**b**) Coarse-grained model of platelet membrane and its flipping dynamics when placed in a linear shear flow (theory of rigid particle dynamics in shear flow can be found in Jeffery 1922). (**c**) Seamless DPD simulation of whole blood flowing over a site of injury in a channel showing RBCs, platelets, and vWF ligands (green particles on the lower wall). Contours of thrombin concentration are plotted as the solution of the coupled ADR equations for the coagulation cascade

$$P_f = 1 - \exp(-k_f \Delta t), \tag{10a}$$

$$P_r = 1 - \exp(-k_r \Delta t), \tag{10b}$$

where k_f and k_r are the rates of formation and dissociation, respectively. Following the Bell model (cf. Bell 1978), the force-dependent k_r is evaluated by

$$k_r = k_r^0 \exp\left(\frac{\sigma_r |F_b|}{k_B T}\right), \tag{11}$$

where k_r^0 is the intrinsic dissociation rate and σ_r is the corresponding reactive compliance derived from atomic force microscope (AFM) experiments and F_b is the pulling force applied to the bond.

2.4   Continuum and Particle/Continuum Modeling of Thrombosis

At the continuum level, macroscopic numerical models have been developed by treating blood as incompressible Newtonian fluid (or non-Newtonian in small arterioles and capillaries), thus leading to continuum fields for blood velocity and pressure and the transport of enzymes, which can be resolved using an Eulerian description. To study thrombosis in both physiological and pathological conditions, some of these models treat platelets as concentration fields similar to chemical species that follow specific ADR transport equations. The governing equations for the incompressible blood flow coupled with the ADR equations related to the biochemistry of coagulation are written as

$$\rho(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u}) = -\nabla p + \nabla \cdot \boldsymbol{\tau} \,, \tag{12a}$$

$$\nabla \cdot \mathbf{u} = 0 \,, \tag{12b}$$

$$\frac{\partial c_i}{\partial t} + \mathbf{u} \cdot \nabla c_i = D_i \nabla^2 c_i + S_i \,, \tag{12c}$$

where $\mathbf{u}$ is the flow velocity, p is the pressure and $\boldsymbol{\tau}$ is the deviatoric viscous stress tensor, c_i and D_i are the concentration and diffusion coefficient for each reactant, respectively, and S_i represents the rate of production or destruction of that reactant. The shear stress tensor can be written as $\boldsymbol{\tau} = 2\mu\mathbf{D}$ where $\mathbf{D}$ is the strain rate tensor. The viscosity μ may be assumed constant for Newtonian blood or a function of the local shear rate using an empirical correlation for non-Newtonian blood. The system of equations (12) are usually solved using a finite element Galerkin formulation subject to proper Dirichlet and Neumann boundary conditions for velocity, pressure, and concentrations at the inlets/outlets and on the wall.

   It is also possible to take individual platelets into account by treating them as Lagrangian particles using a method that is able to couple particle-particle particle-wall interactions with the background flow. Force coupling method (FCM) provides such platform for *two-way coupling* of platelets (treated as rigid spherical particles) with the background flow (cf. Yazdani et al. 2017). This numerical approach has the advantage of tracking thousands of platelets forming aggregates at the site of injury and effectively capturing the shape and extent of thrombus. As a result, the thrombus shape modeled by FCM is affected by the local hydrodynamics and fluid stresses. FCM can be easily incorporated in the Navier-Stokes equation by introducing a body-force term due to the particles

$$\mathbf{f}(\mathbf{x}, t) = \sum_{n=1}^{N} \mathbf{F}^n \, \Delta\, (\mathbf{x} - \mathbf{Y}) \,, \tag{13}$$

where $\mathbf{F}^n$ is the force due to particle n. The effect of platelets is transferred to the flow field through the body force term $\mathbf{f}(\mathbf{x}, t)$ on the right-hand side of the

Navier-Stokes Equation 12a. Further, a phenomenological model based on a Morse potential V_{Morse} may be used to describe the attractive/repulsive forces between platelets. The contribution of each platelet whose center of mass is located at $\mathbf{Y}^n$ to the flow at position $\mathbf{x}$ is smoothed by a Gaussian distribution kernel $\Delta (\mathbf{x} - \mathbf{Y})$.

3 Concurrent Coupling Using a Domain Decomposition Approach

Most current work on multiscale modeling is in the setting of the so-called *concurrent coupling* methods, i.e., the microscale and the macroscale models are linked together on the fly as the computation goes on (Weinan et al. 2007). In most biological problems such as blood clotting, localized processes occur in isolated regions. For these problems, the microscopic model is only necessary near the local biologically active domains. Further away it is adequate to use the macroscopic model. In this case, the macro-micro coupling is localized, and *domain decomposition* can be used by either using overlapping interfaces or embedded subdomains. For example, in modeling of blood clotting in cerebral aneurysm shown in Fig. 10, an approach is adopted in which the flow field is solved in the full geometry using a continuum 3D solver, while small atomistic subdomains are embedded in isolated regions identified with higher platelet residence times and higher propensity of platelet deposition inside the sac of the aneurysm.

The bridging of scales is achieved through designing appropriate interface conditions between the heterogeneous solvers. The interface conditions are derived to respect the physical conservation constraints and the requirements of each solver. Specifically, mass conservation is of primary importance, while momentum and energy conservation are desirable but present several challenges, with the latter being nearly impossible in practice. Interface conditions are concurrently enforced by constructing appropriate projection, interpolation, denoising, and statistical moment-matching techniques that aim to transfer information across the different domains in a compatible way with the requirements and data format of each solver.

Coupling of atomistic and continuum solvers requires the calculation and communication of averaged properties, such as fluid velocity and density, across heterogeneous solver interfaces. For example, atomistic solvers require a local velocity flux to be imposed at each voxel of the atomistic domain. This is achieved through constructing appropriate interpolation and projection operators that are capable of mapping the continuum velocity field onto the atomistic domain. Specifically, to enforce mass conservation, the continuum solver computes the fluxes through the surface interfaces with the atomistic domains, and particles are inserted in the atomistic domain in such a way that these fluxes are preserved.

Accurate computing of averaged fields through processing of non-stationary atomistic data presents several challenges including geometrical complexity of the atomistic domains, thermal fluctuations, and flow unsteadiness. Therefore, effective averaging and filtering techniques are required before the data are transmitted to the continuum domains. In particular, to filter out the stochastic component
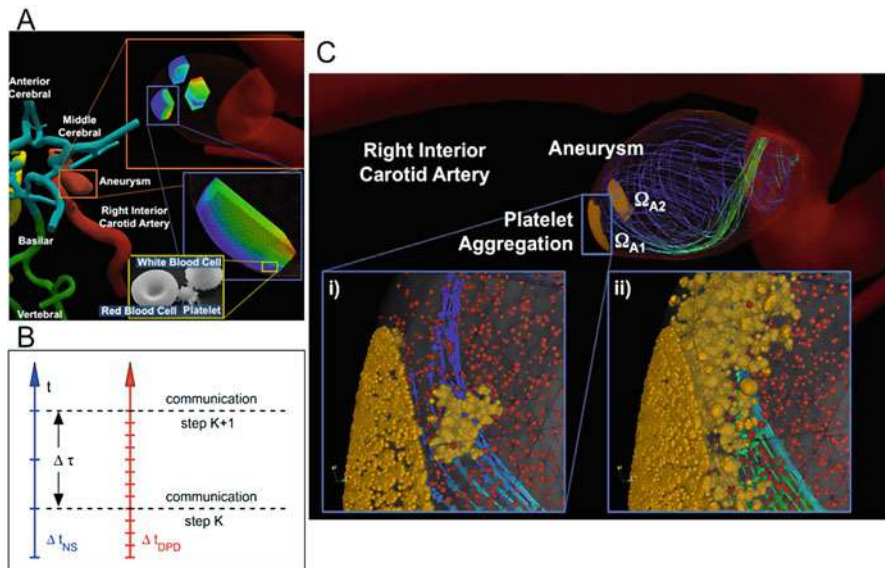
Fig. 10 *Concurrent coupling of continuum-atomistic modeling of platelet aggregation at the wall of an aneurysm.* (**a**) The patient-specific geometry was constructed from MRI imaging and contains three embedded atomistic domains to resolve the microrheology. Continuum blood flow is computed with $\mathcal{O}(0.1\,mm)$ resolution and platelet aggregation with $\mathcal{O}(1\,\mu m)$ resolution. (**b**) Asynchronous time-matching and synchronization of the continuum and atomistic (based on DPD) solvers. (**c**) Platelet aggregation at the wall of the aneurysm. Two atomistic domains $\Omega_{A1,A2}$ in which the platelet aggregation initiates are fully embedded in the continuum domain. Yellow dots correspond to active platelets and red dots to resting platelets (both modeled by hard spheres in DPD). Streamlines depict instantaneous velocity field in the (i) onset of clot formation and in (ii) clot formation as it progresses in time and space. (Reprinted with permission from Grinberg et al. 2013)

(thermal fluctuations) at all interfaces with the continuum domain, a window proper orthogonal decomposition (WPOD) has been used and shown to be very effective (Grinberg et al. 2014).

4 Multiple Time-Stepping (MTS) Algorithm for Efficient Multiscale Modeling

The disparate spatial scales between the DPD flow and the CGMD platelet model are handled by a hybrid force field. However, the disparity in temporal scales between the two models represents a huge challenge that may become prohibitive in the practice of multiscale modeling. A multiple time-stepping (MTS) algorithm is the necessary solution for this kind of multiscale models. Conventional MTS algorithms manage to differentiate multi-stepping for up to one order of magnitude of scale (Han et al. 2007) and are being applied to classical MD simulations. In order to

handle three to four orders of magnitude disparity in the temporal scales between DPD and CGMD, a novel improved MTS scheme hybridizing DPD and CGMD is developed through utilizing four different time-stepping sizes. In this MTS scheme, the DPD-based fluid system advances at the largest time step, the fluid-platelet interface at a middle time step size, and the non-bonded and bonded potentials for the CGMD-based platelet system at two smallest time step sizes (Zhang et al. 2015).

MTS for top-scale DPD uses a modified velocity-Verlet integrator that is derived from stochastic Trotter formula and is written as

$$r(\Delta t) = r(0) + \Delta t \cdot v(0) + \frac{\Delta t^2}{2m} F[r(0), v(0)], \tag{14a}$$

$$\tilde{v}(\Delta t) = v(0) + \lambda \cdot \frac{\Delta t}{m} \cdot F[r(0), v(0)], \tag{14b}$$

$$\text{compute} \quad F[r(\Delta t), \tilde{v}(\Delta t)], \tag{14c}$$

$$\tilde{v}(\Delta t) = v(0) + \lambda \cdot \frac{\Delta t}{m} \cdot \big(F[r(0), v(0)] + F[r(\Delta t), \tilde{v}(\Delta t)] \big), \tag{14d}$$

where the empirical factor λ accounts for the additive effects of stochastic interactions and $\lambda = 0.5$ restores the velocity-Verlet integrator with $\mathcal{O}(\Delta t^2)$. MTS for the bottom-scale CGMD subdivides the pairwise forces into short- and long-ranged forces with different time steps δt and $n\delta t$ where n is a positive integer. It uses the velocity-Verlet integrator as follows:

$$v\big(\frac{\Delta t}{2}\big) = v(0) + \frac{\Delta t}{2m} \cdot F[r(0)], \tag{15a}$$

$$r(\Delta t) = r(0) + \Delta t \cdot v\big(\frac{\Delta t}{2}\big), \tag{15b}$$

$$v(\Delta t) = v\big(\frac{\Delta t}{2}\big) + \frac{\Delta t}{2m} \cdot F[r(\Delta t)]. \tag{15c}$$

DPD-CGMD temporal interfacing scheme decomposes the whole integrator process into four levels as shown in Fig. 11. The topmost two levels use the scheme in Eq. 14, referred to as DPD-MTS, because both of them employ the DPD thermostatting method. The bottommost two levels use the scheme in Eq. 15, referred to as CGMD-MTS, since both of them employ the conservative potentials. In each of DPD-MTS and CGMD-MTS, the integrator is subdivided into two timescales, one for the soft potential with a larger step size and the other for the hard potential with a smaller step size.

MTS algorithm is a trade-off between speed and accuracy for multiscale modeling. Energy conservation and maintenance of adequate precisions of other measures must be verified while accelerating computations. We investigate the microscopic shape changes of platelets in response to macroscopic flow-induced stresses. Thus, in measuring numerical solutions, we focus on the accuracy of characterizing the
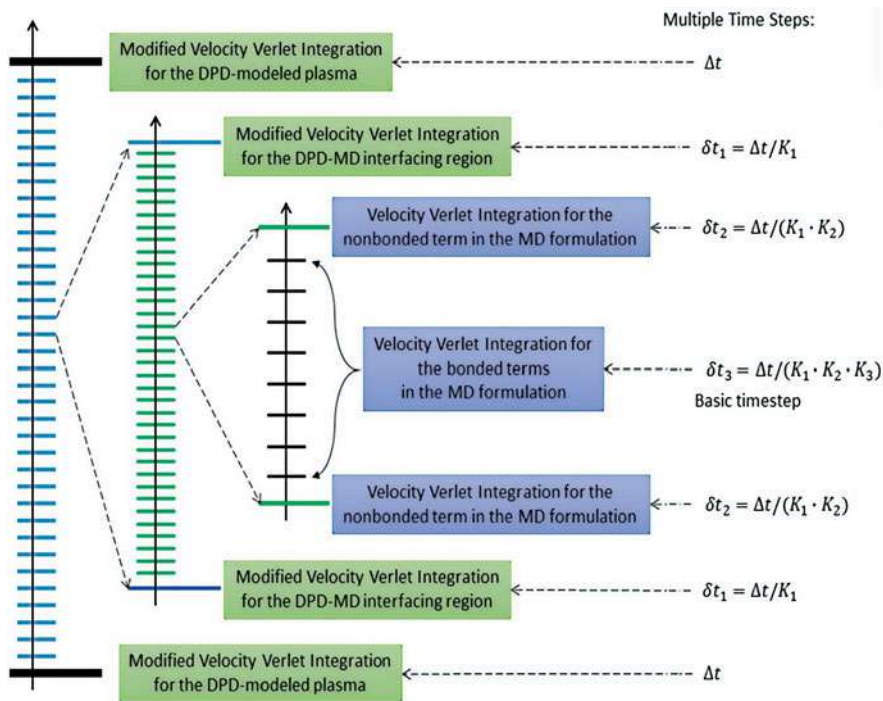
Fig. 11 *Multiple time step sizes in MTS for multiscale DPD-CGMD model.* (Reprinted with permission from Zhang et al. 2015)

hybrid system and the flowing platelets, as well as the accuracy of calculating the dynamic flow-induced stresses on the surface membrane. To compare, the standard time stepping (STS) that operates a single time step size is examined against MTS. The numerical results demonstrated that the microscopic measures for single platelets are more sensitive to the MTS parameters than the macroscopic measures for the hybrid system. The results reaffirm that a separation of temporal scales in MTS considerably improves the efficiency of utilizing parallel computing resources, as compared to conventional single-scale methods in which considerable time is wasted conducting massive unnecessary computations. For example, completing 1-ms multiscale simulation of a 10-million particle system is reduced from 2.6 years to 3.5 days only (Zhang et al. 2015, 2016).

5   Comments on Long-Term Modeling of Thrombus Formation

Any proposed multiscale numerical model for thrombus formation has to account for short-term platelet aggregation, midterm thrombus growth, and long-term clot
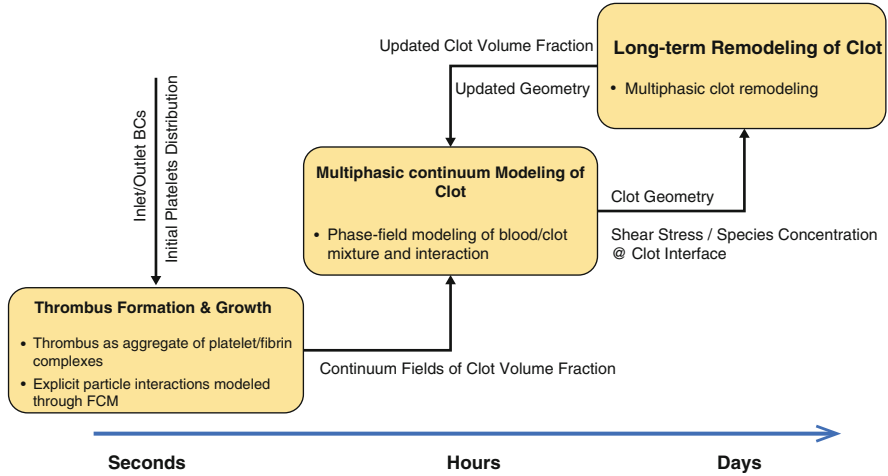
Fig. 12 *Serial/concurrent coupling of numerical models of thrombus formation and growth/remodeling in a dissecting aneurysm.* The platelet activation/aggregation model, based on FCM, is the most expensive subprocess in this framework and, thus, is only used to model the formation and propagation of the clot for several seconds after the injury. Once a stable thrombus is formed, we transform the platelets Lagrangian distribution in the aggregate clusters into continuum fields of clot volume fraction, which can be subsequently used in a phase-field numerical approach to model interaction of blood with thrombus and its further remodeling in a concurrent scheme

stabilization and remodeling (see Fig. 12). The physiologic timescales of blood coagulation dynamics and thrombosis are in the order of minutes, hence making long-time CGMD or DPD computations of such systems at the cellular level impractical. In addition, resorting to coarse-scale simulations at the macroscale may not necessarily resolve this difficulty as acceleration in the chemical reactions in the coagulation model is needed. Despite the importance of temporal scale bridging in biological applications, there has been less attention than spatial scale coupling in the literature.

In contrast to the concurrent coupling, *serial coupling* schemes determine an effective macroscale model from the microscale model in a preprocessing step and use the resulted macroscale model subsequently. Such serial coupling methods are largely limited to *parameter passing* yet are very useful in addressing long-term biological processes provided there is a timescale separation between its relevant subprocesses. For example, in Fig. 12, a possible strategy to tackle the long-term simulation of clot formation and remodeling upon arterial wall injury that normally occurs over days is proposed. Here, the subprocesses are identified as (i) the acute formation of thrombus and platelet aggregation over the first tens of seconds, (ii) the midterm formation of fibrin network at the site of thrombus and stabilization of the clot, and (iii) the long-term remodeling of the clot, which occurs over days. As noted in the figure, due to the separation of timescales between phases (i) and (ii), a serial coupling between the fine-scale particle/continuum (based on FCM)

and the coarse-scale multiphasic continuum models is possible. The challenge is the long-term simulation of particles actively coupled with flow; the complexities in geometry and flow conditions and the large size of injuries may require hundreds of thousands of FCM particles to represent platelets, which imposes a restrictively high computational cost for such simulations in large domains. To overcome this difficulty, it is possible to initially distribute particles in the regions of interest that contain thrombogenic surfaces. The timescales of phases (ii) and (iii) are not separated, which implicates a concurrent coupling scheme between these two subprocesses.

References

Albert R, Barabasi AL, Carle N, Dougherty A (1998) Driven interfaces in disordered media: determination of universality classes from experimental data. Phys Rev Lett 81(14):2926–2929

Alon R, Chen S, Puri KD, Finger EB, Springer TA (1997) The kinetics of l-selectin tethers and the mechanics of selectin-mediated rolling. J Cell Biol 138(5):1169–1180

Alon R, Chen S, Fuhlbrigge R, Puri KD, Springer TA (1998) The kinetics and shear threshold of transient and rolling interactions of l-selectin with its ligand on leukocytes. Proc Natl Acad Sci 95(20):11631–11636

Bell GI (1978) Models for specific adhesion of cells to cells. Science 200(4342):618–627

Berk D, Evans E (1991) Detachment of agglutinin-bonded red-blood-cells. 3. Mechanical analysis for large contact areas. Biophys J 59(4):861–872

Bluestein D, Soares JS, Zhang P, Gao C, Pothapragada S, Zhang N, Slepian MJ, Deng Y (2014) Multiscale modeling of flow induced thrombogenicity with dissipative particle dynamics and molecular dynamics. J Med Devices 8(2):0209541–0209542

Boek ES, Coveney PV, Lekkerkerker HNW (1996) Computer simulation of rheological phenomena in dense colloidal suspensions with dissipative particle dynamics. J Phys Condens Matter 8(47):9509–9512

Boryczko K, Dzwinel W, Yuen DA (2003) Dynamical clustering of red blood cells in capillary vessels. J Mol Model 9(1):16–33

Boryczko K, Dzwinel W, Yuen DA (2004) Modeling fibrin aggregation in blood flow with discrete-particles. Comput Methods Prog Biomed 75(3):181–194

Chen J, Abdelgawad M, Yu L, Shakiba N, Chien WY, Lu Z, Geddie WR, Jewett MA, Sun Y (2011) Electrode formation for single cell mechanical characterization. J Micromech Microeng 21(5):054012

Davis ME (2002) Ordered porous materials for emerging applications. Nature 417(6891):813–821

Dembo M, Torney DC, Saxman K, Hammer D (1988) The reaction-limited kinetics of membrane-to-surface adhesion and detachment. Proc R Soc London Ser B Biol Sci 234(1274):55–83

Dong C, Cao J, Struble EJ, Lipowsky HH (1999) Mechanics of leukocyte deformation and adhesion to endothelium in shear flow. Ann Biomed Eng 27(3):298–312

Dwinzel W, Yuen DA, Boryczko K (2002) Mesoscopic dynamics of colloids simulated with dissipative particle dynamics and fluid particle model. J Mol Model 8(1):33–43

Dzwinel W, Boryczko K, Yuen DA (2003) A discrete-particle model of blood dynamics in capillary vessels. J Colloid Interface Sci 258(1):163–173

Espanol P (1998) Fluid particle model. Phys Rev E 57(3):2930–2948

Espanol P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. Europhys Lett 30(4):191

Evans EA (1985a) Detailed mechanics of membrane-membrane adhesion and separation. I. Continuum of molecular cross-bridges. Biophy J 48(1):175–183

Evans EA (1985b) Detailed mechanics of membrane-membrane adhesion and separation. II. Discrete kinetically trapped molecular cross-bridges. Biophys J 48(1):185–192

Fedosov DA, Karniadakis GE (2009) Triple-decker: interfacing atomistic-mesoscopic-continuum flow regimes. J Comput Phys 228(4):1157–1171

Fogelson AL, Neeves KB (2015) Fluid mechanics of blood clot formation. Ann Rev Fluid Mech 47:377–403

Fogelson AL, Tania N (2005) Coagulation under flow: the influence of flow-mediated transport on the initiation and inhibition of coagulation. Pathophysiol Haemost Thromb 34(2–3):91–108

Gao C, Zhang P, Marom G, Deng YF, Bluestein D (2017) Reducing the effects of compressibility in DPD-based blood flow simulations through severe stenotic microchannels. J Comput Phys 335:812–827

Grinberg L, Fedosov DA, Karniadakis GE (2013) Parallel multiscale simulations of a brain aneurysm. J Comput Phys 244:131–147

Grinberg L, Deng M, Karniadakis GE, Yakhot A (2014) Window proper orthogonal decomposition: application to continuum and atomistic data. In: Quarteroni A, Rozza G (eds) Reduced order methods for modeling and computational reduction. Springer, Cham, pp 275–303

Groot RD, Warren PB (1997) Dissipative particle dynamics: bridging the gap between atomistic and mesoscopic simulation. J Chem Phys 107(11):4423–4435

Grubmüller H, Heymann B, Tavan P (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 271(5251):997–999

Hammer DA, Apte SM (1992) Simulation of cell rolling and adhesion on surfaces in shear flow: general results and analysis of selectin-mediated neutrophil adhesion. Biophys J 63(1):35–57

Hammer DA, Lauffenburger DA (1987) A dynamical model for receptor-mediated cell adhesion to surfaces. Biophys J 52(3):475–87

Han G, Deng Y, Glimm J, Martyna G (2007) Error and timing analysis of multiple time-step integration methods for molecular dynamics. Comput Phys Commun 176(4):271–291

Hou TY, Wu XH (1997) A multiscale finite element method for elliptic problems in composite materials and porous media. J Comput Phys 134(1):169–189

Jackson SP (2007) The growing complexity of platelet aggregation. Blood 109(12):5087–95

Jeffery GB (1922) The motion of ellipsoidal particles immersed in a viscous fluid. Proc R Soc London Ser A 102(715):161–179

Keaveny EE, Pivkin IV, Maxey M, Karniadakis GE (2005) A comparative study between dissipative particle dynamics and molecular dynamics for simple- and complex-geometry flows. J Chem Phys 123(10):104107

Kechagia PE, Tsimpanogiannis IN, Yortsos YC, Lichtner PC (2002) On the upscaling of reaction-transport processes in porous media with fast or finite kinetics. Chem Eng Sci 57(13):2565–2577

Kim J, Hudson NE, Springer TA (2015) Force-induced on-rate switching and modulation by mutations in gain-of-function von Willebrand diseases. Proc Natl Acad Sci 112(15):4648–4653

Koplik J, Banavar JR (1995a) Continuum deductions from molecular hydrodynamics. Ann Rev Fluid Mech 27:257–292

Koplik J, Banavar JR (1995b) Corner flow in the sliding plate problem. Phys Fluids 7(12):3118–3125

Kroll MH, Hellums JD, McIntire LV, Schafer AI, Moake JL (1996) Platelets and shear stress. Blood 88(5):1525–1541

Ladd AJC, Verberg R (2001) Lattice-Boltzmann simulations of particle-fluid suspensions. J Statist Phys 104(5–6):1191–1251

Lauffenburger DA, Linderman J, Berkowitz L (1987) Analysis of mammalian cell growth factor receptor dynamics. Ann N Y Acad Sci 506:147–62

Leung SL, Lu Y, Bluestein D, Slepian MJ (2015) Dielectrophoresis-mediated electrode formation as a means of determining individual platelet stiffness. Ann Biomed Eng 44(4):903–913

Li Z, Yazdani A, Tartakovsky A, Karniadakis GE (2015) Transport dissipative particle dynamics model for mesoscopic advection-diffusion-reaction problems. J Chem Phys 143(1):014101

Linderman JJ, Lauffenburger DA (1986) Analysis of intracellular receptor/ligand sorting. Calculation of mean surface and bulk diffusion times within a sphere. Biophys J 50(2):295–305

Linderman JJ, Lauffenburger DA (1988) Analysis of intracellular receptor/ligand sorting in endosomes. J Theoret Biol 132(2):203–245

Maxwell MJ, Westein E, Nesbitt WS, Giuliano S, Dopheide SM, Jackson SP (2007) Identification of a 2-stage platelet aggregation process mediating shear-dependent thrombus formation. Blood 109(2):566–576

Morse PM (1929) Diatomic molecules according to the wave mechanics. II. Vibrational levels. Phys Rev 34(1):57–64

Moulton JD, Dendy JE, Hyman JM (1998) The black box multigrid numerical homogenization algorithm. J Comput Phys 142(1):80–108

N'Dri NA, Shyy W, Tran-Son-Tay R (2003) Computational modeling of cell adhesion and movement using a continuum-kinetics approach. Biophys J 85(4):2273–2286

Nesbitt WS, Westein E, Tovar-Lopez FJ, Tolouei E, Mitchell A, Fu J, Carberry J, Fouras A, Jackson SP (2009) A shear gradient-dependent platelet aggregation mechanism drives thrombus formation. Nat Med 15(6):665–673

Pivkin IV, Karniadakis GE (2005) A new method to impose no-slip boundary conditions in dissipative particle dynamics. J Comput Phys 207(1):114–128

Pivkin IV, Karniadakis GE (2006) Coarse-graining limits in open and wall-bounded dissipative particle dynamics systems. J Chem Phys 124(18):184101

Pivkin IV, Richardson PD, Karniadakis G (2006) Blood flow velocity effects and role of activation delay time on growth and form of platelet thrombi. Proc Natl Acad Sci 103(46):17164–17169

Pothapragada S, Zhang P, Sheriff J, Livelli M, Slepian MJ, Deng Y, Bluestein D (2015) A phenomenological particle-based platelet model for simulating filopodia formation during early activation. Int J Numer Methods Biomed Eng 31(3):e02702

Qian TZ, Wang XP (2005) Driven cavity flow: from molecular dynamics to continuum hydrodynamics. Multiscale Model Simul 3(4):749–763

Revenga M, Zuniga I, Espanol P (1999) Boundary conditions in dissipative particle dynamics. Comput Phys Commun 122:309–311

Schneider S, Nuschele S, Wixforth A, Gorzelanny C, Alexander-Katz A, Netz R, Schneider M (2007) Shear-induced unfolding triggers adhesion of von willebrand factor fibers. Proc Natl Acad Sci 104(19):7899–7903

Serrano M, Espanol P (2001) Thermodynamically consistent mesoscopic fluid particle model. Phys Rev E 64(4 Pt 2):046115

Sheriff J, Soares JS, Xenos M, Jesty J, Bluestein D (2013) Evaluation of shear-induced platelet activation models under constant and dynamic shear stress loading conditions relevant to devices. Ann Biomed Eng 41(6):1279–1296

Shyy W, Francois M, Udaykumar H, N'Dri N, Tran-Son-Tay R (2001) Moving boundaries in micro-scale biofluid dynamics. Appl Mech Rev 5:405–453

Skalak R, Keller SR, Secomb TW (1981a) Mechanics of blood-flow. J Biomech Eng 103(2):102–115

Skalak R, Zarda PR, Jan KM, Chien S (1981b) Mechanics of rouleau formation. Biophys J 35(3):771–781

Skalak R, Ozkaya N, Skalak TC (1989) Biofluid mechanics. Ann Rev Fluid Mech 21:167–204

Skeel RD, Izaguirre JA (2002) An impulse integrator for Langevin dynamics. Mol Phys 100(24):3885–3891

Soddemann T, Dnweg B, Kremer K (2003) Dissipative particle dynamics: a useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations. Phys Rev E 68(4):046702

Sorensen EN, Burgreen GW, Wagner WR, Antaki JF (1999) Computational simulation of platelet deposition and activation: I. Model development and properties. Ann Biomed Eng 27(4):436–448

Springer TA (2014) Von willebrand factor, jedi knight of the bloodstream. Blood 124(9): 1412–1425

Sweet CR, Chatterjee S, Xu Z, Bisordi K, Rosen ED, Alber M (2011) Modelling platelet-blood flow interaction using the subcellular element Langevin method. J R Soc Interface 8(65): 1760–1771

Weinan E, Engquist B, Li X, Ren W, Vanden-Eijnden E (2007) Heterogeneous multiscale methods: a review. Commun Comput Phys 2(3):367–450

Weiss L (1990) Metastatic inefficiency. Adv Cancer Res 54:159–211

Xu Z, Lioi J, Mu J, Kamocka MM, Liu X, Chen DZ, Rosen ED, Alber M (2010) A multiscale model of venous thrombus formation with surface-mediated control of blood coagulation cascade. Biophys J 98(9):1723–1732

Yazdani A, Li H, Humphrey JD, Karniadakis GE (2017) A general shear-dependent model for thrombus formation. PLoS Comput Biol 13(1):e1005291

Zhang P, Gao C, Zhang N, Slepian MJ, Deng Y, Bluestein D (2014) Multiscale particle-based modeling of flowing platelets in blood plasma using dissipative particle dynamics and coarse grained molecular dynamics. Cell Mol Bioeng 7(4):552–574

Zhang P, Zhang N, Deng Y, Bluestein D (2015) A multiple time stepping algorithm for efficient multiscale modeling of platelets flowing in blood plasma. J Comput Phys 284:668–686

Zhang P, Zhang N, Gao C, Zhang L, Gao Y, Deng Y, Bluestein D (2016) Scalability test of multiscale fluid-platelet model for three top supercomputers. Comput Phys Commun 204: 132–140

Zhang P, Zhang L, Slepian MJ, Deng Y, Bluestein D (2017) A multiscale biomechanical model of platelets: correlating with in-vitro results. J Biomech 50:26–33

Zhu C (2000) Kinetics and mechanics of cell adhesion. J Biomech 33(1):23–33

Cluster-Guided Multiscale Lung Modeling via Machine Learning

112

Ching-Long Lin, Sanghun Choi, Babak Haghighi, Jiwoong Choi, and Eric A. Hoffman

Contents

––––––––––––––––––––

C.-L. Lin (✉)
Department of Mechanical Engineering, The University of Iowa, Iowa City, IA, USA

Department of Radiology, The University of Iowa, Iowa City, IA, USA

IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA, USA
e-mail: ching-long-lin@uiowa.edu

S. Choi
School of Mechanical Engineering, Kyungpook National University, Daegu, Republic of Korea

B. Haghighi · J. Choi
Department of Mechanical Engineering, The University of Iowa, Iowa City, IA, USA

IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA, USA

E. A. Hoffman
Department of Radiology, The University of Iowa, Iowa City, IA, USA

Department of Biomedical Engineering, The University of Iowa, Iowa City, IA, USA

Abstract

Accurate prediction of airflow distribution and aerosol transport in the human lungs, which are difficult to be measured in vivo but important to understand the structure and function relationship, is challenging. It is because the interplay between them spans more than two orders of magnitude in dimension from the trachea to alveoli. This chapter reviews the techniques and strategies for modeling lungs both within and between subjects, viz., subject specificity versus generalization from individuals to populations, with both exhibiting multiscale characteristics. For "within-subject" modeling, a computed tomography (CT)-derived subject-specific computational fluid dynamics (CFD) lung model is presented. The pipeline for building such an imaging-based lung model is composed of image segmentation and processing, geometrical modeling labeled with anatomical information, image registration, three-dimensional (accurate) and one-dimensional (approximate) coupling techniques, and a high-fidelity turbulent flow model. The subject-specific model is essential in predicting local structural and functional interactions. For "between subjects" modeling, machine learning is employed to identify homogeneous subpopulations (clusters), among healthy and diseased populations, aiming to bridge individual and population scales. For this purpose, three major issues need to be addressed. They are intersubject variability (due to, e.g., gender, age, and height), inter-site variability (due to scanner and imaging protocol differences), and definition of quantitative CT imaging-based metrics at multiple scales (due to alterations at different disease stages) needed for clustering. The use of the cluster membership to select representative subjects for detailed CFD analysis enables an examination of the cluster-specific structural and functional relationships.

1   Introduction

Machine learning (ML) is a data-driven technique in the field of artificial intelligence. It can generate a self-learning program to achieve exact knowledge from a cycle of learning and self-adjustment (Magoulas and Prentza 2001; Deo 2015). ML can help interpret patterns (Uppaluri et al. 1997, 1999; Xu et al. 2006a, b) or extract insightful information from heterogeneous data sets. Each data set used in ML algorithms is represented by a set of features. The features may be continuous or categorical. If an algorithm uses features assigned with a priori labels (related correct outputs), it is called supervised learning. On the other hand, it is called unsupervised learning if data are unlabeled (Deo 2015). For supervised learning, the data used can be classified into two types: (1) a training set used for learning and parameter fitting and (2) a test (validation) set used for assessing the performance of a fully trained ML algorithm (prediction). Decision tree, Naïve Bayes, and support vector machine are common ML supervised algorithms. An unsupervised learning algorithm learns features from unlabeled input data and uses these learned features to recognize the class of new data. Clustering methods, such as K-means,

hierarchical, Gaussian mixture models, and self-organizing maps, are unsupervised learning algorithms.

Healthcare enterprise includes the acquisition, processing, and storage of medical information for diagnosis and disease treatment. Applications in healthcare deal with large amounts of data of differing types, including medical imaging (such as computed tomography (CT), X-ray plates, magnetic resonance), DNA sequences, unstructured text, etc. ML can be employed to process and interpret these medical data, providing a critical tool for personalized treatment and ultimately shifting toward the paradigm of precision medicine (Mesko 2017). Precision medicine has been a core theme for future advances in medicine since the Precision Medicine Initiative was announced by the National Institutes of Health (NIH) and President Obama in 2015 and 2016 (Collins and Varmus 2015; Ledford 2016). Precision medicine not only means individualized medicine but also seeks to advance changes in medicine through the active participation of patients.

This chapter is focused on a review of the application of ML to spatially linked medical image data sets to improve the understanding and treatment of human lung diseases. This application is multiscale in two distinct aspects: (1) within and (2) between individuals. First, structural and functional alterations within individuals at different stages of the pathologic processes associated with lung disease can occur regionally (at a segmental or local scale), globally (at a lobar, organ, or global scale), or both. Thus, it is essential to examine variables at both local and global scales to capture a wide spectrum of pathologic states. In addition, some in vivo data are difficult, if not impossible, to be measured; but they are important to better understand the interplay between structural and functional variables at multiple scales over time. Thus, it is desirable to develop subject-specific lung models to capture these missing data. Second, due to heterogeneity of lung diseases, a tailored treatment for one individual may not be applicable to others. To bridge individual and population scales (viz., between individuals), identification of subpopulations (clusters) characterized by similar structural and functional alterations and clinical features may greatly facilitate the design of cluster-specific treatments and assessment of outcomes.

NIH-sponsored multicenter trials have enabled studies of lung diseases by collecting diverse data sets from large cohorts, such as the Severe Asthma Research Program (SARP) (Wenzel and Busse 2007) and the Subpopulations and Intermediate Outcome Measures in Chronic Obstructive Pulmonary Disease (COPD) Study (SPIROMICS) (Couper et al. 2014; Sieren et al. 2016). SARP was established by the National Heart, Lung, and Blood Institute (NHLBI) in 2001 to study severe asthma in multiple centers, aiming to identify phenotypes that differentiate severe asthma from those with mild-moderate asthma. Thus, SARP allows for a comprehensive phenotyping with clinical, biologic, genetic, and imaging data. Likewise, COPD is currently the third leading cause of death in the United States (Gibson and Simpson 2009). SPIROMICS aims to identify and validate surrogate markers of disease severity for intermediate outcome measures. For this purpose, high-quality quantitative data assessment is one of the key elements in these studies. Imaging data, gathered under a protocol common to both SARP and

SPIROMICS (Sieren et al. 2016), have a great potential to advance knowledge of basic lung physiology and pathophysiology of various diseases providing structural and functional characteristics at local and global scales. Due to highly dynamic characteristics of lung functions, intra-subject, intersubject, interpopulation, and inter-subpopulation variations are of important issues. This uniqueness calls for correct understanding of underlying physics and appropriate treatment of variability in parameterization. ML provides a viable approach to parameterization as well as classification.

This chapter is organized as follows. Section 2 reviews the modeling pipeline for building a subject-specific CT imaging-based multiscale computational fluid dynamics (CFD) lung model that spans generations of airways for simulation of airflow and particle deposition. The model can be used to predict missing data that are essential to understand disease progression. Section 3 discusses the ML methods employed for clustering and classification of asthma cohorts, COPD cohorts, or both. Section 4 presents some applications that utilize cluster membership to guide CFD analysis for improved treatment. Finally, future work is presented in regard to utilization of longitudinal data to further advance the above CFD-ML integrated framework.

2   CT-Based Multiscale Lung Model

The characteristic length scale of the human airways changes from centimeters in the mouth-nose and trachea to micrometers at the terminal bronchioles. As the airflow in the proximal end of the tree divides repeatedly leading eventually to tens of thousands of terminal bronchioles, the primary fluid flow parameter, Reynolds number (Re), scales between three orders of magnitude for a peak flow rate of normal breathing: 1000–2000 at the trachea and 1 at the terminal bronchioles. In the proximal sites, the geometrical complexity such as glottal constriction triggers jet-induced turbulence, causing coexistence of laminar, transitional, and turbulent flows. Thus, it is challenging to build a subject-specific lung model that can capture the structure and function relationships at multiple scales. Despite the complex nature of airflow in the lung, the most common clinical measures of lung function are obtained through the use of global lung function metrics, dubbed the pulmonary function test (PFT). PFTs include such measures as the forced expiratory volume for the initial 1 s (FEV1) when a maximal effort is made to exhale from full inspiration (total lung capacity or TLC) and forced vital capacity (FVC) which represents the maximum amount of air that an individual can force out of their lungs when exhaling as rapidly as possible from TLC down to residual volume (RV) or full expiration. Individual differences in lung size are accounted for by relating FEV1 and FVC through the ratio FEV1/FVC. The device through which the patient blows is known as the spirometer.
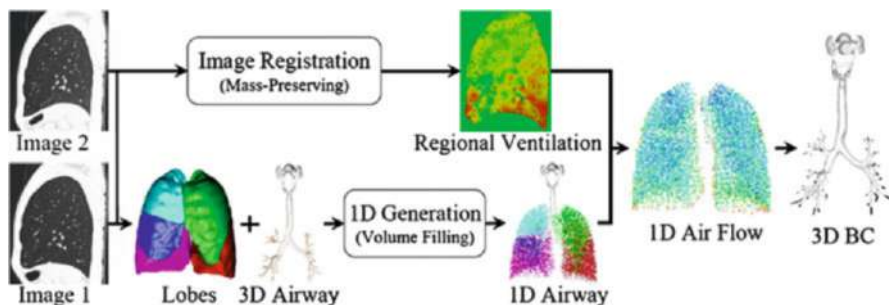
Fig. 1 A flow chart of the imaging-based CFD lung modeling procedure. *LUL* left upper lobe, *LLL* left lower lobe, *RUL* right upper lobe, *RML* right middle lobe, *RLL* right lower lobe (Yin et al. 2010)

2.1   Pipeline for a Lung Model

Recent advances in CT imaging and image processing techniques have been integrated with CFD for simulations of turbulent flows using three-dimensional and one-dimensional (3D-1D) coupled framework (Lin et al. 2009, 2013; Yin et al. 2010; Choi 2011). The pipeline for building a subject-specific imaging-based CFD lung model is illustrated in Fig. 1, comprising five steps of (1) image segmentation and processing, (2) airway geometrical modeling (for building airway structure), (3) image registration (for deriving regional ventilation function), (4) linking structure and function for boundary conditions (BCs), and (5) turbulence modeling and CFD simulation, to predict airflow and particle deposition in the lungs.

2.1.1   Image Segmentation and Processing

State-of-the-art CT provides the best resolution among various imaging modalities when one seeks to capture an image of the integrated lung anatomy. Advanced CT scanners can resolve in plane detail of 0.5×0.5 mm or better, capturing about ten generations of the airways, including segmental and sub-segmental airways. The geometries of lungs, lobes, airways, and vasculature (including arteries and veins) can be segmented from volumetric CT image via an automated software program, such as Apollo (VIDA, Coralville, IA, USA) (Hoffman and McLennan 1997; Hoffman and Chon 2005; Hoffman et al. 2016). Subsequently, structural and functional imaging-based metrics, such as luminal area, airway diameter, wall thickness, air and tissue volumes of the lungs, lobes, and among others, can be extracted. Also airway branches from the trachea to the segmental airways are labeled anatomically, and 1D airway skeleton is generated. With the introduction of dual-energy CT, it is also possible to image regional ventilation and perfusion via the use of radio-dense contrast agents such as xenon gas or iodine, respectively (Fuld et al. 2013; Iyer et al. 2016)
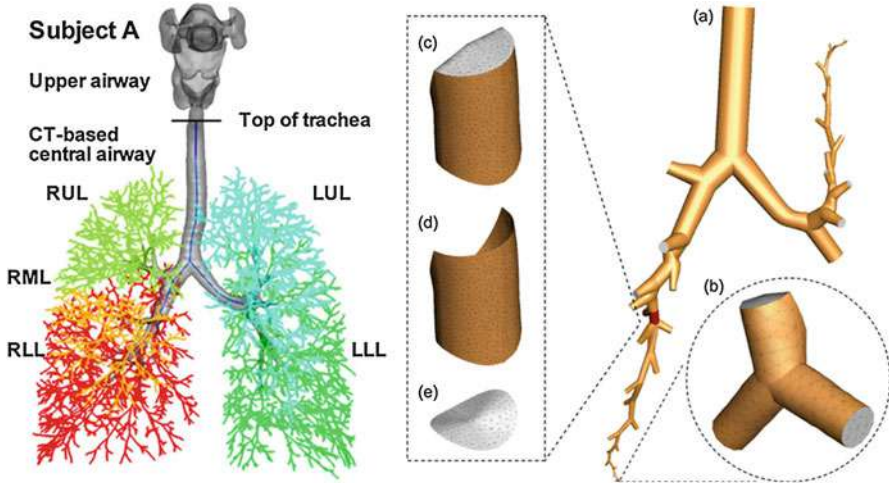
Fig. 2 (Left) A front view of a CT-based 3D-1D coupled geometrical model (Wu et al. 2015). (Right) 3D CFD mesh of airway model, showing (**a**) 3D geometry in the regions of interest, (**b**) computational grid in terminal airways, (**c**) airway segment volume, (**d**) airway segment wall, (**e**) cross-section (Lin et al. 2013)

2.1.2 Airway Geometrical Modeling

The surface geometry and 1D skeleton of segmented airways are edited for airway geometric model for CFD simulation. For CT-unresolved small airways, of which the diameter is usually less than 1 mm, a 1D skeletonized airway model is reconstructed by use of a space filling algorithm within each of five lobes as color-coded in Fig. 2 left (Tawhai et al. 2000, 2004), linking large proximal airways to the lung periphery. The CT image obtained at full inspiration, or TLC is used in this process for the best resolution. Then, a CT-fitting technique is used to complete a 3D-1D coupled airway geometrical model (Miyawaki et al. 2017a), in which a 3D model is used in the regions of interest for a detailed flow simulation (see Fig. 2 right (a)) whereas the 1D model allows bridging of the 3D airways and lung parenchyma where ventilation function takes place and gas exchange occurs with the blood in the capillary beds. In the 3D geometrical model, a volumetric mesh is generated for CFD simulation (see Fig. 2 right (b–e)). The upper airways sit proximal to the tracheal inlet. Through the pharynx and the larynx, respectively, the upper airways precondition the flow and induce a laryngeal jet, which becomes transitional or turbulent flow under normal breathing conditions (Lin et al. 2007; Choi et al. 2009). However, CT scans of the lung which include the upper airways are not commonly acquired as part of clinical protocols. For this reason as well as the need to reduce computational cost, a synthetic eddy model (SEM) is employed in conjunction with a glottal constriction extended from the trachea to produce a laryngeal turbulent jet (Miyawaki et al. 2017b).

2.1.3   Image Registration

With two or more lung CT images obtained at different static lung volumes (typically TLC and one or more expiratory lung volumes), image registration is a key technique to obtaining functional variables, e.g., regional ventilation (distribution of the inspired air bolus), from the images, thus providing BCs for CFD lung simulation. Image registration determines the optimal local-to-local correspondence between medical images collected at different times or using different imaging modalities. Methods for image matching and linking multiple lung volumes have proven useful in linking structure and function to assess their longitudinal alterations (Crum et al. 2004; Choi et al. 2013a, 2017a; Jahani et al. 2017). The similarity measure of the sum of squared tissue volume difference (SSTVD) (Yin et al. 2009) has been demonstrated to successfully improve registration of lung images with large deformation. This approach takes into account changes in the reconstructed Hounsfield unit (HU) due to inflation in lung motion. Also faster and more accurate methods for image registration are essential in population-based studies that require the processing of large numbers of CT data sets. For example, Ellingwood et al. (2016) proposed a computation- and memory-efficient Diffeomorphic Multilevel B-Spline Transform Composite (DMTC) method to implement nonrigid mass-preserving registration of two CT lung images on graphics processing units (GPUs). Later, Haghighi et al. (2017a) developed a symmetric multilevel nonrigid registration by employing an inverse consistent (IC) transformation to impose the inverse consistency (Yang et al. 2014) and improve the accuracy of image registration. Besides providing subject-specific BCs for CFD, registration-derived variables provide functional metrics that can be used in clustering analysis discussed in Sect. 3.

2.1.4   Linking Structure and Function for BCs

Local air volume differences derived from image registration are used to approximate regional flow volume fractions at the ends of terminal bronchioles (Yin et al. 2013; Jahani et al. 2014, 2015, 2017). Then, through the 3D-1D coupled airways, flow rate fractions are computed at the distal ends of the 3D airway (Yin et al. 2010; Choi 2011; Lin et al. 2013; Miyawaki et al. 2016a), allowing calculation of flow BCs at a given total flow rate imposed at the proximal entrance of the mouth and supraglottis. The approach to extract 3D BCs for the terminal branches through the lung accounts for heterogeneous lobar ventilation in healthy subjects. However, a recent study (Wongviriyawong et al. 2013) has argued that static air volume change can predict the effect of regional compliance, but it cannot describe the effects of airway resistance on the distribution of regional ventilation. Using a 1D network structure, Choi et al. (2016) combined the effects of compliance and resistance to predict regional airflows allowing for 3D CFD simulation. The approach demonstrated the effect of airway resistance on regional airflows in asthmatic patients with airway narrowing. Thus, the new 1D network approach can be applied to improve BCs for patients with airway narrowing to avoid unphysiological pressure drops. It is noted

that the 1D resistance model has been sensitively estimated considering branching angle, flow rate, and flow direction (inspiration vs. expiration) (Choi et al. 2018).

2.1.5 Turbulence Modeling and CFD Simulation

The three common methods for modeling 3D turbulent flows are direct numerical simulation (DNS), large eddy simulation (LES), and Reynolds averaged Navier-Stokes models (RANS) (Kleinstreuer and Zhang 2003; Lin et al. 2013; Tawhai et al. 2009; Tawhai and Lin 2011). DNS resolves all turbulent eddies, thus being computationally expensive. In contrast, RANS does not resolve any turbulent eddies but parameterizes them, thus being the most inaccurate but computationally efficient. LES resolves energy-containing large turbulent eddies but parameterizes small eddies. LES has been demonstrated to capture high-fidelity turbulent flow structures in the complex human branching airways (Lin et al. 2007; Choi et al. 2009, 2010; Yin et al. 2010, 2013; Miyawaki et al. 2016a). Thus, LES is adopted in the CFD lung model presented in Sect. 2.2. In addition, temperature, humidity, heat, and water vapor transfer from the oropharyngeal region to the terminal bronchioles in the human lung can be predicted by 3D thermodynamic models (Wu et al. 2014).

2.2 CFD-Predicted Variables in the Human Lungs

2.2.1 Turbulent Flow in the Airways

Contrary to many conventional assumptions, airflow in the human airways features turbulent or transitional characteristics even under normal tidal breathing conditions (Lin et al. 2007; Choi et al. 2010, 2013b). Glottal constriction induces a laryngeal jet, which undergoes turbulent transition during the inspiratory phase of tidal breathing. Turbulence in the trachea efficiently mixes air together with embedded particulate matter before they enter the branching airways. Turbulence decays over several airway generations depending on the individual's breathing pattern and airway diameters (Lin et al. 2007; Choi et al. 2009, 2010; Yin et al. 2013). Figure 3 shows the turbulent jet structure obtained by 3D-1D coupled CFD simulation of tidal breathing in a healthy human lung. Turbulent velocity and pressure fields at a peak inspiratory flow are demonstrated.

2.2.2 Inhaled Particle Transport and Deposition

Lagrangian particle-tracking simulation allows quantification of transport and deposition of inhaled aerosols in the lungs by generation and lobe (Zhang et al. 2002; Lambert et al. 2011; Longest et al. 2012; Miyawaki et al. 2012, 2016b). For example, Miyawaki et al. (2012) compared particle transport characteristics embedded in air, helium (He), He-O_2, xenon (Xe)-O_2, and Xe at the same flow rates to investigate the effect of carrier gas on particle deposition efficiency in the human lungs. Due to the density difference, Re differs by more than tenfold between He and Xe-O_2. Figure 4 shows that the laryngeal jet of the light gas He remains laminar, where particles flow in separate streams. The heavy gas mixture Xe-O_2 becomes turbulent both in the upper airway and the trachea, where particles in the trachea
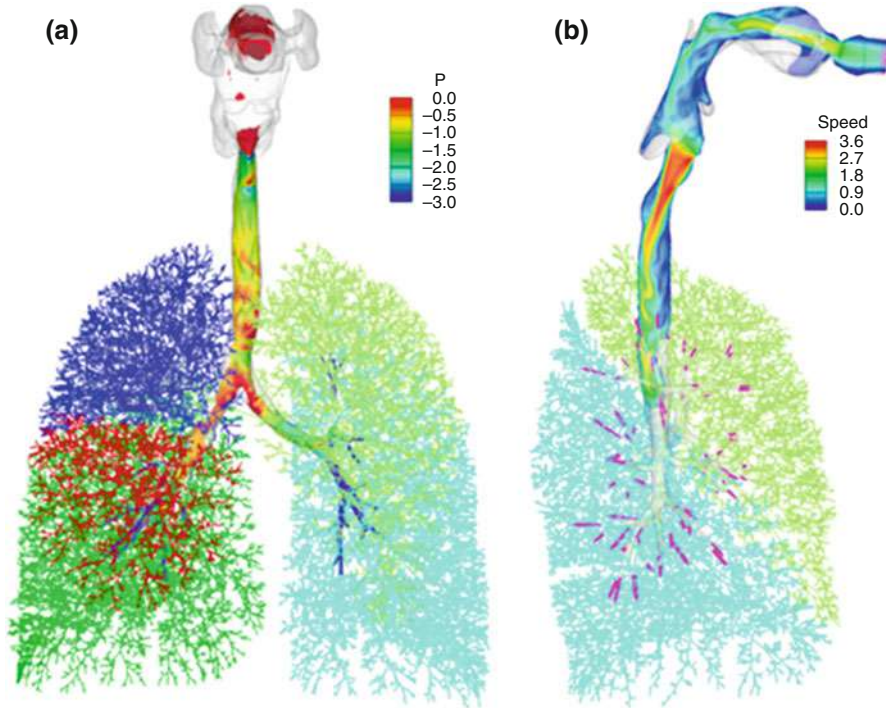
Fig. 3 CFD results of air flow in the multiscale lung model. (**a**) A frontal view of laryngeal jet color-coded with pressure, combined with skeletonized entire conducting airway model color-coded for each lobe; (**b**) A right side view of the laryngeal jet (velocity), with velocity vectors imposed at 3D ending branches (pink) (Lin et al. 2009)

were well mixed and dispersed. Thus, CFD is able to predict airflow structure and particle deposition in different carrier gases for assessment of deposition efficiencies of inhaled aerosols.

3   Machine Learning of CT Imaging Lung Data in Multicenter Setting

3.1   Machine Learning of Asthma Populations

3.1.1   Multiscale Imaging-Based Metrics

Multiscale metrics aim to capture structural and functional phenotypes at different stages of lung diseases. Quantitative CT (QCT) imaging-based metrics are used to characterize asthmatic phenotypes such as airway narrowing and chronic inflammation (Busse and Lemanske 2001). For instance, luminal diameter and wall thickness (Montaudon et al. 2009) measured by QCT imaging indicate the
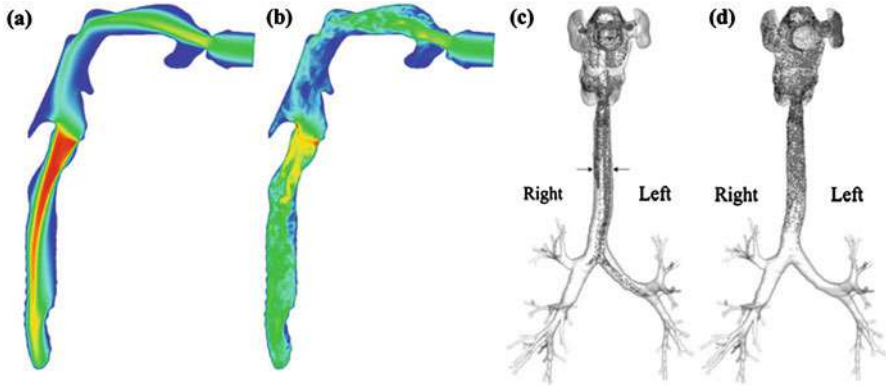
Fig. 4 Instantaneous gas speed distributions in a vertical plane cutting through the center of the trachea looking from right at peak inspiration of normal breathing for (**a**) He and (**b**) Xe-O_2 and the resulting 2.5 μm particle flows in (**c**) He and (**d**) Xe–O_2, viewed from the front (Miyawaki et al. 2012)

quantitative degrees of airway narrowing and airway wall thickening. In addition, an air-trapping metric (voxels < -856 HU on the expiratory image) (Busacker et al. 2009) serves as a measure of what has been termed functional small airway disease or fSAD. These imaging-based metrics have been individually used to identify either structural or functional alteration, but they have not been integrated (until recently) to identify clinically meaningful subgroups. A recent use of ML technique allows for identifying predominant structural and functional variables as well as classifying asthma subpopulations (clusters) acquired by an imaging center (Gupta et al. 2014). However, the results of ML technique could be affected by the quality of given imaging data, as well as the number of samples. For the purpose of deriving unbiased and reproducible clusters with larger data sets, Choi et al. (2017b) developed a multiscale imaging-based cluster analysis (MICA) and employed QCT imaging data of asthmatic subjects acquired by SARP. Furthermore, they improved existing imaging-based metrics by eliminating intersubject and inter-site variations.

SARP (Wenzel and Busse 2007) has contributed to characterize disease phenotypes within a heterogeneous asthmatic population. Besides collecting demography, blood/sputum sample, and PFTs, the multicenter protocols included a CT imaging protocol acquiring images at two static scans at the levels of TLC and functional residual capacity (FRC). The acquisition of two volumetric scans at inspiration and expiration is a critical step in the use of CT for the assessment of both airway structure and lung function. First, registering two static volumetric images provides the means for quantifying peripheral local mechanical strains (Choi et al. 2013c). In addition, airway diameter and wall thickness can be assessed at TLC, while air trapping is evaluated at FRC or RV, and these metrics are integrated through the fact that lung volumes have been linked (Choi et al. 2014, 2015). Such a post-processing
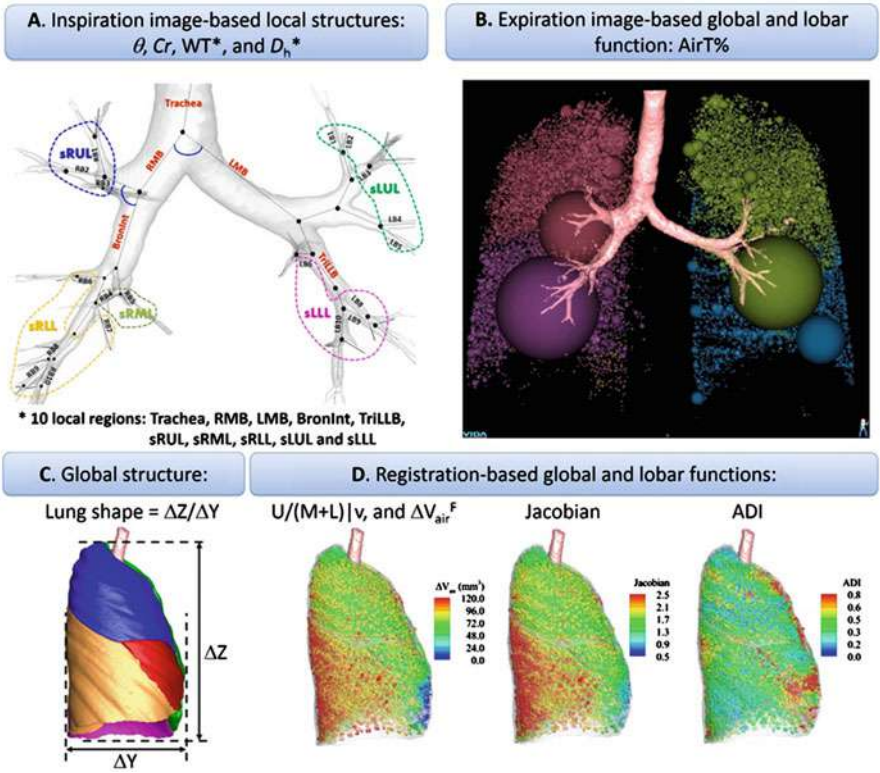
Fig. 5 QCT structural and functional imaging-based metrics obtained at multiscale levels (Choi et al. 2017b)

with image segmentation and image registration successfully identifies structural and functional metrics at multiscale levels.

More specifically, using an automated segmentation software Apollo (VIDA Diagnostics), four local structural variables including branching angle (θ), airway circularity (Cr), wall thickness (WT), and hydraulic diameter (D_h) are extracted from TLC scans. The θ, Cr, WT, and D_h indicate altered skeletal structure, heterogeneous airway shape, airway wall thickening (or thinning), and airway luminal narrowing, respectively. In order to reflect regional features, these metrics are extracted from ten segmental airways (Fig. 5). Using the same software, air trapping (AirT%) is extracted from the FRC image acquired in SARP. (It should be noted that FRC was the lung volume chosen for the expiratory image but more recently SPIROMICS has adopted the more traditional volume of RV for the assessment of air trapping). Furthermore, a lung shape at TLC is extracted to evaluate a global lung structure. Using a customized image matching-based software (Yin et al. 2009), Choi et al. (2017b) obtained three functional variables including fractional regional air volume change (ΔV_{air}^{F}), regional volume change (Jacobian),

and an anisotropic deformation index (ADI) between TLC and FRC. The ΔV_{air}^{F}, Jacobian, and ADI indicate preferential air volume change, local expansion, and nonuniform stretch, respectively, in parenchymal regions. The metrics acquired at the voxel level were reduced to lobar metrics taking fraction (ΔV_{air}^{F}) or medians (Jacobian and ADI) in five lobes and a whole lung (a total of six regions). All of the listed imaging-based metrics at both local (segmental) and global (lobar/whole lung) scales are illustrated in Fig. 5.

3.1.2   Coping with Intersubject Variability

For statistically reliable results from ML, potential bias should be eliminated. In the QCT imaging analysis, the first potential factor causing data bias is intersubject variation. Without using a normalization scheme, there have been conclusions from an asthmatic population suggesting that asthmatics have thinner walls than healthy subjects (Montaudon et al. 2009). This conclusion runs contrary to the general view that asthmatic airways are thicker, at least in severe asthma through a process of airway remodeling Busacker et al. (2009). Choi et al. (2015) observed that airway diameter and wall thickness appear to be significantly associated with gender, age, and height, among them with gender being the most sensitive variable to determine normal airway size. Thus, an unbalanced number of samples due to gender can potentially cause large variations within groups, masking disease-related airway alterations. In the same study, Choi et al. (2015) developed novel normalization methods for airway diameter and wall thickness. Using 61 healthy subjects, they performed multiple linear regression with the given gender, age, and height to derive predicted values of tracheal diameter and wall thickness. These quantities were used to normalize QCT-based D_h and WT, being denoted as D_h^{*} and WT*.

3.1.3   Coping with Inter-site Variability

The second potential factor that may cause statistical bias is associated with inter-site variability. For the purpose of increasing the sample size, a key element of successful machine learning, it is essential to employ multicenter data acquired by multiple scanners. This requires matching imaging protocols (Sieren et al. 2016) since different scanners and associated calibrations may introduced variations in standard CT densities, especially at the lower end of the HU range where air should be represented as -1000 HU and blood should be approximately 55 HU (with water at 0 HU). The difference is critical in evaluating CT density-based measure, e.g., air trapping, and emphysema. Note that HU < -850 (or -856) is commonly used when evaluating air trapping (Busacker et al. 2009). Choi et al. (2014) demonstrated the effects of CT scanners to air-trapping measure between three different imaging centers, namely, University of Iowa, University of Pittsburgh, and Washington University in St. Louis. They concluded that using different CT scanners causes significant variation by data distribution which goes beyond actual patient differences. Therefore, they introduced a normalization scheme to control CT density distribution. This approach employs a fraction-threshold measure rather than density-threshold measure by assuming tracheal CT density on a given scanner

as a representation of air density. The new air-trapping method successfully reduced inter-site variation.

3.1.4   Dimensional Reduction and Clustering

ML techniques, especially unsupervised learning, are very sensitive to the given data distribution. Since Choi et al. (2014, 2015) reduced the potential bias due to intersubject and inter-site variations using newly derived normalization schemes, the use of an unsupervised learning approach has led to the extraction of altered features due to disease rather than due to population and scanner differences. Using QCT imaging-based metrics, Choi et al. (2017b) hypothesized that multiscale imaging-based metrics alone could be employed to derive clinically meaningful clusters. Imaging features are highly correlated such that airway narrowing within the central airway segments may be associated with air trapping in peripheral regions. Since the number of imaging-based metrics is relatively large compared with the number of samples, dimensional reduction via principal component analysis (PCA) was employed. Due to their significant correlations, the 57 imaging-based metrics were reduced to only 7 principal components (PCs). The seven PCs were determined with a parallel analysis with uncorrelated random data. They then used these seven PCs to perform a clustering analysis. K-means, hierarchical, and a Gaussian mixed model with various numbers of clusters were evaluated. They found that K-means with four clusters provided the most reproducible results based on higher Jaccard index than two other methods and other number of clusters.

Fig. 6 A summary of imaging and clinical features between QCT imaging-based clusters (Choi et al. 2017b)

Four imaging-based clusters were successfully identified based upon their unique imaging features (Fig. 6). Cluster 1 demonstrated normal airway feature with hyperinflation measured by Jacobian and ADI. The patients within Cluster 1 were early-onset younger non-severe asthmatics with reversible airflow obstruction. Cluster 2 demonstrated airway narrowing without airway wall thickening. The patients within Cluster 2 were a mix of non-severe and severe asthmatics with marginal inflammation. Cluster 3 demonstrated significant airway wall thickening without airway narrowing. The patients within Cluster 3 were dominated by obese females with reversible airflow obstruction, along with a lack of lymphocyte. Cluster 4 demonstrated most of the alterations such as airway narrowing, wall thickening, and air trapping and reduced lung deformation. The patients within Cluster 4 were late-onset older males with persistent airflow obstruction along with a neutrophil dominant signature.

3.2   Extension of MICA to COPD Populations

With recent advances in unsupervised clustering of patient populations (Paoletti et al. 2009; Cho et al. 2010; Garcia-Aymerich et al. 2011), there has been an increased effort to employ unsupervised learning for the purposes of grouping subpopulations of subjects within COPD communities (Burgel et al. 2012). To extend the above MICA to COPD populations, it is necessary to define new imaging-based variables to capture the COPD heterogeneity due to a complex condition associated with pulmonary and extrapulmonary manifestations.

3.2.1   New QCT Imaging-Based Metrics for COPD

Sixty nine CT imaging-based metrics were employed at both segmental and lobar levels, which is an expanded set of the existing 57 MICA variables used for asthma cluster analysis (Fig. 5). These three new variables are functional small airways disease percentage (fSAD%) to quantify diseased small airways at expiration, emphysema percentage (Emph%) to quantify emphysematous tissues at inspiration, and tissue fraction at TLC (β_{tissue}) indicating the portion of tissue volume in each voxel to assess a possible alteration of local tissue. The new features fSAD% and Emph% can be used to construct a parametric response map (PRM) (Galban et al. 2012) that tracks the progression of a diseased lung. It should be noted that the PRM-based measure of emphysema eliminates emphysematous voxels which on expiration are not air trapped and it also eliminates those voxels on expiration which appear air trapped but on inspiration are found to be emphysematous.

3.2.2   Clustering of Current Smoker Clusters

Using a batch of current smokers ($N = 284$) from SPIROMICS, four statistically stable clusters ($N = 96/45/88/55$) were identified with K-means clustering (Haghighi et al. 2017b). Cluster 1 had relatively normal airway structures; Cluster 2 had wall thickening with an increase of tissue fraction along with an increase of fSAD%; Cluster 3 exhibited decreased wall thickness and luminal narrowing; Cluster 4 had

a significant increase of emphysema and a decrease of luminal diameter along with significant elevation of small airways disease and significant reduction of lung deformation. Association tests of imaging-based clusters with clinical metrics were performed. Cluster 4 showed relatively low FEV1/FVC. In contrast, Cluster 1 showed normal FEV1/FVC with younger age. Association of imaging-based clusters with existing clinical metrics used for diagnosis of COPD suggests the sensitivity of MICA in differentiating clinically meaningful subpopulations.

3.3   Sensitivity of Multiscale Approach in Differentiating Overlap Syndrome

The traditional perspective of differentiating asthma from COPD has relied on reversibility of airflow obstruction by corticosteroids and/or bronchodilators, along with sputum biomarkers to differentiate eosinophilia and neutrophilia. According to recent reports (Gibson and Simpson 2009; De Marco et al. 2013), about 15–45% of patients with COPD may have asthma-COPD overlap, so-called ACO. This implies that two respiratory diseases are difficult to differentiate, because they may share common features. Misdiagnosing these respiratory diseases leads to incorrect treatments. Choi et al. (2017c) hypothesized that multiscale QCT imaging-based metrics could aid in the differentiation of the COPD and asthma.

Choi et al. (2017c) mixed SARP and SPIROMICS data sets together to determine the ability of imaging to differentiate between COPD and asthma. Overall, more significant structural and functional alterations were observed in COPD patients, compared with asthmatic patients. Compared with healthy subjects, asthma and COPD patients demonstrated similar alterations of Cr and $D_h{}^*$. On the other hand, density-based measures between asthma and COPD are significantly different. For example, fSAD% and Emph% are elevated in asthma and COPD but more prominently in COPD subjects. β_{tissue} is significantly increased in asthma and decreased in COPD subjects, compared with healthy subjects.

4   Integration of CFD and MICA for Cluster-Guided Analysis

The information on cluster membership is valuable in several ways because each cluster has unique features of structure and function. For example, it can provide representative patients who can then be used to guide CFD analysis to generate model data on distribution and deposition of inhaled aerosols representative of each individual cluster for the development of improved pulmonary drug delivery approaches. Note that a CFD simulation requires significant computational cost due to the complexity of airway structure and inherent feature of turbulent flows. By studying only small numbers of subjects through their membership in a given cluster, the need for large population CFD efforts can be reduced considerably.

Another example of utilizing the cluster membership is to help understand the success of some interventions, such as bronchial thermoplasty (BT) in severe

asthma. The study of Choi et al. (2017b) identified two severe asthma clusters (Cluster 3 and Cluster 4), one non-severe cluster (Cluster 2) and one severe cluster (Cluster 4) as shown in Fig. 6. Note that Cluster 3 is characterized by airway wall thickening but no airway narrowing, differing from that of Cluster 4 with constricted airways. Proximal airway narrowing may limit transport of inhaled particles into the distal small airways, and that BT enlargement of constricted airway lumens may allow more drugs to penetrate into the distal airways, consequently increasing drug efficacy. Thus, severe asthmatic subjects within Cluster 4 may benefit more from release of airway constriction by BT than those in Cluster 3.

5   Conclusions

This chapter has reviewed the pipeline to construct a subject-specific imaging-based CFD lung model, the development of QCT imaging-based metrics and MICA machine learning to identify homogeneous subpopulations (clusters), and the utilization of cluster membership to guide CFD analysis to generate model data for assessment of therapeutic interventions. The ultimate goal is to advance precision medicine. Both CFD model and MICA are multiscale in nature, within and between individuals, respectively. The former, a realistic personalized lung model, is based on local airway geometry (structure) and global ventilation distribution (function) derived from CT images, spanning from the trachea to parenchyma, as well as physical principles to predict data needed to understand the complex dynamics in the lungs, such as airflow and particle deposition, that are difficult, if not impossible, to be measured in vivo. The latter aims to bridge individual and population scales, enabling application of the former, the CFD lung model, to subpopulations exhibiting similar imaging (structural and functional) and clinical characters. The QCT imaging-based metrics used in ML are extracted at both local and global regions, aiming to capture a wide spectrum of structural and functional phenotypes manifested at different stages of disease progression. As an example, MICA was applied to identify subpopulations in respective asthma and COPD patients. MICA was further applied to a combined cohort of asthma and COPD subjects with overlap syndrome measured by traditional PFTs, demonstrating its sensitivity in identifying significant variables in classification.

Moving forward, it is essential to use longitudinal data to assess stability and transition of cluster membership over time for outcome assessment. It is because all of the methods and analyses presented herein were based on cross-sectional data at one time point. In addition, more advanced machine learning techniques should be employed for more sensitive clustering. For example, deep learning, a machine learning technique based on neural network, can catch the nonlinear features of data and may lead to breakthrough results if equipped with parallel hardware such as GPU and cluster computers to tackle with big data sets. Deep learning can be used to rank the important features in disease severity and then to build a model to predict the severity of each patient in large populations. While PCA can only capture linear structure of the data, t-Distributed Stochastic Neighbor Embedding (t-SNE) focuses

on preservation of the local distance of high-dimensional data and therefore is able to capture nonlinear patterns (van der Maaten and Hinton 2008). Moreover, a self-organizing map, also known as the Kohonen network (Kohonen 1998), is a type of neural network using unsupervised learning for reduction of high-dimensional data and for visualization of data in a two-dimensional space, serving as an alternative method for PCA.

References

Burgel P-R, Roche N, Paillasseur J-L et al (2012) Clinical COPD phenotypes identified by cluster analysis: validation with mortality. Eur Respir J 40:495–496

Busacker A, Newell JD, Keefe T et al (2009) A multivariate analysis of risk factors for the air-trapping asthmatic phenotype as measured by quantitative CT analysis. Chest 135:48–56

Busse WW, Lemanske RF Jr (2001) Asthma. N Engl J Med 344:350–362

Cho MH, Washko GR, Hoffmann TJ, et al (2010) Cluster analysis in severe emphysema subjects using phenotype and genotype data: an exploratory investigation. Respir Res. 11:30. https://doi.org/10.1186/1465-9921-11-30

Choi J (2011) Multiscale numerical analysis of airflow in CT-based subject specific breathing human lungs. The University of Iowa, Iowa City

Choi J, Tawhai MH, Hoffman EA, Lin CL (2009) On intra- and intersubject variabilities of airflow in the human lungs. Phys Fluids 21(10):101901. Epub 2009 Oct 13

Choi J, Xia G, Tawhai MH et al (2010) Numerical study of high-frequency oscillatory air flow and convective mixing in a CT-based human airway model. Ann Biomed Eng 38(12):3550–3571. https://doi.org/10.1007/s10439-010-0110-7. Epub 2010 Jul 8

Choi J, Hoffman EA, Lee CH et al (2013a) MDCT-based image matching for assessment of heterogeneity of regional ventilation and methacholine response in asthmatics. American Thoracic Society International Conference, May 17–22, Philadelphia, PA. Am J Respir Crit Care Med 187:A3740

Choi J, Yin Y, Hoffman EA et al (2013b) Airflow in a multiscale subject-specific breathing human lung model. 66th APS division of fluid dynamics gallery of fluid motion, Nov. 24–26, Pittsburgh, PA. arXiv:1310.5057 [physics.flu-dyn]

Choi J, Hoffman EA, Lin CL et al (2017a) Quantitative computed tomography determined regional lung mechanics in normal nonsmokers, normal smokers and metastatic sarcoma subjects. PLoS One 12(7):e0179812. https://doi.org/10.1371/journal.pone.0179812

Choi S, Hoffman EA, Wenzel SE et al (2013c) Registration-based assessment of regional lung function via volumetric CT images of normal subjects vs. severe asthmatics. J Appl Physiol 115:730–742

Choi S, Hoffman EA, Wenzel SE et al (2014) Improved CT-based estimate of pulmonary gas trapping accounting for scanner and lung volume variations in a multi-center study. J Appl Physiol 117:593–603

Choi S, Hoffman EA, Wenzel SE et al (2015) Quantitative assessment of multiscale structural and functional alterations in asthmatic populations. J Appl Physiol 118:1286–1298

Choi S, Choi J, Hoffman EA et al (2016) Relationship between pulmonary airflow and resistance in patients with airway narrowing using an 1-D network resistance and compliance model. 69th APS Division of Fluid Dynamics, Nov. 20–22, Portland

Choi S, Hoffman EA, Wenzel SE et al (2017b) Quantitative computed tomographic imaging-based clustering differentiates asthmatic subgroups with distinctive clinical phenotypes. J Allergy Clin Immunol 140:690–700

Choi S, Haghighi B, Choi J et al (2017c) Differentiation of quantitative CT imaging phenotypes in asthma versus COPD. BMJ Open Resp Res 4(1):e000252

Choi S, Choi J, Lin CL (2018) Contributions of kinetic energy and viscous dissipation to airway resistance in pulmonary inspiratory and expiratory airflows in successive symmetric airway models with various bifurcation angles. J Biomech Eng 140(1):011010

Collins FS, Varmus H (2015) A new initiative on precision medicine. N Engl J Med 372(9):793–795. https://doi.org/10.1056/NEJMp1500523. Epub 2015 Jan 30

Couper D, LaVange LM, Han M et al (2014) Design of the subpopulations and intermediate outcomes in COPD study (SPIROMICS). Thorax 69(5):491–494. https://doi.org/10.1136/thoraxjnl-2013-203897. Epub 2013 Sep 12

Crum WR, Hartkens T, Hill DL (2004) Non-rigid image registration: theory and practice. Br J Radiol 77 Spec No 2:S140–S153

De Marco R, Pesce G, Marcon A et al (2013) The coexistence of asthma and chronic obstructive pulmonary disease (COPD): prevalence and risk factors in young, middle-aged and elderly people from the general population. PLoS One 8:e62985

Deo RC (2015) Machine learning in medicine. Circulation 132:1920–1930

Ellingwood ND, Yin Y, Smith M, Lin CL (2016) Efficient methods for implementation of multi-level nonrigid mass-preserving image registration on GPUs and multi-threaded CPUs. Comput Methods Prog Biomed 127:290–300. https://doi.org/10.1016/j.cmpb.2015.12.018. Epub 2016 Jan 6

Fuld MK, Halaweish AF, Newell JD Jr et al (2013) Optimization of dual-energy xenon-computed tomography for quantitative assessment of regional pulmonary ventilation. Investig Radiol 48(9):629–637. https://doi.org/10.1097/RLI.0b013e31828ad647

Galban CJ, Han MK, Boes JL (2012) Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. Nat Med 18(11):1711–1715. https://doi.org/10.1038/nm.2971

Garcia-Aymerich J, Gómez FP, Benet M et al (2011) Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. Thorax 66:430–437

Gibson P, Simpson J (2009) The overlap syndrome of asthma and COPD: what are its features and how important is it? Thorax 64:728–735

Gupta S, Hartley R, Khan UT et al (2014) Quantitative computed tomography-derived clusters: redefining airway remodeling in asthmatic patients. J Allergy Clin Immunol 133:729–738

Haghighi B, Ellingwood N, Yin Y et al (2017a) A GPU-based symmetric non-rigid image registration method in human lung. Med Biol Eng Comput. https://doi.org/10.1007/s11517-017-1690-2. Epub 2017 Aug 1 [ahead of print]

Haghighi B, Choi J, Choi S, et al. (2017b) Cluster-specific small airway modeling for imaging-based CFD analysis of pulmonary air flow and particle deposition in COPD smokers. 70th APS Division of Fluid Dynamics, Nov. 19–21, 2017, Denver

Hoffman EA, Chon D (2005) Computed tomography studies of lung ventilation and perfusion. Proc Am Thorac Soc 2(6):492–498, 506

Hoffman EA, McLennan G (1997) Assessment of the pulmonary structure-function relationship and clinical outcomes measures: quantitative volumetric CT of the lung. Acad Radiol 4(11):758–776

Hoffman EA, Lynch DA, Barr RG et al (2016) Pulmonary CT and MRI phenotypes that help explain chronic pulmonary obstruction disease pathophysiology and outcomes. J Magn Reson Imaging 43(3):544–557. https://doi.org/10.1002/jmri.25010. Epub 2015 Jul 22

Iyer KS, Newell JD Jr, Jin D et al (2016) Quantitative dual-energy computed tomography supports a vascular etiology of smoking-induced inflammatory lung disease. Am J Respir Crit Care Med 193(6):652–661. https://doi.org/10.1164/rccm.201506-1196OC

Jahani N, Yin Y, Hoffman EA, Lin CL (2014) Assessment of regional non-linear tissue deformation and air volume change of human lungs via image registration. J Biomech 47(7):1626–1633. https://doi.org/10.1016/j.jbiomech.2014.02.040. Epub 2014 Mar 13

Jahani N, Choi S, Choi J et al (2015) Assessment of regional ventilation and deformation using 4D-CT imaging for healthy human lungs during tidal breathing. J Appl Physiol 119(10):1064–1074. https://doi.org/10.1152/japplphysiol.00339.2015. Epub 2015 Aug 27

Jahani N, Choi S, Choi J et al (2017) A four-dimensional computed tomography comparison of healthy and asthmatic human lungs. J Biomech 56:102–110. https://doi.org/10.1016/j.jbiomech.2017.03.012. Epub 2017 Mar 18

Kleinstreuer C, Zhang Z (2003) Laminar-to-turbulent fluid-particle flows in a human airway model. Int J Multiphase Flow 29(2):271–289

Kohonen T (1998) The self-organizing map. Neurocomputing 21(1):1–6. https://doi.org/10.1016/S0925-2312(98)00030-7

Lambert AR, O'Shaughnessy PT, Tawhai MH et al (2011) Regional deposition of particles in an image-based airway model: large-eddy simulation and left-right lung ventilation asymmetry. Aerosol Sci Technol 45(1):11–25. https://doi.org/10.1080/02786826.2010.517578

Ledford H (2016) Obama's science legacy: betting big on biomedical science. Nature 536(7617):385–386. https://doi.org/10.1038/536385a

Lin CL, Tawhai MH, McLennan G, Hoffman EA (2007) Characteristics of the turbulent laryngeal jet and its effect on airflow in the human intra-thoracic airways. Respir Physiol Neurobiol 157(2–3):295–309. https://doi.org/10.1016/j.resp.2007.02.006. Epub 2007 Feb 14

Lin CL, Tawhai MH, McLennan G, Hoffman EA (2009) Computational fluid dynamics: multiscale simulation of gas flow in subject-specific models of the human lung. IEEE Eng Med Biol Mag 28(3):25–33. https://doi.org/10.1109/memb.2009.932480

Lin CL, Tawhai MH, Hoffman EA (2013) Multiscale image-based modeling and simulation of gas flow and particle transport in the human lungs. Wiley Interdisc Rev Syst Biol Med 5(5):643–655. https://doi.org/10.1002/wsbm.1234. Epub 2013 Jul 10

Longest PW, Tian G, Walenga RL, Hindle M (2012) Comparing MDI and DPI aerosol deposition using in vitro experiments and a new stochastic individual path (SIP) model of the conducting airways. Pharm Res 29(6):1670–1688

Magoulas GD, Prentza A (2001) Machine learning in medical applications. Mach. Learning Appl. (Lecture Notes Comput. Sci.). Berlin/Heidelberg, Springer, 2049:300–307

Mesko B (2017) The role of artificial intelligence in precision medicine. Expert Rev Precis Med Drug Dev 2(5):239–241. https://doi.org/10.1080/23808993.2017.1380516

Miyawaki S, Tawhai MH, Hoffman EA, Lin CL (2012) Effect of carrier gas properties on aerosol distribution in a CT-based human airway numerical model. Ann Biomed Eng 40(7):1495–1507

Miyawaki S, Choi S, Hoffman EA, Lin CL (2016a) A 4DCT imaging-based breathing lung model with relative hysteresis. J Comput Phys 326:76–90. https://doi.org/10.1016/j.jcp.2016.08.039. Epub 2016 Aug 31

Miyawaki S, Hoffman EA, Lin CL (2016b) Effect of static vs. dynamic imaging on particle transport in CT-based numerical models of human central airways. J Aerosol Sci 100:129–139. https://doi.org/10.1016/j.jaerosci.2016.07.006. Epub 2016 Jul 16

Miyawaki S, Tawhai MH, Hoffman EA et al (2017a) Automatic construction of subject-specific human airway geometry including trifurcations based on a CT-segmented airway skeleton and surface. Biomech Model Mechanobiol 16:583–596. https://doi.org/10.1007/s10237-016-0838-6

Miyawaki S, Hoffman EA, Lin CL (2017b) Numerical simulations of aerosol delivery to the human lung with an idealized laryngeal model, image-based airway model, and automatic meshing algorithm. Comput Fluids 148:1–9. https://doi.org/10.1016/j.compfluid.2017.02.008. Epub 2017 Feb 10

Montaudon M, Lederlin M, Reich S et al (2009) Bronchial measurements in patients with asthma: comparison of quantitative thin-section CT findings with those in healthy subjects and correlation with pathologic findings. Radiology 253:844–853

Paoletti M, Camiciottoli G, Meoni E et al (2009) Explorative data analysis techniques and unsupervised clustering methods to support clinical assessment of chronic obstructive pulmonary disease (COPD) phenotypes. J Biomed Inform 42:1013–1021

Sieren JP, Newell JD Jr, Barr RG et al (2016) SPIROMICS protocol for multicenter quantitative computed tomography to phenotype the lungs. Am J Respir Crit Care Med 194(7):794–806

Tawhai MH, Lin CL (2011) Airway gas flow. Compr Physiol 1:1135–1157

Tawhai MH, Pullan AJ, Hunter PJ (2000) Generation of an anatomically based three-dimensional model of the conducting airways. Ann Biomed Eng 28(7):793–802. https://doi.org/10.1114/1.1289457

Tawhai MH, Hunter P, Tschirren J et al (2004) CT-based geometry analysis and finite element models of the human and ovine bronchial tree. J Appl Physiol 97(6):2310–2321. https://doi.org/10.1152/japplphysiol.00520.2004. Epub 2004 Aug 20

Tawhai MH, Hoffman EA, Lin CL (2009) The lung physiome: merging imaging-based measures with predictive computational models of structure and function. Wiley Interdiscip Rev Syst Biol Med 1(1):61–72

Uppaluri R, Mitsa T, Sonka M et al (1997) Quantification of pulmonary emphysema from lung computed tomography images. Am J Respir Crit Care Med 156(1):248–254

Uppaluri R, Hoffman EA, Sonka M et al (1999) Computer recognition of regional lung disease patterns. Am J Respir Crit Care Med 160(2):648–254

van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605

Wenzel SE, Busse WW (2007) Severe asthma: lessons from the severe asthma research program. J Allergy Clin Immunol 119:14–21

Wongviriyawong C, Harris RS, Greenblatt E et al (2013) Peripheral resistance: a link between global airflow obstruction and regional ventilation distribution. J Appl Physiol 114(4):504–514

Wu D, Tawhai MH, Hoffman EA, Lin CL (2014) A numerical study of heat and water vapor transfer in MDCT-based human airway models. Ann Biomed Eng 42(10):2117–2131. https://doi.org/10.1007/s10439-014-1074-9

Wu D, Miyawaki S, Tawhai MH et al (2015) A numerical study of water loss rate distributions in MDCT-based human airway models. Ann Biomed Eng 43(11):2708–2721. https://doi.org/10.1007/s10439-015-1318-3. Epub 2015 Apr 14

Xu Y, Sonka M, McLennan G et al (2006a) MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. IEEE Trans Med Imaging 25(4):464–475

Xu Y, van Beek EJ, Hwanjo Y et al (2006b) Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM). Acad Radiol 13(8):969–978

Yang X, Pei J, Shi J (2014) Inverse consistent non-rigid image registration based on robust point set matching. Biomed Eng Online. 13 Suppl 2:S2. https://doi.org/10.1186/1475-925X-13-S2-S2. Epub 2014 Dec 11

Yin Y, Hoffman EA, Lin CL (2009) Mass preserving nonrigid registration of CT lung images using cubic B-spline. Med Phys 36:4213–4222

Yin Y, Choi J, Hoffman EA et al (2010) Simulation of pulmonary air flow with a subject-specific boundary condition. J Biomech 43(11):2159–2163. https://doi.org/10.1016/j.jbiomech.2010.03.048

Yin Y, Choi J, Hoffman EA et al (2013) A multiscale MDCT image-based breathing lung model with time-varying regional ventilation. J Comput Phys 244:168–192

Zhang Z, Kleinstreuer C, Kim CS (2002) Micro-particle transport and deposition in a human oral airway model. J Aerosol Sci 33(12):1635–1652

Computational Crystal Structure Prediction: An Introduction

113

Cai-Zhuang Wang and Christopher Wolverton

Contents

Abstract

Accurate and efficient computational tools for crystal structure prediction play an important role in accelerating the pace of new materials discovery and technological advances. In the last decade, there has been considerable progress in this area due to advances in both the development of computational algorithms and the rapid growth in the computational power of modern computers. The chapters in this section will describe the development and current status of algorithms and methods for crystal structure prediction and present several successful examples of utilization of these tools for the discovery of materials across a variety of applications.

C.-Z. Wang (✉)
Ames Laboratory – USDOE and Department of Physics and Astronomy, Iowa State University, Ames, IA, USA
e-mail: wangcz@ameslab.gov

C. Wolverton
Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA
e-mail: c-wolverton@northwestern.edu

1   Introduction

Advanced materials are essential for energy and economic security. The urgent demand for new energy technologies has greatly exceeded the capabilities of today's materials and chemical processes. Accurate and fast computational structure/property determination would complement traditional experimental efforts in materials design and discovery and accelerate the pace of technological advances. Additionally, the combinatorics of multicomponent compounds suggests that there are large numbers of currently unknown phases awaiting discovery. Hence, the opportunity for new materials discovery through a synergy of computational crystal structure prediction and experimental synthesis is very promising.

Knowledge about the structure of a crystal or nanoparticle at the atomic level is essential for the accurate calculation and understanding of materials properties. However, there are large numbers of experimentally synthesized compounds for which the crystal structure has not been solved. From the materials design and discovery perspective, it is also desirable to have reliable computational tools that can use modern high-performance computers to predict material structures based solely on chemical compositions with no prior assumptions about the system. However, to predict crystal or nanoparticle structures solely from the knowledge of chemical composition has long been recognized as a grand challenge in condensed matter physics, materials science, and solid-state chemistry (Maddox 1988; Eberhart and Clougherty 2004; Woodley and Catlow 2008). This "crystal structure prediction" problem is particularly challenging because it involves optimization (energy minimization) over a complex energy landscape with many local minima, and the number of these minima grows exponentially with the size of the system.

Early attempts for finding the lowest-energy structure of a given number of atoms and chemical composition involved simulated annealing using molecular dynamics (MD) or Monte Carlo (Kirkpatrick et al. 1983; Wille 1986; Doll et al. 2007). However, these approaches often prove computationally inefficient, since an effective exploration of the structure-energy landscape requires easily overcoming the transition barriers among different local minima in the system. In the past two decades, a number of algorithms and methods have been developed to tackle this challenging problem. As early as 1995, Ho et al. introduced a genetic algorithm (GA) scheme for global structure optimization by adopting a real space representation of the structures and a geometrical cut-and-paste operation to generate offspring from parent structures (Deaven and Ho 1995). This type of GA approach was successfully applied to structure optimization for atomic clusters (Deaven and Ho 1995; Ho et al. 1998; Lu et al. 2010), surfaces (Chuang et al. 2004, 2005), and interfaces (Zhang et al. 2009; Chua et al. 2010), as well as nanowires (Chan et al. 2006; Lu et al. 2007). In 2008, Oganov extended the GA method to the case of crystal structure prediction (Oganov and Glass 2008; Lyakhov et al. 2013) and consequently developed a computational crystal structure prediction package called Universal Structure Predictor: Evolutionary Xtallography (USpex)

(http://uspex-team.org/en/) which is now widely used in the community. Another widely used package for crystal structure prediction called CALYPSO (http://www.calypso.cn/) was introduced by Ma et al. in 2010 (Wang et al. 2010, 2012) which is based on a particle swarm optimization strategy. In both GA and particle swarm optimizations, the most time-consuming part of the computation is the structural relaxation and energy evaluation. These portions of the algorithm can be accelerated using classical empirical potentials but with the caveat of limited accuracy and predictive power. On the other hand, first-principles calculations based on density functional theory have increased accuracy, but with concomitant increased computational cost. In order to overcome this problem, in 2014 Wang et al. proposed an adaptive genetic algorithm (Wu et al. 2014; Zhao et al. 2014) which uses classical potentials for the structural optimization and energy evaluation, but the potentials are adaptively trained on the fly by first-principles calculations. In parallel to methods based on GA or particle swarm optimization, other approaches based on basin or minima hopping strategies have also been developed (Wales and Doye 1997; Goedecker 2004; Amsler and Goedecker 2010). These approaches modify the traditional simulated annealing method to more efficiently accelerate the sampling of the energy landscape of a system through MD simulations. Recently, Liu et al. 2017 extended the basin hopping strategy and introduced a stochastic surface walking (SSW) method targeting both global potential energy surface (PES) exploration and reaction pathway sampling (Shang and Liu 2013; Shang et al. 2014). In addition to the methods mentioned above, other methods such as ab initio random structure search method (Pickard and Needs 2011) and methods combining experimental information with computational approaches for crystal structure solution (Meredig and Wolverton 2013) have also been developed.

The methods for crystal structure prediction mentioned above have been successfully applied to materials discovery in a large variety of energy-related areas in the past decade, including superconducting materials, superhard materials, magnetic materials, and energy storage, solar energy, and thermal energy materials. For example, combining the CALYPSO method for crystal structure prediction with first-principles electron-phonon-coupling calculations has led to predictions of many unexpected superconducting phases in H-rich compounds at high pressure. Strikingly, a T_c of about 80 K predicted for dense H_2S (Li et al. 2014) stimulated the experimental discovery and synthesis of H_2S and H_3S phases above megabar pressures with superconducting T_c values in the range of 30–203 K (Drozdov et al. 2015). Very recently, a theoretical prediction of a high pressure superconducting phase of LaH_{10} using the methods mentioned above has also been successfully confirmed by experimental synthesis of this compound (Liu et al. 2017). Evidence for record-breaking superconductivity of $T_c \sim$ 260 K for this hydrogen-rich compound has also been reported (Somayazulu et al. 2019). These examples, among others given in the following chapters in this section, demonstrate that advances in computational methods for crystal structure prediction can greatly accelerate the pace of new materials discovery and technological advances.

2   Contributions

In the first edition of the *Handbook of Materials Modeling* published in 2005, the topic of crystal structure prediction and materials discovery was not included since most of the exciting development in this area has just emerged in the last 10 years. The purpose of this section is to give overviews and tutorials on some of the important algorithms and methods developed in this area along with their applications in design and discovery for various energy-related materials. The contributions to this part are briefly outlined below.

▶ Chapter 114, "CALYPSO Method for Structure Prediction and Its Applications to Materials Discovery," provides an overview of the basic theory and main features of the CALYPSO approach, which is based on a particle swarm optimization strategy, for atomic level of structure prediction based only on chemical composition of the system. Applications of the method to design and discovery of superconductors and superhard materials are also discussed.

▶ Chapter 115, "Adaptive Genetic Algorithm for Structure Prediction and Application to Magnetic Materials," introduces an adaptive genetic algorithm (AGA) method for atomic-level structure prediction. This method uses a genetic algorithm (GA) to explore the atomic configuration space with computationally fast classical interatomic potentials, but the potentials are adaptively trained on the fly by first-principles calculations along with the GA search. Therefore, the method holds the speed of classical potential searches and at the same time retains the accuracy of first-principles calculations for energy evaluations. Applications of the method to discovery of novel magnetic materials are discussed. Note that the concept of on-the-fly adaptive interatomic potential fitting can also be applied to other optimization methods (e.g., particle swarm in CALYPSO).

▶ Chapter 116, "Multi-objective Optimization as a Tool for Material Design" deals with materials design where multi-objective optimizations are required. It explains the concepts of Pareto optimality and Pareto dominance and uses these concepts in solving multi-objective (MO) optimization problems. Several practical case studies are presented, and the efficiency of this method is assessed for optimization of double-objective problems.

▶ Chapter 117, "Minima Hopping Method for Predicting Complex Structures and Chemical Reaction Pathways" describes the minima hopping method (MHM) for predicting complex structures and chemical reaction pathways. The MHM is a structure prediction method that uses short molecular dynamic escape trials to explore potential energy landscapes and a fingerprinting approach to avoid sampling previously visited minima. The MHM has been successfully employed in studying molecular clusters, bulk solids, and surfaces. The application of MHM to the discovery of high pressure materials is also discussed.

▶ Chapter 118, "Stochastic Surface Walking Method and Applications to Real Materials" introduces a stochastic surface walking (SSW) method for both global potential energy surface (PES) exploration and reaction pathway sampling. The SSW PES sampling is automated and unbiased and takes into account second derivative information which allows the structural evolution to follow low-energy

pathways while being able to overcome high barriers. Several examples of PES exploration by combining SSW with first principles and neural network potential calculations are presented. These examples illustrate the power of SSW for unbiased PES exploration and reaction pathway searching.

▶ Chapter 119, "First-Principles-Assisted Structure Solution: Leveraging Density Functional Theory to Solve Experimentally Observed Crystal Structures" discusses the first-principles-assisted structure solution (FPASS) method. FPASS is a genetic algorithm-based approach, but optimizes based on a hybrid of computational (minimization of energies from density functional theory) and experimental (refinement to known, experimental powder diffraction pattern) information. The goal of FPASS is to solve crystal structures with reduced guidance from experts typically required for Rietveld refinement alone. A number of case studies are presented that show the breadth of structure types that have been solved using this method, including those run in a high-throughput mode.

▶ Chapter 120, "Computational Modeling and the Design of Perovskite Solar Cells" focuses on materials modeling for solar energy applications. It discusses the key material parameters for photovoltaic application such as the bandgap, effective mass, and optical absorptions. It also takes perovskite solar cell as an example, to show how materials modeling and high-throughput computational screening of crystal structure-property relationship can help understand and engineer practical materials for solar energy applications.

3 Future Perspectives

It should be pointed out that despite considerable progress and success in methodology development and applications of crystal structure prediction and materials discovery over the past 10 years, there still a lot of room for further exploration. For example, the rapid progress in artificial intelligence and data science has opened the door to possible new approaches to computational crystal structure prediction and materials discovery. Future work in integrating crystal structure prediction algorithms with the advances in machine learning and materials informatics will be very fruitful.

It is also worthwhile to note that knowledge of stable structures does not provide detailed synthesis pathways. It is highly desirable to know whether and how any identified new materials can be eventually synthesized experimentally. The theoretical prediction of synthesis conditions leading to a desired structure remains a grand challenge in computational materials science. Prediction of synthesis pathways will require not only thermodynamic free energies at finite temperatures for the phases of interest but also for competing phases and decomposition pathways.

The future of structure prediction methods should also address other limitations of current approaches. Structure prediction of defects and interfaces remains a difficult problem for most methods. Also, nearly all the methods in this section deal with perfectly ordered, stoichiometric compounds. A large fraction of known material phases are partially ordered or disordered, and incorporation of these

aspects into structure prediction methods will enable a much larger search space of potential new compounds. Finally, most of the methods in this section are predicted on energy minimization as a route to materials discovery. However, a large number of known compounds are either metastable or stabilized at non-ambient conditions (e.g., high temperatures) and hence are not zero-temperature ground states. Crystal structure prediction for these non-ground-state phases (where straightforward energy minimization will not be sufficient) remains another challenge for future advances.

References

Amsler M, Goedecker G (2010) Crystal structure prediction using the minima hopping method. J Chem Phys 133:224104

Chan TL, Ciobanu CV, Chuang F-C, Lu N, Wang CZ, Ho KM (2006) Magic structures of H-passivated 110 silicon nanowires. Nano Lett 6(2):277

Chua ALS, Benedek NA, Chen L, Finnis MK, Sutton AP (2010) A genetic algorithm for predicting the structures of interfaces in multicomponent systems. Nat Mat 9:418

Chuang FC, Ciobanu CV, Shenoy VB, Wang CZ, Ho KM (2004) Finding the reconstructions of semiconductor surfaces via a genetic algorithm. Surf Sci Lett 573:L375

Chuang FC, Liu B, Wang CZ, Chan T-L, Ho KM (2005) Global structural optimization of Si magic clusters on the Si(111) 7 × 7 surface. Surf Sci Lett 598:L339–L346

Deaven DM, Ho KM (1995) Molecular geometry optimization with a genetic algorithm. Phys Rev Lett 75:288–291

Doll K, Schön JC, Jansen M (2007) Global exploration of the energy landscape of solids on the ab initio level. Phys Chem Chem Phys 9:6128–6133

Drozdov AP, Eremets MI, Troyan IA, Ksenofontov V, Shylin SI (2015) Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system. Nature 525:73–76

Eberhart ME, Clougherty DP (2004) Looking for design in materials design. Nat Mater 3:659–661

Goedecker S (2004) Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. J Chem Phys 120:9911–9917

Ho KM, Shvartsburg A, Pan BC, Lu ZY, Wang CZ, Wacker J, Fye JL, Jarrold MF (1998) Structures of medium-sized silicon clusters. Nature 392:582

Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Li Y, Hao J, Liu H, Li Y, Ma Y (2014) The metallization and superconductivity of dense hydrogen sulfide. J Chem Phys 140:174712

Liu H, Naumov II, Hoffmann R, Ashcroft NW, Hemley RJ (2017) Potential high-T_c superconducting lanthanum and yttrium hydrides at high pressure. Proc Natl Acad Sci U S A 114:6990–6995

Lu N, Ciobanu CV, Chan TL, Chuang F-C, Wang CZ, Ho KM (2007) The structure of ultrathin H-passivated [112] silicon nanowires. J Phys Chem C 111:7933

Lu WC, Wang CZ, Zhao LZ, Zhang W, Qin W, Ho KM (2010) Appearance of bulk-like motifs in Si, Ge, and Al clusters. Phys Chem Chem Phys 12:8511

Lyakhov AO, Oganov AR, Stokes H, Zhu Q (2013) New developments in evolutionary structure prediction algorithm USPEX. Comput Phys Commun 184:1172–1182

Maddox J (1988) Crystals from first principles. Nature 335:201

Meredig B, Wolverton C (2013) A hybrid computational–experimental approach for automated crystal structure solution. Nat Mater 12:123–127

Oganov AR, Glass CW (2008) Evolutionary crystal structure prediction as a tool in materials design. J Phys Condens Matter 20:064210

Pickard CJ, Needs RJ (2011) Ab initio random structure searching. J Phys C Condens Matter 23:053201

Shang C, Liu ZP (2013) Stochastic surface walking method for structure prediction and pathway searching. J Chem Theory Comput 9:1838–1845

Shang C, Zhang XJ, Liu ZP (2014) Stochastic surface walking method for crystal structure and phase transition pathway prediction. Phys Chem Chem Phys 16:17845–17856

Somayazulu M, Ahart M, Mishra AK, Geballe ZM, Baldini M, Meng Y, Struzhkin VV, Hemley RJ (2019) Evidence for superconductivity above 260 K in lanthanum superhydride at megabar pressures. Phys Rev Lett 122:027001

Wales D, Doye J (1997) Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J Phys Chem A 101:5111–5116

Wang Y, Lv J, Zhu L, Ma Y (2010) Crystal structure prediction via particle-swarm optimization. Phys Rev B 82:094116

Wang Y, Lv J, Zhu L, Ma Y (2012) CALYPSO: a method for crystal structure prediction. Comput Phys Commun 183:2063–2070

Wille LT (1986) Searching potential energy surfaces by simulated annealing. Nature 324:46–48

Woodley SM, Catlow R (2008) Crystal structure prediction from first principles. Nat Mater 7: 937–946

Wu SQ, Ji M, Wang CZ, Nguyen MC, Zhao X, Umemoto K, Wentzcovitch RM, Ho KM (2014) Adaptive genetic algorithm for crystal structure prediction. J Phys Condens Matter 26:035402

Zhang J, Wang CZ, Ho KM (2009) Finding the low-energy structures of Si[001] symmetric tilted grain boundaries with a genetic algorithm. Phys Rev B 80:174102

Zhao X, Shu Q, Nguyen MC, Wang Y, Ji M, Xiang H, Ho KM, Gong X, Wang CZ (2014) Interface structure prediction from first-principles. J Phys Chem C 118:9524–9530

CALYPSO Method for Structure Prediction and Its Applications to Materials Discovery

114

Yanchao Wang, Jian Lv, Quan Li, Hui Wang, and Yanming Ma

Contents

Abstract

Atomic-level structure prediction for condensed matters, given only a chemical composition, is a major challenging issue in a broad range of science (e.g., physics, chemistry, materials, and planetary science, etc.). By combining the global swarm optimization algorithm with a number of specially designed structure-dealing techniques (e.g., symmetry constraints, structure fingerprints, etc.), we developed the CALYPSO (Crystal structure AnaLYsis by Particle Swarm Optimization) structure prediction method that is able to predict the

Y. Wang (✉) · J. Lv · H. Wang
State Key Laboratory of Superhard Materials and Innovation Center of Computational Physics Methods and Software, College of Physics, Jilin University, Changchun, China
e-mail: wyc@calypso.cn

Q. Li · Y. Ma (✉)
State Key Laboratory of Superhard Materials and Innovation Center of Computational Physics Methods and Software, College of Physics, Jilin University, Changchun, China

International Center of Future Science, Jilin University, Changchun, China
e-mail: mym@jlu.edu.cn

structures of a wide range of materials including isolated clusters/nanoparticles, two-dimensional layers and reconstructed surfaces, and three-dimensional bulks and holds the promise for the functionality-driven design of materials (e.g., superhard, electride, and optical materials, etc.). It has been demonstrated in a wide range of applications that CALYPSO is highly efficient when searching for the structures of materials and becomes an invaluable tool for aiding materials discovery. In this chapter, we provide an overview of the basic theory and main features of the CALYPSO approach, as well as its versatile applicability to the design of superconductors and superhard materials. Finally, the conclusion and opportunities for further research on CALYPSO method are presented.

1 Introduction

The traditional way of discovering new materials is by Edisonian iterations of trial and error, which choose materials to synthesize and test depending on chemical intuition, experience, or theoretical knowledge until some combination of success and exhaustion is achieved. This approach is painstaking, long, and costly due to the extraordinary number of experimental attempts required. Advances in computing now allow new materials with target properties to be simulated and analyzed theoretically before their synthesis (Agrawal and Choudhary 2016; Jain et al. 2013), significantly reducing the necessary trials (and errors) and thus the expected cost of discovery (Jain et al. 2016; Oganov et al. 2010; Needs and Pickard 2016).

A material's structure determines its properties and functionalities. Therefore, theoretical structure prediction is critical in aiding materials discovery (Wang and Ma 2014). The goal of crystal structure prediction is to develop an efficient computational scheme to explore the high-dimensional potential-energy surface (PES) in order to identify the global energy minimum that corresponds to the global stable ground-state structure observable in experiment. Empirical observations and heuristic estimates indicate that the number of local minima grows exponentially as the system size increases (Stillinger 1999; Gavezzotti 1994). Therefore, an efficient computational approach is urgently needed to find quickly the most stable structure of a large assembly of atoms in the vast configuration space of the PES.

Various strategies to generate hypothetical structures have been recently pursued in the field of crystal structure prediction, and several efficient structure prediction methods (e.g., data mining (Gavezzotti 1994), simulated annealing (Kirkpatrick et al. 1983; Schön and Jansen 1996), genetic algorithms (Deaven and Ho 1995; Oganov and Glass 2006; Abraham and Probert 2006; Trimarchi and Zunger 2007; Kolmogorov et al. 2010), minima hopping (Goedecker 2004; Amsler and Goedecker 2010), basin hopping (Wales and Doye 1997; Doye and Wales 1998), metadynamics (Laio and Parrinello 2002), random sampling (Pickard and Needs 2011; Shang and Liu 2013), and CALYPSO (Wang et al. 2010)) have been developed. Their applications to materials science have led to a number of breakthroughs, as exemplified by a series of theory-driven discoveries of novel functional materials (Ceder 2010; Ma et al. 2009; Oganov et al. 2009; Zhu et al. 2011; Pickard and Needs 2006; Drozdov

et al. 2015; Ceder et al. 1998; Teter 1998; Li et al. 2014). Structure prediction is now becoming an irreplaceable tool to accelerate the discovery of materials. Several reviews consider structure prediction methods and their applications (Wang and Ma 2014; Woodley and Catlow 2008; Oganov 2011; Ciobanu et al. 2013; Wang et al. 2015; Zhang et al. 2017a), but this work focuses only on our CALYPSO method and its application to superconductive and superhard materials.

This chapter is organized as follows. Section 2 briefly introduces the basic theory and general features of the CALYPSO methodology, including swarm intelligence algorithm. Several recent applications of CALYPSO in discovering superconductive and superhard materials are provided in Sect. 3, followed by general conclusions and future prospects in Sect. 4.

2 CALYPSO Structure Prediction Method

2.1 General Theory of the CALYPSO Method

The CALYPSO method for structure prediction run comprises several main steps (Fig. 1). Initial structures are randomly generated with symmetry constraints. Once a new structure is generated, structure characterization is performed to examine its similarity to all the previous ones, where duplicate or similar structures are eliminated. After all the structures of each population have been generated, local structural optimizations are performed to reduce the noise of the energy surface and drive the systems to local minima. Structural evolution via the particle swarm optimization algorithm is then carried out to produce new structures for the next generation. These steps are iterated until a termination criterion (such as a prescribed threshold or a fixed number of iterations) is attained (Wang et al. 2012a). Below we give an introduction to the general theory of the CALYPSO method.

2.1.1 Generation of Random Structures with Symmetry Constraints

Most natural structures of materials are governed by symmetry rules; any crystal structure must belong to one of the 230 crystallographic space groups. For a given space group, only a subset of the optimized variables (lattice parameters and atomic coordinates) is independent, while the rest are determined via symmetry rules. It is found that the majority of randomly generated structures (particularly for

Fig. 1 The flow chart of CALYPSO

large systems, e.g., >30 atoms per simulation cell) without symmetry constraints exhibits unreasonably low symmetries, while proper consideration of symmetry constraints can significantly reduce the optimization variables and enrich the diversity of structures during structure evolution. Symmetry constraints on random structure generation are therefore implemented in CALYPSO to improve the starting structures for the subsequently evolutionary structure searches. Different strategies for dealing with 0D-isolated molecule/clusters, 2D layers or surfaces, and 3D crystals are designed to generate random symmetric structures. In particular, the generation of random structures is constrained by $\sim$40 point groups for isolated systems (e.g., molecules, clusters, and nanoparticles), 17 planar space groups for 2D layers or surfaces, and 230 space groups for 3D crystals. For example, the space group of initial crystal structures is randomly selected among 230 ones for a crystal structure prediction. Once the space group is selected, the lattice parameters are generated within the chosen symmetry according to the defined volume, and the corresponding atomic coordinates are obtained by a combination of a set of symmetrically related coordinates (Wyckoff positions) in accordance to the number of atoms in the simulation cell. Notably, a backup database of space groups of all generated structures is established in order to compare them with the space groups of newly generated structures. The appearance of identical space group is forbidden with a certain probability (80%). This allows the initial samplings to cover different regions of the search space and generate the diverse structures, which are found to be crucial for a high efficiency of a global minimization.

To demonstrate the effects of symmetric constraints on structure generation, bulk crystals (Wang et al. 2012a) and isolated clusters (Lv et al. 2012) are chosen as two test systems, where crystalline TiO_2 consists of 16 formula units per cell, while the cluster contains 100 atoms described by the Lennard–Jones potential. For each system, a large number of structures (3,250 for the crystal and 10,000 for the cluster) were randomly generated with and without symmetric constraints, respectively, and then locally optimized using the GULP code (Gale 1997). The obtained energy distributions of crystal and cluster are shown in Fig. 2a and b, respectively. The results (Fig. 2a) showed that the rutile structure, i.e., the global stable structure, cannot be generated without symmetry constraints. However, once symmetric constraints were employed in the generation of random structures, 203 ($\sim$6.2% of the total) rutile structures were successfully produced. The generated cluster structures (Fig. 2b) without symmetric constraints have a Gaussian-like energy distribution with a well-defined sharp peak centered at a high energy of 0.3 ε/atom. Many of the cluster structures appeared disordered or liquid-like; their generation greatly reduced structural diversity. Instead, once point group symmetry was applied, a much broader energetic distribution emerged with the generation of a large number of lower-energy structures.

2.1.2   Structural Characterization Techniques

During the structural evolution process, many trial structures are produced for each generation, and some may be very similar or even identical. The existence of similar structures decreases the diversity of the population, potentially leading to stagnation.
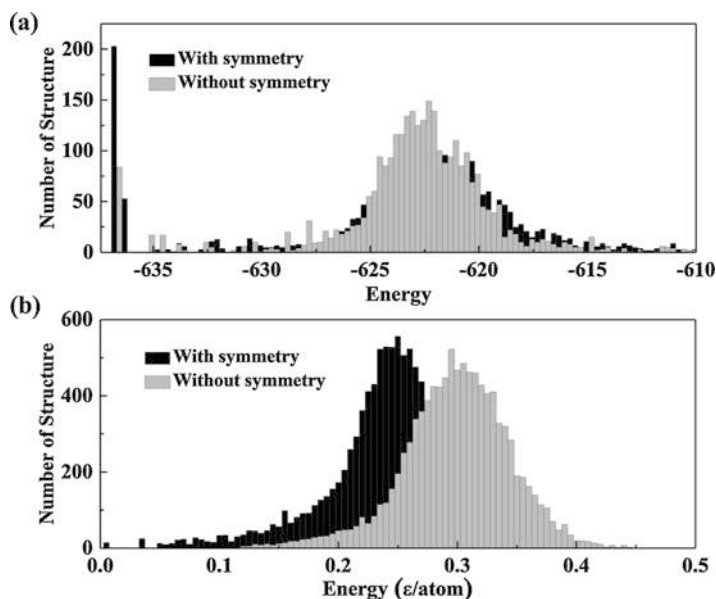
Fig. 2 Energetic distributions of randomly generated structures for (**a**) TiO_2 and (**b**) LJ_{100} clusters with and without symmetries

Moreover, given that the most computationally expensive parts of the algorithm are the energy calculations and local structural optimizations, repetitive optimizations of similar structures will significantly increase the computational cost and impede the structure searching efficiency. Thus, efficient searching must employ techniques to fingerprint each structure, measure its similarities to other structures, and eliminate similar structures. Two effective methods, the bond characterization matrix (BCM) (Lv et al. 2012) and the coordination characterization function (CCF) (Su et al. 2017), have been developed based on the geometrical information of a structure.

The BCM is an advanced version of the bond-orientational order parameter technique originally introduced by Steinhardt et al. (Steinhardt et al. 1983). For a given structure, a bond vector r_{ij} between atoms i and j is defined if the interatomic distance is less than a given cutoff distance. This vector is associated with the spherical harmonics $Y_{lm}(\theta_{ij}, \varphi_{ij})$, where θ_{ij} and φ_{ij} are the polar angles. A weighted average over all bonds formed by, for instance, the type A and B atoms is then evaluated by Eq. 1:

$$Q_{lm}^{\delta_{AB}} = \frac{1}{N_{\delta_{AB}}} \sum_{i \in A, j \in B} e^{-\alpha(r_{ij} - b_{AB})} Y_{lm}\left(\theta_{ij}, \phi_{ij}\right), \tag{1}$$

where δ_{AB} and $N_{\delta_{AB}}$ denote the type and the number of a bond, respectively. To avoid dependence on the choice of reference frame, it is important to consider the rotationally invariant combinations:

$$Q_l^{\delta_{AB}} = \sqrt{\frac{4\pi}{2l+1}\sum_{m=-l}^{l}\left|Q_{lm}^{\delta_{AB}}\right|^2}, \tag{2}$$

where each series of $Q_l^{\delta_{AB}}$ for $l = 0, 2, 4, 6, 8,$ and 10 can be used to represent a type of bond and is thus an element of the BCM. Only even-l spherical harmonics are used to guarantee invariant bond information with respect to the direction of the bonds. By including the complete structural information of bond lengths and angles, this technique unambiguously fingerprints the structures. As a result, the similarity of any two structures (u and v) can be quantitatively evaluated by the Euclidean distance between their BCMs:

$$D_{uv} = \sqrt{\frac{1}{N_{type}}\sum_{\delta_{AB}}\sum_l\left(Q_l^{\delta_{AB},u} - Q_l^{\delta_{AB},v}\right)^2}, \tag{3}$$

where N_{type} is the number of bond types.

Two LJ_{38} clusters with different structural motifs are chosen as examples to test the BCM. The first structure is the global stable structure, having a face-centered cubic structural motif (LJ_{38}-O_h), while the second one is a metastable incomplete Mackay icosahedron (LJ_{38}-C_{5v}). Their BCMs, shown by histograms in Fig. 3a, are dramatically different. Except for Q_2 and Q_{10}, the LJ_{38}-O_h structure has obviously larger Q_l values than those in the LJ_{38}-C_{5v} structure. The resultant distance between these two structures is a very large value of 0.618, characterizing their large structural differences. The distances to the isotropic system are 0.728 for LJ_{38}-O_h and 0.163 for LJ_{38}-C_{5v}, illustrating a high degree of order in LJ_{38}-O_h. Randomly distorting LJ_{38}-O_h, and plotting its distances to the unaltered structure as a function of distortion (Fig. 3b), illustrates that the distances increase linearly with the distortions at the initial stage, reflecting correctly the structural deviations. Eventually, the curves become flat at a distortion of about $0.4r_0$. At this stage, the distances are approaching those distances relative to the isotropic limits favorable for disordered systems.

Although the BCM can provide a precise measurement of structures, it suffers from heavy computational costs for large systems (e.g., the number of atoms per simulation cell >30) since all the geometrical information (bond lengths and angles) is considered and evaluated. To reduce the computational cost, CCF, an alternative and fast method depending only on the two-body correlation function (interatomic distance), was developed to fingerprint large structures. The CCF method employs a matrix involving different atomic types:

$$M = \begin{bmatrix} ccf_{11} & \cdots & ccf_{1Nt} \\ \vdots & \ddots & \vdots \\ ccf_{Nt1} & \cdots & ccf_{NtNt} \end{bmatrix}. \tag{4}$$
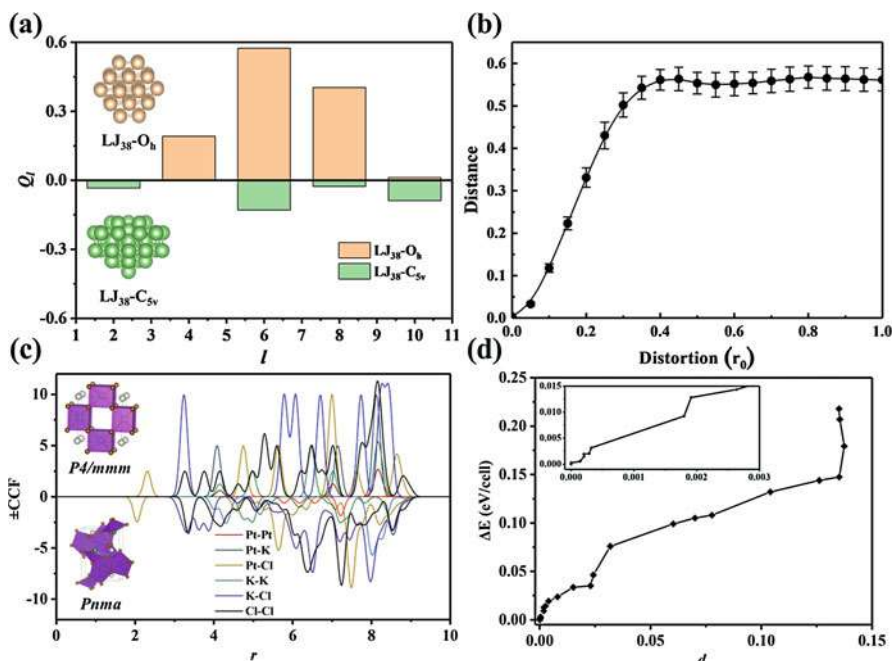
Fig. 3 (**a**) Histograms of BCMs for LJ$_{38}$-O_h and LJ$_{38}$-C_{5v}. (**b**) BCM distances with respect to the unaltered structures as a function of distortions for LJ$_{38}$-O_h. Each data point was obtained by averaging over the distances of 100 randomly distorted structures, and the error bar denotes the standard deviation. (**c**) Comparisons of the CCFs of different structures of $P4/mmm$ and $Pnma$ for PtK$_2$Cl$_4$. For ease of comparison, the plots show -CCFs for the $Pnma$ structure. (**d**) Calculated ΔE and structural distances between the optimized structure of CuInS$_2$ and all intermediate structures in each ionic step. Inset shows the detailed optimization procedure approaching the equilibrium point

Here, each matrix element related to different pairs of atomic types i and j is calculated as follows:

$$
\mathrm{ccf}_{ij}(r) =
\begin{cases}
\frac{1}{N}\sum_{n_i}\sum_{n_j} f\left(r_{n_i n_j}\right)\sqrt{\frac{a_{pw}}{\pi}}\exp\left[-a_{pw}\left(r - r_{n_i n_j}\right)^2\right], & (i \neq j), \\
\frac{1}{2N}\sum_{n_i}\sum_{n_j} f\left(r_{n_i n_j}\right)\sqrt{\frac{a_{pw}}{\pi}}\exp\left[-a_{pw}\left(r - r_{n_i n_j}\right)^2\right], & (i = j),
\end{cases}
\tag{5}
$$

where n_i runs over all the atoms of the i-th type within the cell, n_j runs over all the atoms of the j-th type within the extended cell, $r_{n_i n_j}$ is the interatomic distance less than the cutoff radius (usually 9.0 Å), $f\left(r_{n_i n_j}\right)$ is the weighting function for different interatomic distances, N is the number of atoms in the cell, a_{pw} is an empirical parameter that controls the peak width of the Gaussian function, and $\sqrt{\frac{a_{pw}}{\pi}}\exp\left[-a_{pw}\left(r - r_{n_i n_j}\right)^2\right]$ is the normalized Gaussian function. As a result, the similarity of two structures can be estimated by the distance between their CCF

matrices. The Pearson correlation coefficient was employed to measure the degree of similarity between two matrices, and the corresponding distance (d) is defined as $d = 1 - R$ from their correlation coefficient (R). A detailed description of the CCF can be found in Su et al. (2017).

Two distinct crystal structures of PtK_2Cl_4 with space groups of *P4/mmm* and *Pnma* are chosen to test the CCF method. The CCFs calculated for the two structures (Fig. 3c) are clearly rather different, with a large calculated distance of 0.655, giving a good measure on the degree of dissimilarity between the two structures. Note that our tests recommend a user-defined threshold of 0.075 as a good number to identify dissimilar structures. CCF is a continuous function with respect to the motion of the atoms. Through the optimization of a $CuInS_2$ structure with *Pbam* symmetry, it is seen from the calculated ΔE and structural distances between the optimized structure and the intermediate structures of each ionic step (Fig. 3d) that the structure closest to the equilibrium point tends to have the smaller ΔE and distance values.

2.1.3 Structural Evolution via Particle Swarm Optimization Algorithm

The CALYPSO method adopts a "self-improving" strategy to locate the global minimum of the PES via particle swarm optimization (PSO) (Kennedy 2011). The PSO algorithm is a typical swarm-intelligence scheme inspired by natural biological systems (e.g., ants, bees, or birds) and has been applied to a variety of fields in engineering and chemistry (Ourique et al. 2002; Call et al. 2007). Its application to extended systems for structure prediction started only recently (Wang et al. 2010, 2012b). Within the CALYPSO method, the structures at the $t + 1^{th}$ generation (x^{t+1}) evolve in the energy landscape through velocity. In practice, the lattice parameters (unit cell) of new structures remain unaltered or are randomly generated, while the atomic positions are updated according to the following formula:

$$x^{t+1} = x^t + v^{t+1}. \tag{6}$$

It is noteworthy that the velocity plays a significant role in governing the speed and direction of particle movement. The updated velocity (v^{t+1}) is calculated based on its previous location (x^t), previous velocity (v^t), current location $(pbest^t)$ derived from geometrical optimization of this individual, and the global best location $(gbest^t)$ for the entire population as follows:

$$v^{t+1} = wv^t + c_1r_1 \left(pbest^t - x^t \right) + c_2r_2 \left(gbest^t - x^t \right), \tag{7}$$

where w denotes the inertia weight controlling the momentum of the particle, which dynamically varies and decreases linearly from 0.9 to 0.4 during the iteration in our methodology; the self-confidence factor (c_1) and swarm confidence factor (c_2) are equal to 2, which give the best overall performance; r_1 and r_2 are random numbers distributed in the range [0, 1]. The velocity formula includes random parameters (r_1) and (r_2) that ensure good coverage of the searching space and avoid entrapment in local optima. The initial velocity of each structure is generated randomly, while
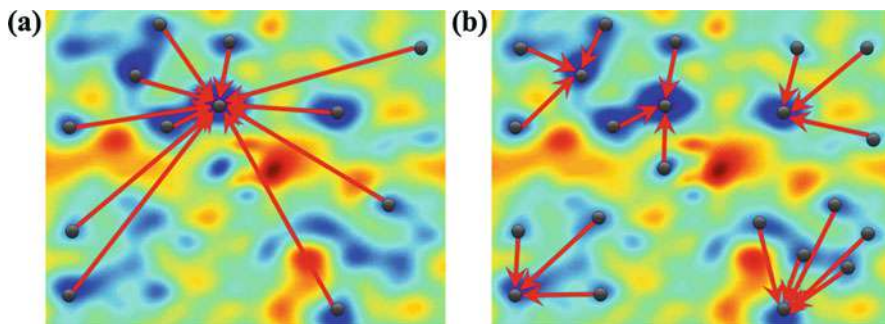
Fig. 4 The schematic diagram of global (**a**) and local (**b**) PSO

the updated one is calculated using both its individual properties and those of the entire population.

Two versions of the PSO algorithm (local and global) have been implemented in CALYPSO (Lv et al. 2012; Wang et al. 2011). The global PSO, outlined in Fig. 4a, has only one global best structure acting as the learning example or attractor for the entire structure population, and all particles seek new positions only in the regions close to the uniquely overall best position. This method converges quickly for small systems (i.e., those with fewer than $\sim$30 atoms in the simulation cell), but it may be less effective for larger systems because the PES becomes much more complex. The local PSO method is outlined in Fig. 4b. Each particle (i.e., a candidate structure) selects a set of other particles as its neighbors, and its velocity is adjusted according to both its position and the best position achieved so far in the community formed by its neighborhood. Thus, at each iteration, the particle will move toward its own best position and the best position of its local neighborhood, rather than the overall best position in the swarm. By maintaining multiple attractors corresponding to different regions of the PES, the local PSO allows for a finer exploration of the PES and can effectively avoid stagnation during the structure searches.

2.1.4   Local Structure Optimization

The PES of a material can be regarded as a multidimensional system of many peaks and valleys connected by saddle points. Given that the global minimum is one of the possibly many local minima, local structure searching is an inevitable part of the global search and therefore should be included in any efficient structure searching method. The CALYPSO method currently has interfaces with various ab initio and force-field-based total-energy packages (e.g., VASP (Kresse and Furthmüller 1996), SIESTA (Soler et al. 2002), Quantum ESPRESSO (Giannozzi et al. 2009), CASTEP (Segall et al. 2002), CP2K (Vandevondele et al. 2005), LAMMPS (Plimpton 1995), and GULP (Gale 1997)) for local structural optimization. Other external total-energy programs can also be easily interfaced with CALYPSO as required.

2.2   Features of the CALYPSO Method

CALYPSO can be used to perform unbiased searches of the energetically sta-
ble/metastable structures of a given chemical composition (Wang et al. 2015,
2016a). Several attractive features have been implemented in CALYPSO to resolve
various structure prediction problems, including 3D solids (Wang et al. 2010), 2D
layers (Wang et al. 2012b; Luo et al. 2011) and atomic adsorption (Gao et al.
2015), 2D surfaces (Lu et al. 2014), 0D nanoclusters or molecules (Lv et al.
2012), X-ray diffraction data-assisted structure searches (Gao et al. 2017a), and the
inverse design of novel functional materials (e.g., superhard, electride, and optical
materials) (Zhang et al. 2013, 2017b; Xiang et al. 2013).

2.2.1   3D Crystals

A crystallographic structure can be regarded as an infinitely repeating array of 3D
boxes, termed unit cells, represented by lattice parameters and atomic coordinates.
As shown in Fig. 5a, there are six lattice parameters: three lengths and three angles
of the lattice vectors. Each atom has three coordinates coded as a fraction of each of
the lattice vectors. Structure searches of 3D crystals evolve both types of variables
simultaneously to achieve the global minimum on the PES with $3N + 3$ dimensions,
where $3N - 3$ degrees of freedom are the atomic positions, and the remaining six
dimensions are the lattice parameters. To reduce the search space, several hard
constraints, including the minimum interatomic distances, the minimum lattice
lengths, and constraints on the range of angles between the lattice vectors, are
imposed in the CALYPSO method. For example, the distance between two atoms
should not be smaller than a threshold or the sum of their covalent radii, and the
angles between lattice vectors are free to vary between $30°$ and $160°$. An advantage
gained by using these constraints is to force attention onto regions containing
physically reasonable minima. The CALYPSO method has been widely applied
to the prediction of crystal structures for various systems (Wang and Ma 2014;
Zhu et al. 2011; Lv et al. 2011; Nishio-Hamane et al. 2012; Li et al. 2016; Chen
et al. 2013), some of which have already been experimentally confirmed (see, e.g.,
Guillaume et al. (2011), Bai et al. (2015), Ma et al. (2012), and Yang et al. (2017)).
Section 3 describes its applications to the discovery of novel superconductive and
superhard materials.

2.2.2   2D Layers and Atomic Adsorption

A 2D material is defined as having a finite thickness in one dimension and an
infinite extent in the other two. The slab model is employed to simulate 2D-layered
structures in the CALYPSO method (Wang et al. 2012b). As illustrated in Fig. 5b
and c, these models contain two regions: the layered material and vacuum. The
vacuum ensures that the studied layers are isolated from their periodic images.
Any 2D-layered structure can be characterized by 1 of 17 planar space groups. To
ensure unbiased sampling of the energy landscape, the lattice parameters and atomic
positions are randomly generated with 2D symmetry constraints. A distortion
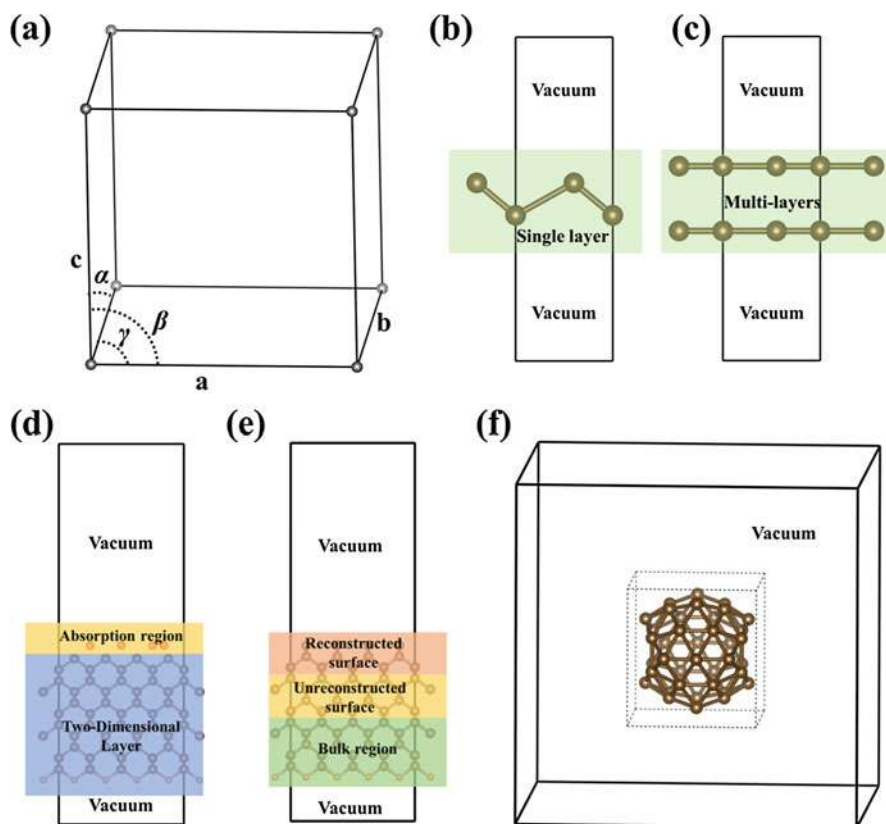
Fig. 5 The models used in CALYPSO for (**a**) 3D crystal structure, (**b**) single-layer structure, (**c**) multilayer structure, (**d**) 2D atomic adsorption structure, (**e**) surface reconstruction structure, and (**f**) cluster structure

parameter perpendicular to the in-plane layer (z in Fig. 5b) can also be activated to search for buckled layers, and a van der Waals gap parameter (i.e., the distance between two adjacent layers, as in Fig. 5c) is introduced for multilayered systems. Structural evolution in 2D space is achieved by a constrained PSO algorithm. The CALYPSO method has been widely applied to various 2D materials with exotic structures, including Be_5C_2 (Wang et al. 2016b), FeB_2 (Zhang et al. 2016), and Si (Luo et al. 2014). Because of their intriguing structures and properties, these materials are expected to be promising candidates for future applications in specific fields (Gu et al. 2017).

The adsorption of atoms can efficiently functionalize two-dimensional layer materials with desirable properties. Therefore, structural information about the atoms adsorbed on 2D layers is crucial for understanding their functional properties. A new method was developed to predict the structures of atoms adsorbed on 2D-layered materials based on the CALYPSO methodology. The structural model

used is shown in Fig. 5d; it contains three regions: vacuum, adsorption region, and 2D substrate. Several specially designed techniques (e.g., fixed adsorption sites, symmetry constraints, and a constrained PSO algorithm) were employed to improve the search efficiency. Application of our method to investigate fully and partially hydrogenated graphene and graphene oxide successfully predicts their energetically most favorable structures (Gao et al. 2015). In general, our method is promising for the prediction and design of structures of atoms adsorbed on 2D-layered materials (Li et al. 2015a; Zhou et al. 2016).

2.2.3 2D Surfaces

Surfaces play a decisive role in determining the properties and processing of almost all engineering materials, especially at the nanoscale. However, surface structures are often elusive, impeding significantly the engineering of devices. To predict surface structures, we developed an efficient method based on the CALYPSO methodology (Lu et al. 2014). Structure searches for surfaces are conducted using a slab model comprising three regions (Fig. 5e): the bulk material region, the unreconstructed surface, and the reconstructed surface. Generally, the bulk region (6–8 layers) remains fixed to preserve the bulk nature of the material, and the bottom side of the slabs is passivated by hydrogen atoms with integer or partial charges. The atoms in the unreconstructed surface region (usually composed of 2–4 layers) are subject to local structural relaxation, but they are not involved in the structural evolution. In contrast, the atoms in the reconstructed surface region are fully evolved during structure searching, and the choice of thickness for this region depends strongly on the specific energetic situation of the system. The surface excess free energy is calculated as the fitness during structure evolution by $\triangle \gamma \left(\mu \right) = \frac{1}{A} \left(E_{surf}^{tot} - E_{ideal}^{tot} - \sum_i n_i * \mu_i \right)$, where A is the area of the studied surface (usually a 1×1 planar unit cell), E_{surf}^{tot} and E_{ideal}^{tot} are the total energies of the reconstructed and unreconstructed surfaces, respectively; and n_i and μ_i denote the number and chemical potential of the ith species in the surface region. Additionally, several specially designed methods have been incorporated into the search procedure to improve its efficiency. For example, symmetry operations of various two-dimensional space groups are applied, while the initial surface structures and surfaces of semiconductors are constructed using the electron-counting rules (Lu et al. 2014). Our approach is evaluated via its application to various semiconductor surface reconstructions (Gao et al. 2017b; Xu et al. 2017), including C, Si, SiC, AlN, and ZnO. Experimentally observed surfaces are readily reproduced by the CALYPSO method from input knowledge of only the chemical compositions, thus validating our approach (Lu et al. 2014).

It is noteworthy that application of new method to a simple diamond (100) surface reveals an unexpected surface reconstruction featuring self-assembled carbon nanotubes arrays (Lu et al. 2014). Such a surface is energetically competitive with the known dimer structure under normal conditions, but it becomes more favorable under a small compressive strain or at high temperatures. The novel surface

structure exhibits a unique feature of carrier kinetics (i.e., one dimensionality of hole states, while two dimensionality of electron states) that could lead to novel design of superior electronics. The finding of a previously unknown formation of self-assembled carbon nanotubes on the diamond (100) surface highlights the power of our intelligent surface structure searching method.

2.2.4   0D Nanoclusters or Molecules

Clusters or molecules belong to 0D nonperiodic materials systems. They usually exhibit geometrical frustration due to the competition between surface and bulk, leading to various structural motifs that do not exist in periodic system. The CALYPSO method has been generalized to perform structure searches for these 0D systems (Call et al. 2007). In contrast to periodic systems with translational symmetry, only the point group symmetries are utilized in generating candidate structures of isolated systems, and Cartesian coordinates are used straightforwardly to represent structures. To comply with usual total energy calculations requiring periodic boundary conditions, a big box (Fig. 5f) is built, and the cluster is located at its center. The vacuum surrounding the clusters should be large enough to avoid interactions between the cluster and its periodic images. The main algorithm of PSO is properly revised to allow the structural evolution of 0D systems. The module for cluster structure searching has been extensively benchmarked using Lennard–Jones cluster systems of various sizes (up to 150 atoms), and high search efficiency was achieved, demonstrating the reliability of the current methodology (Lv et al. 2012). Many cluster systems have been investigated using the CALYPSO method (Lv et al. 2014, 2015; Lu et al. 2016; Li et al. 2013a), leading to the discovery of several intriguing structures with interesting chemical bonding. For example, we have used CALYPSO to investigate the structures for neutral boron clusters containing 38 atoms. A symmetric cage-like configuration is found to be the global minimum structure, which can be seen as an all-boron fullerene. Furthermore, a quasi-planar structure with a double-hexagonal vacancy is also revealed with nearly degenerate energy. This quasi-planar structure is recently confirmed by experiments as the ground-state structure for the anionic B_{38} cluster (Chen et al. 2017).

2.2.5   X-Ray Diffraction Data-Assisted Structure Searches

X-ray diffraction (XRD) is the most powerful technique for determining crystal structure information at the atomic level. However, it remains challenging to determine the crystal structure from only experimental powder XRD data because estimated structural information (e.g., unit cell parameters and space group) is required. This is exemplified by about half of the 300,000 Powder Diffraction Files having some unrefined atomic coordinates (Meredig and Wolverton 2012). Traditional structure prediction uses only energy as the fitness function, and the development of a structure prediction method that can effectively use XRD data is highly desirable. This issue is recently addressed by a first-principles-assisted structure solution (FPASS) method (Meredig and Wolverton 2012), which will be
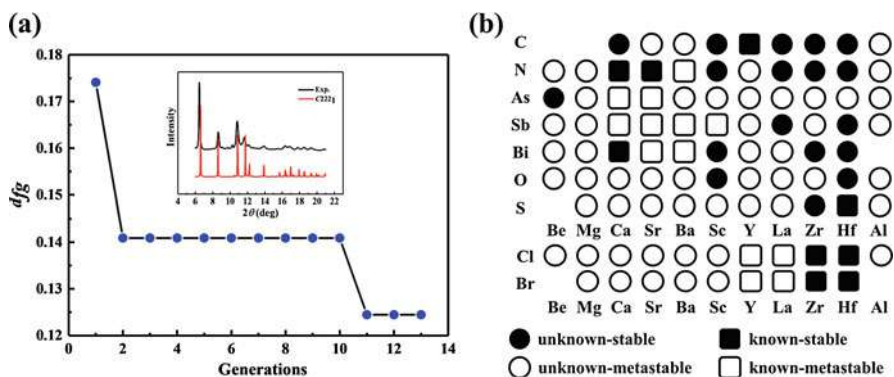
Fig. 6 (**a**) CALYPSO structure search history for $CaLi_2$. Inset shows experimental and simulated XRD patterns of predicted structure. (**b**) Inverse-design results for binary electrides. Stability map of A_2B (top) and AB (bottom) electrides. A and B elements are shown in horizontal and vertical directions, respectively

given in detail in the other chapter of this book. We also propose a versatile global search method based on the CALYPSO method to determine crystal structures from experimental powder XRD data without guessed structural information (Gao et al. 2017a). This search method uses the degree of dissimilarity between the simulated and experimental XRD patterns as the fitness (instead of energy), and a weighted cross-correlation function is used to compare the dissimilarity of two powder diffraction patterns. The efficiency and robustness of the new method are demonstrated by exploring the high-pressure phase of a binary compound, $CaLi_2$. The resultant (Fig. 6a) candidate structure of $C222_1$ for $CaLi_2$ at 54 GPa gave a simulated XRD pattern matching with the experimental data (Debessai et al. 2008).

2.2.6 Inverse Design of Functional Materials

Metastable structures can exist in nature even though they are not thermodynamically stable. However, traditional structure prediction methods that concentrate on finding the ground-state structure of the PES may miss these metastable structures. It is therefore highly desirable to develop an inverse-design approach to design functional materials directly from certain target properties (Franceschetti and Zunger 1999; Zhang et al. 2015a). The inverse design aims to establish an efficient computational scheme to explore the structure–functionality landscape. The fitness function here is the intrinsic material property such as hardness, bandgap, or interstitial electron localization rather than energy as used in traditional structure searching methods. Our CALYPSO method implements a number of computer-assisted inverse-design techniques to search for functional materials (Zhang et al. 2013, 2017b; Xiang et al. 2013) (e.g., superhard, optical, and electride materials). Here, we demonstrate the performance of the CALYPSO inverse-design method by applying it to the design of inorganic electride materials by screening 99 binary compounds (Zhang et al. 2017b). Before our work, the known inorganic electrides

at ambient pressure are only limited to several of them, e.g., C12A7 (Patel 2003) and Ca_2N and its variants (Lee et al. 2013; Inoshita et al. 2014; Tada et al. 2014). With the use of the advanced inverse-design method, we are able to identify 89 new electrides, among which 18 are existing compounds that have not been identified as electrides (solid squares, Fig. 6b), while the other 71 electrides (circles, Fig. 6b) are unknown compounds that have not been synthesized, of which 18 (solid circles, Fig. 6b) are thermodynamically stable in their lowest-energy states and dynamically stable with the absence of any imaginary frequency in phonon spectra; the other 53 (void circles, Fig. 6b) are metastable with positive formation energies. In reality, these metastable electrides might be synthesizable, as various metastable electrides have already been synthesized (void squares, Fig. 6b). Our work provides a useful tool for aiding the discovery of electride materials and reveals the rich abundance of inorganic electrides in nature.

3 Materials Discovery Using CALYPSO

3.1 Superconductors

Superconductivity is among the most exciting properties in condensed-matter physics, and considerable effort in the past century has been devoted to the discovery of new superconductors (Bardeen et al. 1957). A significant boost came with the establishment of the microscopic theory of superconductivity by Bardeen, Cooper, and Schrieffer (BCS) (Bardeen et al. 1957) in 1957, which provides important guidance for the design of superconductors with high T_c (critical temperature for superconductivity). According to BCS theory, T_c is intimately tied to one key parameter, the electron–phonon coupling (EPC) constant, which can be estimated by first-principles calculations based on the band structure and Eliashberg theory (Eliashberg 1960). This EPC constant is clearly the critical factor in the design of high-T_c superconductors and is strongly correlated with crystal structure. The determination of crystal structure is therefore a critical step in the design of novel superconductors.

Advances in crystal structure prediction in the past decades have greatly aided the design of superconductors; many have been predicted theoretically, and some confirmed experimentally. Here, we provide an overview of recent advancements in the discovery of superconductors aided by the CALYPSO method. As most of the discoveries are related to hydrides, the overview is divided into two parts: one concerning hydrogen-rich compounds and the other non-hydrogen compounds.

3.1.1 Hydrogen-Rich Compounds

It has been suggested that the lightest element, H, forms metallic solids with sufficiently strong EPC necessary for a high-T_c phonon-mediated superconductivity at high pressures (Ashcroft 1968). However, the metallization of hydrogen had not been observed at low temperature at up to 388 GPa (Dalladay-Simpson et al. 2016). Metallization of solid H at 495 GPa has recently been reported (Dias and Silvera
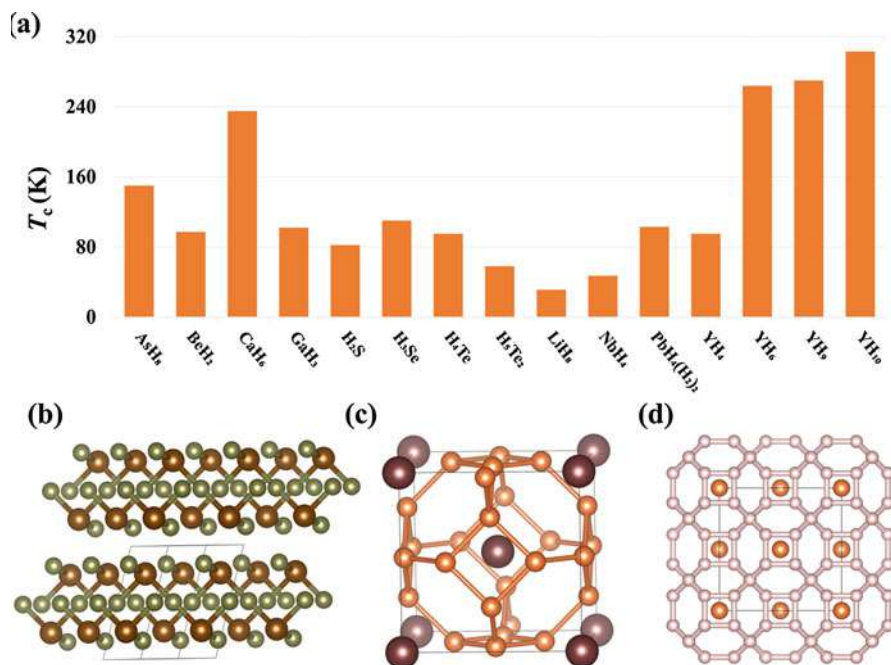
Fig. 7 (**a**) Histogram of calculated T_c of some hydrogen-rich compounds predicted by CALYPSO. Crystal structures of (**b**) H_2S, (**c**) CaH_6, and (**d**) YH_{10}

2017), but additional experimental measurements are required to verify this claim (Liu et al. 2017a; Silvera and Dias 2017). Hydrogen-rich compounds have been considered as an alternative because they are expected to metallize at considerably lower pressures owing to the chemical "precompression" caused by other elements (Ashcroft 2004). Combining the CALYPSO method with first-principles EPC calculation leads to predictions of many unexpected superconducting phases in H-rich compounds (Zhang et al. 2017a; Wang et al. 2017). Theoretically calculated T_c values for some hydrogen-rich compounds are summarized in Fig. 7a. These works suggest the possibility of superconductivity with estimated T_c values up to 303 K. Strikingly, a T_c of about 80 K predicted for dense H_2S has stimulated experimental observation of hydrogen sulfide superconductors above megabar pressures with T_c in the range of 30–200K (Drozdov et al. 2015). Results for sulfur hydrides (Li et al. 2014), calcium polyhydrides (Wang et al. 2012c), and yttrium polyhydrides (Peng et al. 2017; Liu et al. 2017b) will be introduced here in detail.

H_2S is a typical molecular solid at ambient pressure. At high pressure, it transforms into three high-pressure phases, whose crystal structures are the subject of intense debate. H_2S is not considered a superconductor as it has been proposed to dissociate into its constituent elements before metallization (Rousseau et al. 2000). Li et al. performed extensive structure searches on solid H_2S at pressures of

10–200 GPa (Li et al. 2014). In addition to the identification of candidate structures for the nonmetallic phases IV and V, two metallic structures with P-1 (Fig. 7b) and $Cmca$ symmetry were predicted to be stable above 80 GPa, thus contradicting with the traditional belief (Rousseau et al. 2000) on elemental dissociation at high pressure. The subsequent EPC calculations revealed T_c values of 80 K for the P-1 structure at 158 GPa and 82 K for the $Cmca$ structure at 160 GPa. Motivated by this prediction, electrical measurements of compressed H_2S observed the appearance of superconductivity with a T_c in a range of 30–150 K in a sample prepared at 100–150 K, in a good accordance with our prediction in the low T_c regime of 30–80 K; a higher T_c of ~200 K was observed in another sample prepared at above 220 K (Drozdov et al. 2015; Duan et al. 2015), which is later on suggested as H_3S through the decomposition of H_2S (see, e.g., (Li et al. 2016; Duan et al. 2015; Einaga et al. 2016; Errea et al. 2015)).

Extensive structure searching on calcium polyhydrides (CaH_x) at high pressure (Wang et al. 2012c) showed the CaH_6 stoichiometry (Fig. 7c) to be stable at above 150 GPa: it had a body-centered cubic structure, with hydrogen forming unusual "sodalite" cages containing enclathrated Ca at the center of the cage. Its stability is derived from the acceptance by two H_2 of electrons donated by Ca to form an "H_4" unit as the building block for the construction of the three-dimensional sodalite cage. This unique structure has partially occupied degenerate orbitals at the zone center. The resultant dynamic Jahn–Teller effect helps to enhance the EPC, leading to the superconductivity of CaH_6. A T_c of 220–235 K at 150 GPa obtained by solving the Eliashberg equation is the highest among all hydrides studied thus far. Similar sodalite-like structures of MH_6 (M = Mg and Y) compounds (Li et al. 2015b; Feng et al. 2015) are also high-T_c superconductors at high pressure, with T_c reaching as high as 264 K. These findings suggest that metal polyhydrides with clathrate structures are potential high-T_c superconductors.

Two recent works conducted structure searches using CALYPSO for stable H-rich clathrate structures in rare earth hydrides at high pressures (Peng et al. 2017; Liu et al. 2017b). Hydrogen clathrate cage structures with stoichiometries of H_{24}, H_{29}, and H_{32}, in which H atoms are weakly covalently bonded to one another, forming cages around rare earth atoms are predicted to be thermodynamically stable at high pressures, some of which exhibit high superconductivity. It is found that high-T_c superconductivity is closely associated with H clathrate structures, with large H-derived electronic densities of states at the Fermi level and strong electron–phonon coupling related to the stretching and rocking motions of H atoms within the cages. In particular, an yttrium H_{32} clathrate structure of stoichiometry YH_{10} (Fig. 7d) is predicted to be a potential room-temperature superconductor with an estimated T_c of up to 303 K at 400 GPa (Peng et al. 2017). In the same structure of LaH_{10}, the estimated T_c reaches 288 K at 200 GPa (Peng et al. 2017; Liu et al. 2017b). But it is unstable with respect to LaH_3 and LaH_{11} at 200 GPa. There might be a need of very high temperature to synthesize it (Geballe et al. 2018). Encouragingly, two recent independent experiments seem support the predicted superconductivity of LaH10 at high pressures, where the measured T_c can reach

as high as 260-280 K at 190 GPa (Maddury et al. 2019) and 215 K at 150 GPa (Drozdov et al. 2019)

3.1.2   Non-hydrogen Compounds

Despite a wealth of high-T_c hydrogen-rich hydrides theoretically predicted at high pressure, experimental exploration of these compounds remains a great challenge. The potentially broader ranges of stability make non-hydrogen compounds attractive. Combining CALYPSO with first-principles EPC calculations has predicted many unexpected superconducting compounds of light elements.

Tin telluride is an established high-pressure superconductor, but understanding the underlying mechanism of its transition from semiconductor has been impeded by unsettled issues concerning its structural identification and phase boundary at high pressure. Zhou et al. investigated the high-pressure phase transitions of SnTe using angle-dispersive synchrotron X-ray diffraction combined with the CALYPSO method (Zhou et al. 2013). Three coexisting intermediate phases of *Pnma*, *Cmcm*, and GeS-type structures were identified. The *Pnma* and *Cmcm* phases were predicted to be superconducting by first-principles calculations, with T_c values of 0.70–0.37 K and 0.01–0.03 K, respectively.

As a sister of CO_2, CS_2 is a transparent liquid under ambient conditions and found to transform a molecular solid with a *Cmca* structure at 1 GPa. Upon compression, it became a superconductor with a T_c of 6 K which remains almost constant from 60 to 170 GPa (Dias et al. 2013). However, the lack of an accurate structural determination impedes further understanding the origin of this superconductivity. The later theoretical work using the particle swarm optimization and genetic algorithm methods investigated the crystal structures of CS_2 at high pressures (Zarifi et al. 2015). The predicted structure with $P2_1/m$ symmetry was found to be the most stable from 60 to 120 GPa. The calculated pair distribution functions are in agreement with the experimental assignment. By calculating the electron–phonon coupling parameters, the estimated T_c at 60 GPa is 13 K that is slightly higher than observed 6 K in experiment. The computed Eliashberg spectral function reveals that the S-S vibrations play a critical role in determining the superconductivity of CS_2.

The binary semiconductor boron phosphide (BP) with the zinc blende crystal structure has attracted tremendous attentions due to novel properties such as high hardness, high temperature stability, and high thermoelectric powers for direct energy conversion. In order to understand the pressure-induced structural behavior of BP, the CALYPSO method was employed to determine the crystal structures of boron phosphide at high pressure. A novel *C2/m* structure, which possesses zigzag phosphorus chain structure, was uncovered at 113 GPa, followed by another $P4_2/mnm$ structure above 208 GPa (Zhang et al. 2015b). Theoretical calculations indicate that the *C2/m* phase was superconductor with T_c (9.4–11.5 K). This work reveals that pressure-induced zigzag phosphorus chain in BP exhibits high superconducting temperature, opening a new route to design superconductors with zigzag phosphorus chains.

3.2   Superhard Materials

To be superhard, a material should have a Vickers hardness exceeding 40 GPa. Superhard materials are widely used industrially in cutting, polishing, abrasives, protective coatings, etc. The origin of their excellent mechanical properties is strong directional covalent bonds with high electron density that resist both elastic and plastic deformation. Diamond has the highest Vickers hardness of 60–120 GPa but is incompatible with ferrous metals and high temperatures due to its low thermal and chemical stability (Haines et al. 2001). The theoretical design of new superhard materials has become a major research focus and is greatly desirable to assist experimental investigations (Kaner et al. 2005; Zhao et al. 2016). Over the past several decades, covalent compounds formed by light elements, namely, boron (B), carbon (C), nitrogen (N), and oxygen (O), have been the preferred targets owing to their ability to form strong and densely packed 3D covalent bonding networks (Kaner et al. 2005). A new family of materials formed by heavy transition metals and light elements has recently been proposed as potential superhard materials, given that heavy transition metals can introduce high valence electron density into the compounds to resist both elastic and plastic deformations. Our CALYPSO method has helped to design new superhard materials comprising light elements or heavy transition metals and light elements (Fig. 8a, Zhang et al. 2017a).
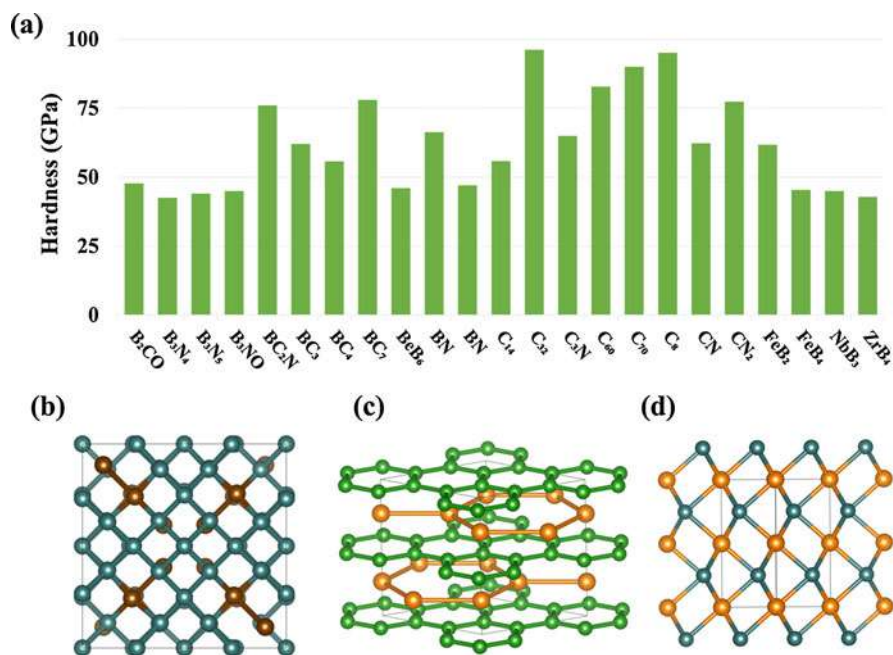
Fig. 8 (**a**) Histogram of calculated hardness of some compounds predicted by CALYPSO. Crystal structures of (**b**) BC_3, (**c**) WB_3, and (**d**) WN

3.2.1 Light-Element Compounds

Boron-doped diamond has been expected to show better oxidation and ferrous resistance than diamond, thus expanding its applicability in electronic devices (Zinin et al. 2012; Dubrovinskaia et al. 2007; Liu et al. 2015; Solozhenko et al. 2004). However, it remains a huge challenge to introduce a high amount of B into the cubic diamond cell. Encouragingly, diamond-like BC_3 with one-quarter B content has been synthesized at high pressure and high temperature. Several structures have been proposed for this novel diamond-like BC_3; however, none of the proposed sp^3 structural models satisfy the experimentally observed cubic symmetry, thus severely impeding further exploration of its properties. The CALYPSO method has solved the crystal structure of recently synthesized cubic BC_3 (Zhang et al. 2015c). In contrast to previously proposed tetragonal and orthorhombic structures, our predicted highly symmetric BC_3 phase for the cubic diamond structure has a space group of I-43m symmetry and 64 atoms per unit cell, denoted as d-BC_3 (Fig. 8b). It becomes stable above 41.3 GPa, in good agreement with the reported synthesis pressure of 39 GPa. Simulated X-ray diffraction and Raman spectra of the predicted d-BC_3 phase agree well with the experimental data (Zinin et al. 2012). The calculated hardness (62 GPa) and ideal strength (52.5 GPa) demonstrate d-BC_3 as an intrinsic superhard material with intriguing bond elongation and sequential bond-breaking processes that lead to remarkable extended ductility and elastic response. These results represent a significant advance in the understanding of a distinct type of superhard material that exhibits superior ductility to diamond or c-BN.

The prediction of several superhard materials with C_3N_4 stoichiometry has motivated intense experimental interest to synthesize and characterize them and also intense debate. To design new stoichiometric carbon nitrides with four-coordinated carbon and three-coordinated nitrogen atoms, a body-centered tetragonal CN_2 with I-42d space group is predicted at high pressure using CALYPSO (Li et al. 2012). The current predicted structure is built up by strong covalent C–N bonds and N–N bonds, denoted bct-CN_2, which resist decomposition into a mixture of diamond + N_2 or $1/3(C_3N_4 + N_2)$ above 45.4 GPa. The strong covalent C–N bonds, N–N bonds, and non-bonding lone-pair are together the driving force for its high bulk (407 GPa) and shear (386 GPa) modulus and simulated hardness (77 GPa) at equilibrium. Our results show that the lone-pair non-bonding states are much more flexible and mobile than the covalent bonds under large strain, and thus bct-CN_2 possesses lower ideal strength (47 GPa).

We used CALYPSO method to design two potential superhard ternary materials, B_2CO and B_3NO, that are isoelectronic with diamond (Li et al. 2015c). For B_2CO, two energetically competitive diamond-like structures with space groups of P-4m2 (tP4) and I-42d (tI16) with all sp^3 bonding states are predicted. Our simulated results show that they exhibit excellent and similar mechanical properties: e.g., high bulk modulus (~31 GPa), high shear modulus (~260 GPa), high hardness (~50 GPa), and low Poisson's ratio (~0.17). For B_3NO, we further performed a superhard-driven search, finding a variety of new structures with short and strong 3D covalent bonds. Among them, two orthorhombic structures with space groups of Imm2 (oI20)

and *Pmn*2$_1$ (oP20) were found to be dynamically stable. They are both superhard materials with simulated Vickers hardnesses above 45 GPa, exceeding the criterion for superhardness. Electronic results show that the oI20 and oP20 structures are semiconductors with optimal bandgaps of 0.87 and 0.12 eV, respectively, and thus have broad prospects for industrial application.

3.2.2   Transition-Metal–Light-Element Compounds

Tungsten borides are transition-metal–light-element compounds that exhibit diverse polymorphism, excellent functional properties, and potential industrial applicability; for example, they have strong mechanical properties, inexpensive components, and feasible synthesis conditions achievable in a cubic anvil apparatus at relatively low pressures. The conventional expectation is their high boron concentration forms a strong covalent bonding network to enhance their hardness/strength. However, the latest theoretical studies show the structural determination and even the chemical composition of these synthesized tungsten borides still to be open questions. Using CALYPSO, we identified the thermodynamically stable structures as well as many metastable structures over a wide range of boron concentrations for tungsten borides (Li et al. 2013b). Comparison of experimental and simulated X-ray diffraction patterns leads to the identification of previously synthesized *I*4/*m*-4u W$_2$B (γ-phase), *I*4$_1$/*amd*-8u WB (α-phase), *P*6$_3$/*mmc*-4u WB$_2$ (ε-phase), and *P*6$_3$/*mmc*-4u WB$_3$ (Fig. 8c). Based on the calculated convex hull, *P*6$_3$/*mmc*-2u WB$_2$, *R*-3*m*-6u WB$_3$, and *P*6$_3$/*mmc*-2u WB$_4$ are thermodynamically stable and thus viable for experimental synthesis. Our first-principles calculations reveal that, contrary to common expectations, increasing the boron content does not raise the mechanical strength of boron-rich tungsten borides; instead, ideal strength show little improvement or even decreases with rising boron content in most cases. This intriguing behavior is fundamentally rooted in boron's ability to form versatile bonding states, which produces distinct structural configurations (Li et al. 2015a). This work represents a comprehensive study of boron-rich tungsten borides based on a global structure search, and the results have important implications for a large class of transition-metal borides that share the same boron concentration range, along with likely similar bonding configurations and deformation mechanisms as revealed in the present study.

Transition-metal nitrides are another class of materials that can be tailored into a new generation of superhard solids. CALYPSO structural searching yields many crystal structures for tungsten nitrides (Lu et al. 2017). X-ray diffraction indicates the presence of hP4-WN in a recently synthesized specimen. Here we report findings from first-principles calculations for two tungsten nitrides, hP4-WN (Fig. 8d) and hP6-WN$_2$, which exhibit extraordinary strain stiffening that remarkably enhances their indentation strengths above 40 GPa, raising exciting prospects for nontraditional superhard solids. Calculations show that hP4-WN is metallic both at equilibrium and under indentation, making it the first known intrinsic superhard metal.

4   Conclusions and Prospects

This chapter briefly introduced the basic theory and general features of our CALYPSO approach. Its validity has been extensively demonstrated via successful applications to various material systems, including 3D bulk crystals, 2D layers and surfaces, and 0D clusters. Its successful application to the discovery of superconductive and superhard materials was presented, demonstrating its great promise for designing functional materials.

Design-orientated experiments guided by structure prediction offered, for example, by CALYPSO have greatly expedited materials discoveries. However, they are sometimes hampered by the high computational costs needed for structural optimizations, especially for multicomponent or large systems. Machine learning as a data-driven method for making prediction, decision, or classification is now starting to enter the heart of computational physics, chemistry, and material sciences in a manifold way. Several machine learning potentials have been developed, which show accuracy comparable to those of quantum mechanical simulations but require less computational effort by many orders of magnitude (Behler and Parrinello 2007; Behler 2016). The replacement of the heavy computational cost of DFT calculations with a state-of-art machine learning potential is a viable route for structure prediction in large systems (Tong et al. 2018; Deringer et al. 2018).

The structural prediction at finite temperature remains challenging due to the requirement of accurate calculation of free energy. The harmonic/quasi-harmonic lattice dynamics models (Van de Walle and Ceder 2002) can be used to estimate the free energy. However, numerous trial structures are required to estimate free energy for structure prediction. It is computationally unaffordable to calculate phonons of all structures. In principle, metadynamics or molecular dynamics simulations (Martoňák et al. 2003; Parrinello and Rahman 1981) can be widely used to predict structures at finite temperatures. However, in reality, metadynamics or molecular dynamics simulations may suffer frequently from the constraint of short simulation run, especially when time-consuming first-principles simulations are performed. Short simulation runs naturally lead to inadequate sampling of free-energy landscape, limiting the wide use of these methods on prediction of high-temperature structures. Future development of cheaper and more efficient methods in calculations of free energy is essential to accomplish the goal of structure prediction at high temperature.

Knowledge of stable structures does not automatically provide detailed synthesis pathways. It is highly desirable to know whether the identified new materials can be eventually synthesized experimentally, but the theoretical prediction of synthesis conditions leading to a desired structure remains challenging. This requires us to compute the stability of a material with respect to alternative atomic configurations. For doing so, it is worth developing relevant methods to address the minimum-energy path of phase transition or chemical reaction after new materials are predicted.

References

Abraham NL, Probert MIJ (2006) A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. Phys Rev B 73:224104

Agrawal A, Choudhary A (2016) Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science. APL Mater 4:053208

Amsler M, Goedecker S (2010) Crystal structure prediction using the minima hopping method. J Chem Phys 133:224104

Ashcroft NW (1968) Metallic hydrogen: a high-temperature superconductor? Phys Rev Lett 21:1748–1749

Ashcroft NW (2004) Hydrogen dominant metallic alloys: high temperature superconductors? Phys Rev Lett 92:187002

Bai L et al (2015) Pressure-induced cation-cation bonding in V_2O_3. Phys Rev B 92:134106

Bardeen J, Cooper LN, Schrieffer JR (1957) Theory of superconductivity. Phys Rev 108:1175–1204

Behler J (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145:170901

Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 98:146401

Call ST, Zubarev DY, Boldyrev AI (2007) Global minimum structure searches via particle swarm optimization. J Comput Chem 28:73–86

Ceder G (2010) Opportunities and challenges for first-principles materials design and applications to Li battery materials. MRS Bull 35:693–701

Ceder G et al (1998) Identification of cathode materials for lithium batteries guided by first-principles calculations. Nature 392:694–696

Chen Y et al (2013) High-pressure phase transitions and structures of topological insulator BiTeI. J Phys Chem C 117:25677–25683

Chen Q et al (2017) Planar B_{38-} and B_{37-} clusters with a double-hexagonal vacancy: molecular motifs for borophenes. Nanoscale 9:4550–4557

Ciobanu CV, Wang C-Z, Ho K-M (2013) Atomic structure prediction of nanostructures, clusters and surfaces. Wiley-VCH Verlag GmbH & Co. KGaA

Dalladay-Simpson P, Howie RT, Gregoryanz E (2016) Evidence for a new phase of dense hydrogen above 325 gigapascals. Nature 529:63–67

Deaven DM, Ho K-M (1995) Molecular geometry optimization with a genetic algorithm. Phys Rev Lett 75:288–291

Debessai M et al (2008) Superconductivity under high pressure in the binary compound $CaLi_2$. Phys Rev B 78:1–10

Deringer VL, Pickard CJ, Csányi G (2018) Data-driven learning of total and local energies in elemental boron. Phys Rev Lett 120:156001

Dias RP, Silvera IF (2017) Observation of the Wigner-Huntington transition to metallic hydrogen. Science 355:715–718

Dias RP et al (2013) Superconductivity in highly disordered dense carbon disulfide. Proc Natl Acad Sci U S A 110:11720–11724

Doye JPK, Wales DJ (1998) Thermodynamics of global optimization. Phys Rev Lett 80:1357–1360

Drozdov AP, Eremets MI, Troyan IA, Ksenofontov V, Shylin SI (2015) Conventional super-conductivity at 203 kelvin at high pressures in the sulfur hydride system. Nature 525: 73–76

Drozdov AP, Minkov VS, Besedin SP, Kong PP, Kuzovnikov MA, Knyazev DA, Eremets MI (2019) Superconductivity at 215 K in Lanthanum hydride at high pressures, Nature 569(7757):528–531

Duan D et al (2015) Pressure-induced metallization of dense $(H_2S)_2H_2$ with high-T_c superconductivity. Sci Rep 4:6968

Dubrovinskaia N et al (2007) Superhard nanocomposite of dense polymorphs of boron nitride: noncarbon material has reached diamond hardness. Appl Phys Lett 90:38–41

Einaga M et al (2016) Crystal structure of the superconducting phase of sulfur hydride. Nat Phys 12:835–838

Eliashberg GM (1960) Interactions between electrons and lattice vibrations in a superconductor. Soviet Phys Jetp 11:696–702

Errea I et al (2015) High-pressure hydrogen sulfide from first principles: a strongly anharmonic phonon-mediated superconductor. Phys Rev Lett 114:157004

Feng X, Zhang J, Gao G, Liu H, Wang H (2015) Compressed sodalite-like MgH_6 as a potential high-temperature superconductor. RSC Adv 5:59292–59296

Franceschetti A, Zunger A (1999) The inverse band-structure problem of finding an atomic configuration with given electronic properties. Nature 402:60–63

Gale JD (1997) GULP: a computer program for the symmetry-adapted simulation of solids. J Chem Soc Faraday Trans 93:629–637

Gao B, Shao X, Lv J, Wang Y, Ma Y (2015) Structure prediction of atoms adsorbed on two-dimensional layer materials: method and applications. J Phys Chem C 119:20111–20118

Gao P, Tong Q, Lv J, Wang Y, Ma Y (2017a) X-ray diffraction data-assisted structure searches. Comput Phys Commun 213:40–45

Gao B et al (2017b) Novel structures of oxygen adsorbed on a Zr(0001) surface predicted from first principles. Appl Surf Sci 393:422–427

Gavezzotti A (1994) Are crystal structures predictable? Acc Chem Res 27:309–314

Geballe ZM et al (2018) Synthesis and stability of lanthanum superhydrides. Angew Chem Int Ed 57:688–692

Giannozzi P et al (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys-Condens Mat 21:395502

Goedecker S (2004) Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. J Chem Phys 120:9911–9917

Gu T, Luo W, Xiang H (2017) Prediction of two-dimensional materials by the global optimization approach. Wires Comput Mol Sci 7:e1295

Guillaume CL et al (2011) Cold melting and solid structures of dense lithium. Nat Phys 7:211–214

Haines J, Leger JM, Bocquillon G (2001) M Aterials. Recherche 1955:1–23

Inoshita T, Jeong S, Hamada N, Hosono H (2014) Exploration for two-dimensional electrides via database screening and ab initio calculation. Phys Rev X 4:031023

Jain A et al (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:011002

Jain A, Shin Y, Persson KA (2016) Computational predictions of energy materials using density functional theory. Nat Rev Mater 1:15004

Kaner RB, Gilman JJ, Tolbert SH (2005) Materials Science. Designing superhard materials. Science 308:1268–1269

Kennedy J (2011) Particle Swarm Optimization. In: Sammut C, Webb GI (eds) Encyclopedia of machine learning. Springer, Boston, pp 760–766

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Kolmogorov AN et al (2010) New superconducting and semiconducting Fe-B compounds predicted with an ab initio evolutionary search. Phys Rev Lett 105: 217003–217006

Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci U S A 99:12562

Lee K, Kim SW, Toda Y, Matsuishi S, Hosono H (2013) Dicalcium nitride as a two-dimensional electride with an anionic electron layer. Nature 494:336–340

Li Q et al (2012) A novel low compressible and superhard carbon nitride: body-centered tetragonal CN_2. Phys Chem Chem Phys 14:13081

Li Q, Zhou D, Zheng W, Ma Y, Chen C (2013b) Global structural optimization of tungsten borides. Phys Rev Lett 110:136403

Li Q, Zhou D, Zheng W, Ma Y, Chen C (2015a) Anomalous stress response of ultrahard WB_n compounds. Phys Rev Lett 115:185502

Li Q, Wang J, Zhang M, Li Q, Ma Y (2015c) Superhard-driven search of the covalent network in the B_3NO system. RSC Adv 5:35882–35887

Li SF, Zhao XJ, Xu XS, Gao YF, Zhang Z (2013a) Stacking principle and magic sizes of transition metal nanoclusters based on generalized Wulff construction. Phys Rev Lett 111:115501

Li Y, Hao J, Liu H, Li Y, Ma Y (2014) The metallization and superconductivity of dense hydrogen sulfide. J Chem Phys 140:174712

Li Y et al (2015b) Pressure-stabilized superconductive yttrium hydrides. Sci Rep 5:9948

Li Y et al (2016) Dissociation products and structures of solid H_2S at strong compression. Phys Rev B 93:2–6

Liu G et al (2015) Submicron cubic boron nitride as hard as diamond. Appl Phys Lett 106:121901

Liu H, Naumov II, Hoffmann R, Ashcroft NW, Hemley RJ (2017b) Potential high-T_c superconducting lanthanum and yttrium hydrides at high pressure. Proc Natl Acad Sci U S A 114:6990–6995

Liu X-D, Dalladay-Simpson P, Howie RT, Li B, Gregoryanz E (2017a) Comment on "Observation of the Wigner-Huntington transition to metallic hydrogen". Science 357:eaan2286

Lu S, Wang Y, Liu H, Miao M, Ma Y (2014) Self-assembled ultrathin nanotubes on diamond (100) surface. Nat Commun 5:3666

Lu S-J, Cao G-J, Xu X-L, Xu H-G, Zheng W-J (2016) The structural and electronic properties of $NbSi_n^{-/0}$ (n = 3–12) clusters: anion photoelectron spectroscopy and ab initio calculations. Nanoscale 8:19769–19778

Lu C, Li Q, Ma Y, Chen C (2017) Extraordinary indentation strain stiffening produces superhard tungsten nitrides. Phys Rev Lett 119:115503

Luo X et al (2011) Predicting two-dimensional boron-carbon compounds by the global optimization method. J Am Chem Soc 133:16285–16290

Luo W, Ma Y, Gong X, Xiang H (2014) Prediction of silicon-based layered structures for optoelectronic applications. J Am Chem Soc 136:15992–15997

Lv J, Wang Y, Zhu L, Ma Y (2011) Predicted novel high-pressure phases of lithium. Phys Rev Lett 106:015503

Lv J, Wang Y, Zhu L, Ma Y (2012) Particle-swarm structure prediction on clusters. J Chem Phys 137:084104

Lv J, Wang Y, Zhu L, Ma Y (2014) B_{38}: an all-boron fullerene analogue. Nanoscale 6: 11692–11696

Lv J et al (2015) Stabilization of fullerene-like boron cages by transition metal encapsulation. Nanoscale 7:10482–10489

Ma Y et al (2009) Transparent dense sodium. Nature 458:182–185

Ma Y et al (2012) Determinations of the high-pressure crystal structures of Sb_2Te_3. J Phys-Condens Mat 24:475403

Maddury S, Muhtar A, Ajay K. M, Zachary M. G, Maria B, Yue M, Viktor V. S, Russell J. H, (2019) Evidence for superconductivity above 260 K in lanthanum superhydride at megabar pressures, Phys Rev Lett 122:027001

Martoňák R, Laio A, Parrinello M (2003) Predicting crystal structures: the Parrinello-Rahman method revisited. Phys Rev Lett 90:075503

Meredig B, Wolverton C (2012) A hybrid computational–experimental approach for automated crystal structure solution. Nat Mater 12:123–127

Needs RJ, Pickard CJ (2016) Perspective: role of structure prediction in materials discovery and design. APL Mater 4:053210

Nishio-Hamane D, Zhang M, Yagi T, Ma Y (2012) High-pressure and high-temperature phase transitions in $FeTiO_3$ and a new dense $FeTi3O_7$ structure. Am Mineral 97:568–572

Oganov AR (ed) (2011) Modern methods of crystal structure prediction. Wiley-VCH Verlag GmbH & Co. KGaA

Oganov AR, Glass CW (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. J Chem Phys 124:244704

Oganov AR et al (2009) Ionic high-pressure form of elemental boron. Nature 457:863–867

Oganov AR, Ma Y, Lyakhov AO, Valle M, Gatti C (2010) Evolutionary crystal structure prediction as a method for the discovery of minerals and materials. Rev Mineral Geochem 71:271–298

Ourique CO, Biscaia EC, Pinto JC (2002) The use of particle swarm optimization for dynamical analysis in chemical processes. Comput Chem Eng 26:1783–1793

Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 52:7182–7190

Patel JS (2003) High-density electron anions in a nanoporous single crystal. Science 301:626–630

Peng F et al (2017) Hydrogen clathrate structures in rare earth hydrides at high pressures: possible route to room-temperature superconductivity. Phys Rev Lett 119:107001

Pickard CJ, Needs RJ (2006) High-pressure phases of silane. Phys Rev Lett 97:045504

Pickard CJ, Needs RJ (2011) Ab initio random structure searching. J Phys-Condens Mat 23:053201

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117:1–19

Rousseau R, Boero M, Bernasconi M, Parrinello M, Terakura K (2000) Ab initio simulation of phase transitions and dissociation of H_2S at high pressure. Phys Rev Lett 85:1254–1257

Schön JC, Jansen M (1996) First step towards planning of syntheses in solid-state chemistry: determination of promising structure candidates by global optimization. Angew Chem Int Ed 35:1286–1304

Segall MD et al (2002) First-principles simulation: ideas, illustrations and the CASTEP code. J Phys-Condens Mat 14:2717–2744

Shang C, Liu ZP (2013) Stochastic surface walking method for structure prediction and pathway searching. J Chem Theory Comput 9:1838–1845

Silvera I, Dias R (2017) Response to comment on "observation of the Wigner-Huntington transition to metallic hydrogen.". Science 357:eaan1215

Soler JM et al (2002) The SIESTA method for ab initio order-N materials simulation. J Phys-Condens Mat 14:2745–2779

Solozhenko VL, Dubrovinskaia NA, Dubrovinsky LS (2004) Synthesis of bulk superhard semiconducting B-C material. Appl Phys Lett 85:1508–1510

Steinhardt PJ, Nelson DR, Ronchetti M (1983) Bond-orientational order in liquids and glasses. Phys Rev B 28:784–805

Stillinger FH (1999) Exponential multiplicity of inherent structures. Phys Rev E 59:48–51

Su C et al (2017) Construction of crystal structure prototype database: methods and applications. J Phys-Condens Mat 29:165901

Tada T, Takemoto S, Matsuishi S, Hosono H (2014) High-throughput ab initio screening for two-dimensional electride materials. Inorg Chem 53:10347–10358

Teter DM (1998) Computational alchemy: the search for new superhard materials. MRS Bull 23:22–27

Tong Q, Xue L, Lv J, Wang Y, Ma Y (2018) Accelerating CALYPSO structure prediction by data-driven learning of a potential energy surface. Faraday Discuss 211:31–43. https://doi.org/10.1039/C8FD00055G

Trimarchi G, Zunger A (2007) Global space-group optimization problem: finding the stablest crystal structure without constraints. Phys Rev B 75:104113

Van de Walle A, Ceder G (2002) The effect of lattice vibrations on substitutional alloy thermodynamics. Rev Mod Phys 74:11–45

Vandevondele J et al (2005) Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. Comput Phys Commun 167:103–128

Wales DJ, Doye JPK (1997) Global optimization by Basin-Hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J Phys Chem A 101:5111–5116

Wang H, Tse JS, Tanaka K, Iitaka T, Ma Y (2012c) Superconductive sodalite-like clathrate calcium hydride at high pressures. Proc Natl Acad Sci U S A 109:6463–6466

Wang H et al (2016a) CALYPSO structure prediction method and its wide application. Comput Mater Sci 112:406–415

Wang H, Li X, Gao G, Li Y, Ma Y (2017) Hydrogen-rich superconductors at high pressures. Wires Comput Mol Sci:e1330. https://doi.org/10.1002/wcms.1330

Wang Y, Ma Y (2014) Perspective: crystal structure prediction at high pressures. J Chem Phys 140:040901

Wang Y, Lv J, Zhu L, Ma Y (2010) Crystal structure prediction via particle-swarm optimization. Phys Rev B 82:094116

Wang Y et al (2011) High pressure partially ionic phase of water ice. Nat Commun 2:563

Wang Y, Lv J, Zhu L, Ma Y (2012a) CALYPSO: a method for crystal structure prediction. Comput Phys Commun 183:2063–2070

Wang Y et al (2012b) An effective structure prediction method for layered materials based on 2D particle swarm optimization algorithm. J Chem Phys 137:224108

Wang Y et al (2015) Materials discovery via CALYPSO methodology. J Phys-Condens Mat 27:203203

Wang Y, Li F, Li Y, Chen Z (2016b) Semi-metallic Be_5C_2 monolayer global minimum with quasi-planar pentacoordinate carbons and negative Poisson's ratio. Nat Commun 7:11488

Woodley SM, Catlow R (2008) Crystal structure prediction from first principles. Nat Mater 7:937–946

Xiang HJ, Huang B, Kan E, Wei SH, Gong XG (2013) Towards direct-gap silicon phases by the inverse band structure design approach. Phys Rev Lett 110:13–16

Xu M et al (2017) Anatase (101)-like structural model revealed for metastable rutile $TiO_2(011)$ surface. ACS Appl Mater Interfaces 9:7891–7896

Yang X et al (2017) Novel superhard sp_3 carbon allotrope from cold-compressed C_{70} peapods. Phys Rev Lett 118:27–29

Zarifi N, Liu H, Tse JS (2015) Structures of the metallic and superconducting high pressure phases of solid CS_2. Sci Rep 5:10458

Zhang H, Li Y, Hou J, Du A, Chen Z (2016) Dirac state in the FeB_2 monolayer with graphene-like boron sheet. Nano Lett 16:6124–6129

Zhang L, Wang Y, Lv J, Ma Y (2017a) Materials discovery at high pressures. Nat Rev Mater 2:17005

Zhang M et al (2015c) Superhard BC_3 in cubic diamond structure. Phys Rev Lett 114:1–5

Zhang X, YanchaoWang JL, Zhu C, Li Q, Zhang M, Li Q, Ma Y (2013) First-principles structural design of superhard materials. Phys Chem Chem Phys 15:20894

Zhang X et al (2015b) Pressure-induced zigzag phosphorus chain and superconductivity in boron monophosphide. Sci Rep 5:8761

Zhang Y-Y, Gao W, Chen S, Xiang H, Gong X-G (2015a) Inverse design of materials by multi-objective differential evolution. Comput Mater Sci 98:51–55

Zhang Y, Wang H, Wang Y, Zhang L, Ma Y (2017b) Computer-assisted inverse design of inorganic electrides. Phys Rev X 7:1–11

Zhao Z, Xu B, Tian Y (2016) Recent advances in superhard materials. Annu Rev Mater Res 46:383–406

Zhou D, Li Q, Ma Y, Cui Q, Chen C (2013) Unraveling convoluted structural transitions in SnTe at high pressure. J Phys Chem C 117:5352–5357

Zhou L, Hou ZF, Gao B, Frauenheim T (2016) Doped graphenes as anodes with large capacity for lithium-ion batteries. J Mater Chem A 4:13407–13413

Zhu L et al (2011) Substitutional alloy of Bi and Te at high pressure. Phys Rev Lett 106:145501

Zinin PV et al (2012) Phase transition in BCx system under high-pressure and high-temperature: synthesis of cubic dense BC_3 nanostructured phase. J Appl Phys 111:114905

Adaptive Genetic Algorithm for Structure Prediction and Application to Magnetic Materials

115

Xin Zhao, Shunqing Wu, Manh Cuong Nguyen, Kai-Ming Ho, and Cai-Zhuang Wang

Contents

Abstract

We present an adaptive genetic algorithm which can be used to predict the atomistic structures of crystals, interfaces, and nanoparticles based on the chemical composition of the materials. The method combines the speed of structure exploration by classical potentials with the accuracy of total energy calculations using first-principles density functional theory (DFT). This method increases the efficiency of structure prediction based on DFT calculations by several orders of magnitude and allows considerable increase in size and complexity of systems that can be studied. The performance of the method is demonstrated by successful structure identifications of complex binary and ternary intermetallic compounds

X. Zhao · M. C. Nguyen · K.-M. Ho
Ames Laboratory and Department of Physics, Iowa State University, Ames, IA, USA

C.-Z. Wang
Ames Laboratory – USDOE and Department of Physics and Astronomy, Iowa State University, Ames, IA, USA
e-mail: wangcz@ameslab.gov

S. Wu
Department of Physics, Xiamen University, Xiamen, China

with the number of atoms in the unit cell as large as 150 atoms. Applications of the method to discovery of novel magnetic materials are discussed.

1   Introduction

Crystal structure prediction based solely on chemical composition information of the materials has been one of the long-standing challenges in theoretical solid state physics, chemistry, and materials science (Maddox 1988; Eberhart and Clougherty 2004; Woodley and Catlow 2008). Advances in this area are highly desirable as computational approaches have been playing a more and more important role in materials discovery and design in the digital age. In the past two decades, several computational methods have been proposed to tackle this challenging problem. These methods include simulated annealing (Kirkpatrick et al. 1983; Wille 1986; Doll et al. 2007), genetic algorithm (GA) (Deaven and Ho 1995; Woodley et al. 1999; Harris et al. 1998; Woodley 2004; Oganov and Glass 2008; Lyakhov et al. 2013; Ji et al. 2011; Wu et al. 2011), basin (or minima) hopping (Wales and Doye 1997; Goedecker 2004), particle swarm optimization (Wang et al. 2010, 2012), and ab initio random structure search (Pickard and Needs 2011). While considerable progress has been achieved, exploration of complex systems requires more advanced algorithms that can provide fast and efficient configuration space samplings and at the same time retain the accuracy of energy evaluation.

The most time-consuming step in all methods for structure predictions mentioned above is the local structure relaxation and energy evaluation since many structures need to be evaluated during the structure search. While first-principles density functional theory (DFT) calculations can offer total energies for many (although not all) crystal structures with acceptable accuracy, its computational cost imposes the bottleneck to fast structure exploration of complex materials. Although size of the order of 1000 atoms can now be handled by many DFT codes in modern computers, determining the low-energy crystal structures for one given chemical composition using genetic algorithm (or similar algorithms) generally would require structure relaxations and energy evaluations for 10^3 to 10^5 structures or more. Therefore, DFT calculations to search for the crystal structures with unit cells containing a few hundred atoms and with variable stoichiometry will be very time-consuming and expensive. By contrast, calculations based on classical empirical potentials are fast even for very large systems, but they are limited in accuracy. Moreover, reliable classical empirical potentials are not even available for many systems.

In order to overcome this bottleneck, we have recently developed an adaptive genetic algorithm (AGA) method (Wu et al. 2014; Zhao et al. 2014a) which combines the speed of classical potential searches and the accuracy of first-principles DFT calculations. The AGA method enables us to investigate crystal structures previously intractable within current computer capabilities.

In this chapter, we will describe the concept and practical operations of the AGA for atomistic structure prediction of crystals, interfaces, and nanoparticles, using

the information of the chemical compositions of the materials. Applications to the discovery of novel materials for magnetic applications will also be discussed.

2   Adaptive Genetic Algorithm

The main idea of the AGA is to use much inexpensive empirical interatomic potentials for the labor intensive local structure relaxation during the genetic algorithm (GA) search, while these classical auxiliary potentials are supervised by accurate DFT calculations through the training and correction iterations (Wu et al. 2014; Zhao et al. 2014a).

The flowchart of the AGA scheme is illustrated in Fig. 1. The left-hand side of the flowchart is the traditional GA loop. The GA is an optimization strategy inspired by the Darwinian evolutionary process and has been widely adopted for atomistic structure optimization in the last 20 years (Deaven and Ho 1995; Woodley et al. 1999; Harris et al. 1998; Woodley 2004; Oganov and Glass 2008; Lyakhov et al. 2013; Ji et al. 2011; Wu et al. 2011). During the GA optimization process, inheritance, mutation, selection, and crossover operations are included to produce new structures and select most fit survivors from generation to generation. For structure optimization, the total energy of the system is commonly used as the fitness function for the survival during the GA process, although other properties of the materials can also be included in the fitness function. The most time-consuming step in the traditional GA-loop is the local optimization of the new off-springs by DFT calculations. For complex structures, GA search usually iterates over more than several hundred generations to converge so that many structures need to be evaluated. In the AGA scheme, this most time-consuming step is performed by using auxiliary classical potentials. An *adaptive*-loop is added to the conventional GA loop as shown in the right side of Fig. 1 to guide the adjustment of the auxiliary classical potentials to fit the accurate results of DFT calculations from iteration to iteration. Only single point DFT calculations performed on a small set of candidate structures obtained in the GA-loop by the auxiliary classical potentials are needed at each iteration. Energies, forces, and stresses of these structures from first-principles DFT calculations are used to update the parameters of the auxiliary classical potentials, e.g., by force-matching method with stochastic simulated annealing algorithm as implemented in the *potfit* code (Brommer and Gahler 2006, 2007). Another cycle of GA search is then performed using the newly adjusted potentials, followed by the re-adjustment of the potential parameters, and the AGA iteration process is then repeated. In this way, the auxiliary classical potentials can help in fast sampling of configuration space though GA and expensive DFT calculations are kept at the minimal without losing the accuracy of the structure search.

In practice, the numbers of parent N_p in the GA pool and off-spring structures N_o to be set depend on the complexity of the system investigated. For systems with 10–50 atoms, $60 < N_p$ or $N_o < 200$ would be a reasonable choice. Therefore, the use of classical auxiliary potentials for the structure relaxations reduces the computational load by approximately five to six orders of magnitude. It usually takes 30–50
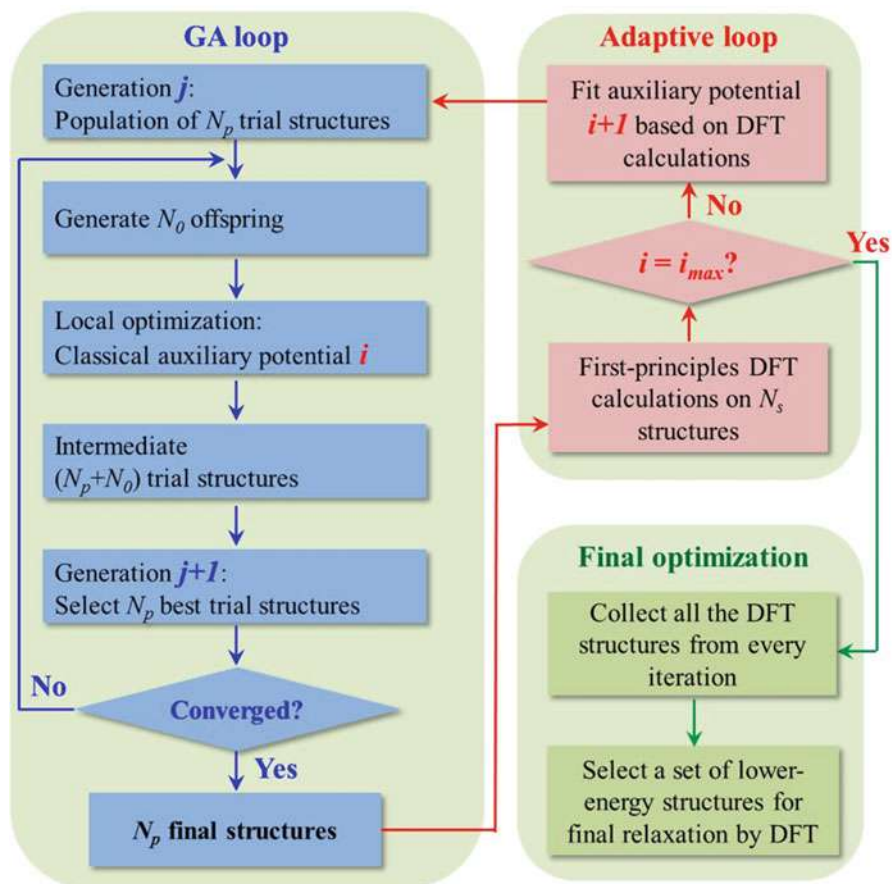
Fig. 1 Flowchart of the adaptive genetic algorithm. An adaptive loop (with i as the iteration counter) is added to the regular GA-loop (with j as the generation counter). Optimization of offspring structures in the GA loop is performed using auxiliary classical potentials whose parameters are adjusted adaptively according to the DFT calculations in the adaptive loop. When the preset maximum number of iterations i_{max} is reached, all structures calculated by DFT from every iteration are collected and ranked according to their energies. Finally, a set of low-energy structures selected from the DFT pool are fully relaxed by DFT calculations to locate the ground-state structure

AGA-iterations to obtain the final structures and the net computational time of the entire AGA search can be reduced by more than three orders of magnitude. Since the classical potentials are adjusted according to DFT results, the AGA can explore the configuration space more effectively and accurately. Structures collected over all AGA iterations can be used to select a subset of low-energy metastable structures which can be finally refined by DFT relaxations. Therefore, the AGA can essentially search for structures almost with the efficiency of classical potentials but

with DFT accuracy. Not only the ground state structures but also a set of low-energy metastable structures can be obtained for a given chemical composition.

In order to show how AGA works in practice and the power of AGA for complex crystal structure prediction, we present two benchmark examples of searching for structures of the Hf_2Co_7 binary alloy with 36 atoms per unit cell and of the $ZrCo_3B_2$ ternary alloy with 54 atoms per unit cell, respectively. Ground-state structures of these alloys have been known experimentally (Buschow et al. 1978; Voroshilov et al. 1971). The AGA searches were performed using only chemical compositions as input information. Initial structures were generated by placing atoms randomly in the unit cell. Classical empirical potentials based on Embedded-Atom Method (EAM) (Daw and Baskes 1984) were used in the GA-loop. As shown in Fig. 2, initially the unit cell shapes, atomic positions, and energies of Hf_2Co_7 and $ZrCo_3B_2$ structures are far from the ground-state. AGA searches then quickly locate the correct ground-state structures for these alloys within 8 and 15 iterations, respectively. These results demonstrate that the AGA is a powerful computational tool for predicting complex crystal structures.

We note that the AGA approach is different from the conventional approach of combining classical potentials with DFT calculations for structure optimization which use a *single* set of classical potentials to screen *all* candidate structures followed by a refinement by DFT calculations. The conventional approach would require accurate and transferable classical potentials to capture the global complex energy landscape of the system. In comparison, the auxiliary classical potentials used in each iteration in AGA only need to be able to represent some local basins of potential energy surface. The system during the AGA search can hop from one basin to another through the adjustment of the auxiliary classical potentials from iteration to iteration through the adaptive loop. From the energies of the final structures at each iteration as plotted in the Fig. 2, we can see that the AGA uses different adjusted potentials to sample structures located in different basins of the energy landscape. Each auxiliary classical potential may not just sample the structures in the same basin, it can sample the structures in a subset of the basins in the energy landscape and some of the basins may overlap with those from other potentials. While it is very difficult or even impossible to fit a classical potential to accurately describe a system under various bonding environments, especially for binary and ternary systems, it is possible to adjust auxiliary potentials to describe structures located within different subset of basins in the energy landscape with DFT accuracy. This is why AGA can work efficiently and accurately in the global low-energy structure search.

Finally, we note that the AGA described above can be applied not only to the structures of periodic crystals, but also to the structures of nanoparticles (or cluster) and interfaces (or surfaces). The only difference is in the operations for the offspring structure generation. In all cases, the cut-and-paste operation proposed by Deven and Ho in 1995 (Deaven and Ho 1995) for GA on clusters seems to be the most simple yet efficient operation for generating offspring structures from the existing pool of parent structures. Such a cut-and-paste operation can be applied straightforwardly to the nanoparticle search in AGA scheme. For periodic
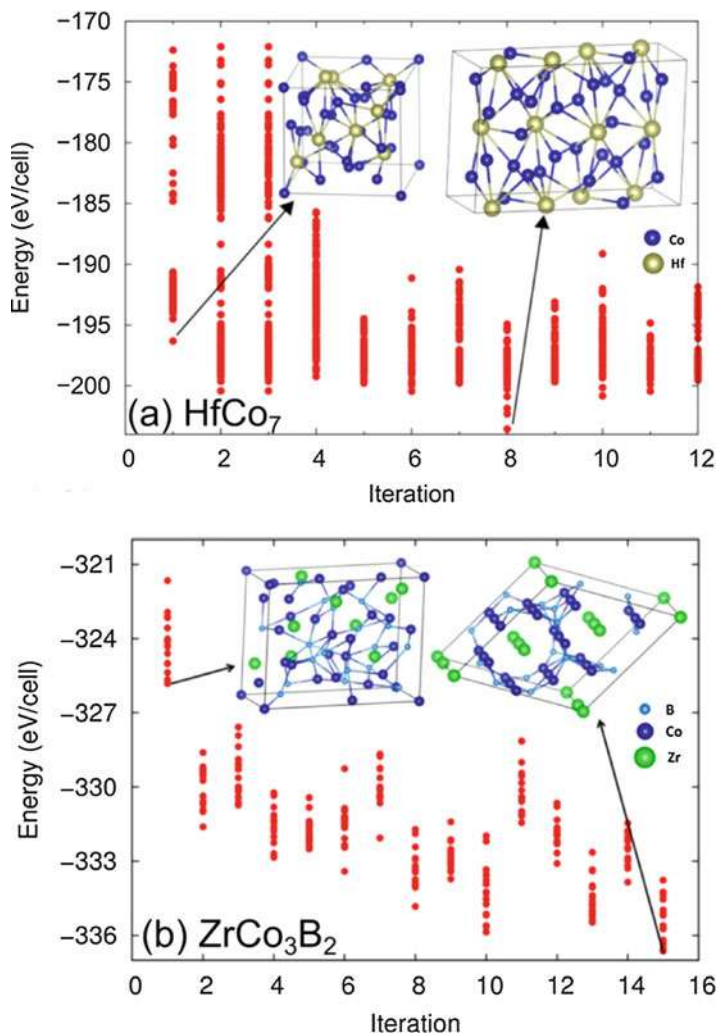
Fig. 2 Structural and energetic evolution versus iteration number of the AGA-loop for (**a**) Hf$_2$Co$_7$ and (**b**) ZrCo$_3$B$_2$ alloys. EAM-type potentials were used in inner GA-loop. Plots show only DFT energies obtained at the end of each AGA-iteration

crystal structures, the atomic positions can be generated by the same cut-and-paste operation, while the shape of the unit cell of the offspring structure can be obtained by averaging the unit-cell matrix elements (which represented the 3 lattice vectors of the unit cell) of the parent structures. For structure prediction of interfaces and surfaces, the unit cell can be divided into several zones as shown in Fig. 3. The cut-and-paste operation can be applied only to those atoms in the interfaces (or surface) zone, while the atoms in the bulk zones can only be rigid shift and relaxed to better match the interfaces (Zhao et al. 2014a).
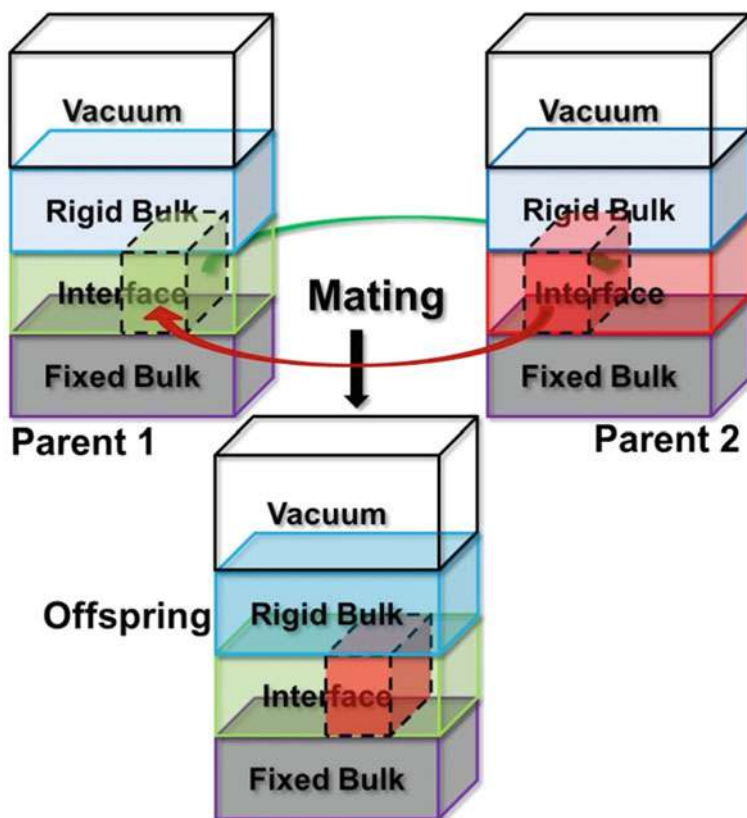
Fig. 3 Schematic representation of the interface model and mating operation in our adaptive genetic algorithm. Four parts are included in the model: fixed bulk where all atoms are fixed; interface where atom positions are to be optimized; rigid bulk which can move as a rigid body relative to the fixed bulk during the search; and a vacuum region to avoid interactions between the two surfaces of the slab. During the mating process, part of the parent structures as indicated by the dashed cuboids will be exchanged to generate the offspring structure

The AGA method has been applied to the prediction and discovery of a range of novel stable and metastable materials (Nguyen et al. 2013, 2018; Zhao et al. 2014b, c, 2015a, 2016a, b, 2017a, b; Dong et al. 2017; Wu et al. 2017). The readers are referred to these references for more details. In the following, we will describe the application of AGA to the prediction and discovery of magnetic materials.

3 Application to Magnetic Materials

As one of the oldest functional materials, permanent magnets are widely used in modern technologies. Permanent magnets are an essential component in electrical

generators using wind, water, geothermal, and carbon-based fuels. Permanent magnets are also essential in computer hard drives and in electric motors for vehicles and other electro-mechanical devices, including levitators. Because of the role of such devices in new energy economies, there has been an increase in demand for stronger permanent magnet materials. Currently, the most widely used permanent-magnet materials are $Nd_2Fe_{14}B$ and $SmCo_5$ discovered in 1980s and 1960s respectively, both containing rare-earth (RE) elements. Owing to limited RE mineral resources and limited RE supplies, there is a strategic need to discover replacement magnet materials for $Nd_2Fe_{14}B$ and $SmCo_5$ to meet performance and cost goals for advanced electric drive motors and generators (Bauer et al. 2011).

It should also be noted that materials synthesis under far-from equilibrium conditions has become an important route for materials design and discovery. Metastable phases formed during the fast cooling process often possess remarkable physical properties. Atomistic structures of some of these metastable phases would be very complex and hard to be resolved by experimental techniques, impeding the understanding and further optimization of the materials for applications. In many cases, metastable structures with desirable properties are hard to be predicted *a priori* by studying the equilibrium phase diagram of the alloys and compounds. With modern computing power, advances in computational algorithms and methods for ab initio crystal structure prediction can speed up the investigation of complex materials and aid the discovery and development of technologically relevant new materials (e.g., magnetic materials).

Using the AGA described above, the crystal structures of several RE-free binary and ternary intermetallic compounds have been systematically explored for possible candidates in magnetic materials applications (Zhao et al. 2014c, 2015a, 2016a, b, 2017a, b). In what follows, we will choose two examples from our recent studies to illustrate the usefulness and the power of the AGA in material design and discovery. The first example is on the determination of the atomistic structures and magnetic properties of complex Zr_2Co_{11} polymorphs with the number of atoms in the unit cell up to 150 atoms (Zhao et al. 2014c). Another example describes the discovery of novel phases of Co-N compound for RE-free magnets by the AGA prediction (Zhao et al. 2016a, b) which leads to the success in experimental synthesis of one of the new phases for potential magnetic materials application (Balasubramanian et al. 2018).

3.1   Unraveling the Structural Mystery of Zr_2Co_{11} Polymorphs

As a potential candidate for RE-free magnetic materials, various phases around the composition of Zr_2Co_{11} have received a lot of interests (Demczyk and Cheng 1991; Gabay et al. 2001; Ivanova et al. 2007; Ivanova and Shchegoleva 2009; Zhang et al. 2013). Studies have shown that some of these metastable phases with composition close to the Zr_2Co_{11} intermetallic compound, resulting from the rapid quenching, exhibit large magnetocrystalline anisotropy energy (MAE) and

with Curie temperatures around 500 °C (Demczyk and Cheng 1991; Gabay et al. 2001). However, the crystal structures of these phases remain unsolved. Mixed phases with small grain sizes in experimental samples make it difficult to determine the atomic positions in the crystal structures of these phases using standard X-ray techniques. Even the exact compositions, the shape and size of the unit cells of the observed phases are under debate. The uncertainty in the crystal structures greatly hinders further development and optimization of the material for practical applications.

In 2007 and 2009, Ivanova et al. (2007; Ivanova and Shchegoleva 2009) were able to obtain samples with larger gains which enable the determination of the crystal Bravais lattice of the phases using X-ray diffraction (XRD). A high temperature phase was identified to be rhombohedral structure with $a_{rhomb} = 4.76$ Å and $c_{rhomb} = 24.2$ Å (in the hexagonal setting). Evidence for a hexagonal, high-temperature phase with $a_{hex} = a_{rhomb}$ and $c_{hex} = 2c_{rhomb}/3$ was also observed from their transmission electron microscopy (TEM) analysis. As the temperature is lowered, a transformation to an orthorhombic structure with $a_{orth} = 4.71$ Å, $b_{orth} = 16.7$ Å, and $c_{orth} = 24.2$ Å takes place. These phases were later confirmed by Zhang et al. (2013). However, these experiments were not able to determine the Wyckoff positions of the proposed crystal structures.

In order to resolve the long-standing structural puzzle of Zr_2Co_{11} polymorphs, we have performed a systematic crystal structure search for $ZrCo_{5+x}$ compounds with x = 0.0, 0.1, 0.2, 0.25, and 0.5 using the AGA (Wu et al. 2014; Zhao et al. 2014c). The structure search was performed without any assumptions on the Bravais lattice type, atom basis, or unit cell dimensions. The auxiliary classical potentials used in the AGA search take the form described by EAM formalism (Daw and Baskes 1984). The potential parameters were adjusted adaptively according to the AGA scheme described in the previous section by fitting to the DFT energies, forces, and stresses of selected structures. For this system, the ab initio calculations were performed using spin-polarized density functional theory within generalized-gradient approximation (GGA) in the form proposed by Perdew-Burke-Ernzerhof (PBE) (Perdew et al. 1996) and with the projector-augmented wave (PAW) method (Blochl 1994; Kresse and Joubert 1999) as implemented in the VASP code (Kresse and Furthmüller 1996a, b). We note that for different systems, exchange-correlation energy functional and other parameter settings need to be properly chosen in order to achieve desirable accuracy. For systems with weak interactions such as layered transition metal dichalcogenides (Ci et al. 2017), van-der-Waals corrections in Chap. 13, "Van der Waals Interactions in Material Modelling" by Hermann and Tkatchenko also need to be included.

Many crystal structures with closely competitive energies were obtained from the AGA search. In particular, three low-energy structures with lattice vectors and simulated X-ray diffraction spectra match well with experimental measurement are obtained. These structures are identified as an orthorhombic structure shown in Fig. 4a with composition of $ZrCo_{5.25}$ and lattice parameters a = 4.68 Å, b = 16.54 Å, c = 24.08 Å, a $ZrCo_5$ rhombohedral structure shown in Fig. 4b with space group R32 and lattice parameters a = 4.69 Å and c = 24.02 Å, and a
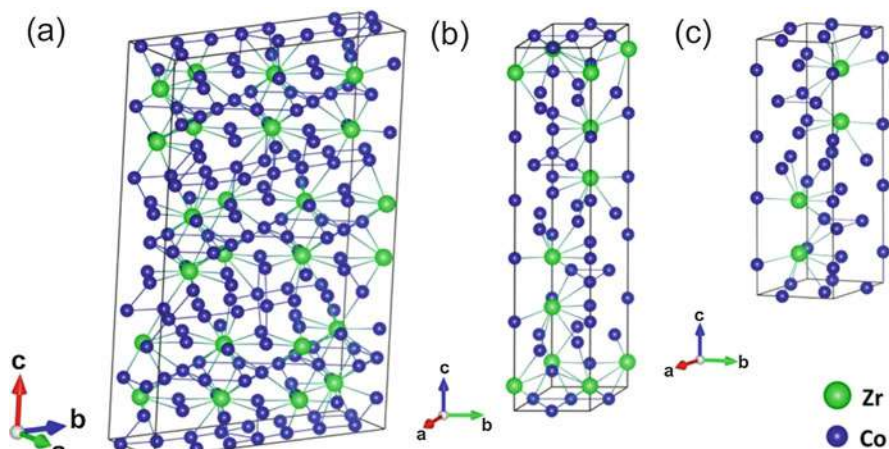
Fig. 4 Three low-energy structures with lattice vectors and simulated X-ray diffraction spectra match well with experimental measurement: (**a**) the orthorhombic structure with a composition of $ZrCo_{5.25}$; (**b**) the rhombohedral structure with the composition of $ZrCo_5$; (**c**) the hexagonal structure with the composition of $ZrCo_5$

$ZrCo_5$ hexagonal structure shown in Fig. 4c with space group P-62c, a = 4.69 Å, and c = 16.1 Å, respectively. The lattice parameters of these structures agree well with those determined by experimental studies of Ivanova et al. (2007; Ivanova and Shchegoleva 2009) and Zhang et al. (2013). It is worth noting that the orthorhombic structure is very complicated with 150 atoms in the unit cell yet it can be determined by the AGA search. The simulated XRD spectrum of orthorhombic structure obtained from the AGA search agrees well with the experimental data as can be seen from Fig. 5a. The simulated TEM images also agree well with experiment, indicating the atomic arrangement in the unit cell predicted from the AGA are accurate. Energy-dispersive X-ray spectroscopy analysis by experiment shows that Co/Zr ratio in the orthorhombic phase is very close to 5.25, agreeing well with our theoretical prediction.

The main features of the simulated XRD patterns of the rhombohedral and hexagonal structures from the AGA prediction also agree with the experimental data. However, due to overlapping reflections from multiple low symmetry phases and broadened peaks from nanoscale grains, the resolution of experimental spectra is not good enough for detail comparison. On the other hand, the structure of the rhombohedral and hexagonal phases can also be revealed by selected-area electron diffraction (SAED) pattern and high-resolution electron microscopy (HREM) image. The simulated SAED patters and HREM images using the AGA predicted structures are in good agreement with experimental ones, indicating the atomistic structures obtained from the AGA search are correct (Zhao et al. 2014c).

The atomistic structures of the above mentioned three phases are also validated by comparing the magnetic properties calculated using the predicted structures
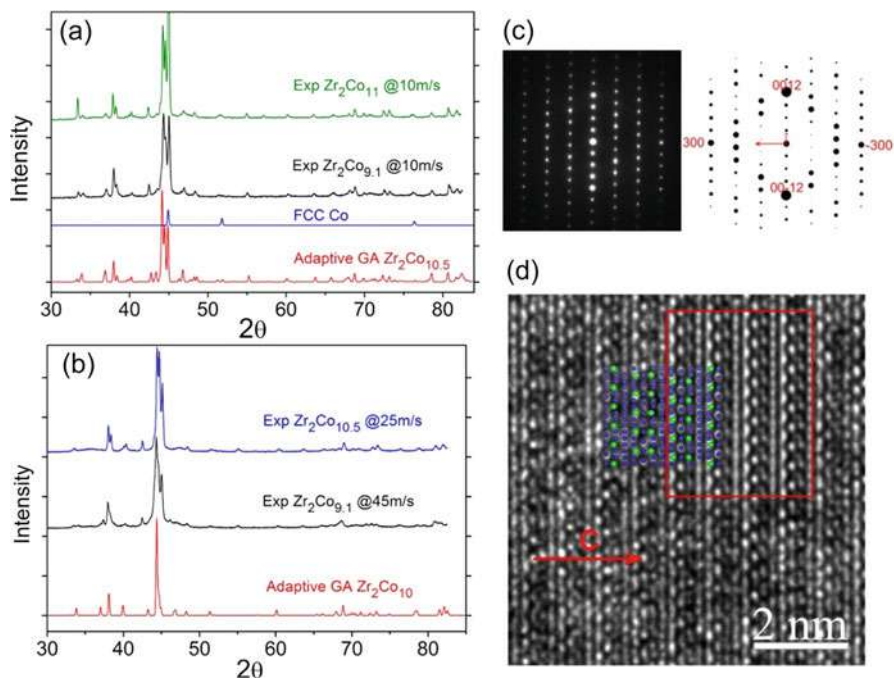
Fig. 5 Comparison between the simulation and the experimental measurement: (**a**) XRD spectrum of the low-temperature orthorhombic phase; (**b**) XRD spectrum of the high-temperature rhombohedral phase; (**c**) selected-area electron diffraction (SAED) pattern; and (**d**) high-resolution electron microscopy (HREM) image with the simulated results shown in the red box

with those measured by experiment. The magnetization in these structures (0.9–1.1 μ_B/atom) is much smaller than that of elemental hcp-Co (1.6 μ_B/atom), in agreement with experimental results. Magnetocrystalline anisotropy energies (MAE) are very different for different structures. The largest anisotropy energy was found in the rhombohedral and hexagonal structures ($\sim$1.3 MJ/m^3), while the MAE is almost zero for the orthorhombic structure. The calculation results are consistent with the experimental measurements and indicate that the high temperature rhombohedral/hexagonal phases correspond to the hard magnetic phase in Zr_2Co_{11} compounds.

This example demonstrates that the AGA can help experiment in resolving complex structures which allows us to elucidate the physical origin of high coercivity observed in this system and provides useful insight guiding further development of these materials for use as high performance permanent magnets without rare earth elements. Resolving complicated atomistic structures with up to 150 atoms per cell by a first-principle computational approach in such a complex multiple-phase system demonstrates a new capability to aid the accelerating of materials discovery through the use of state-of-the-art supercomputers.

3.2   Prediction of Metastable Transition Metal Nitride Structures with High Magnetic Anisotropy

Nitrides of *3d* transition metals (TM) have interesting magnetic properties and have been used in various magnetic devices. For example, it has been shown that Fe atoms in $Fe_{16}N_2$ thin film exhibit magnetic moment larger than that of pure Fe in BCC crystal (Kim and Takahashi 1972; Sugita et al. 1991). Recently, Co_4N thin films prepared using a reactive magnetron sputtering process were also reported to have magnetic moment higher than that of pure Co bulk structure (Gupta et al. 2015). Discovery of new phases in *3d* TM nitrides for magnetic materials applications has attracted considerable interest. In this regard, structure and property prediction based on computational approaches to provide guidance to experimental synthesis will be greatly beneficial to accelerating the materials design and discovery.

Using the AGA (Wu et al. 2014) and first-principles calculations by VASP code (Kresse and Furthmüller 1996a, b), we have systematically investigated the structures and magnetic properties of pure and Fe-doped cobalt nitrides with different compositions (Zhao et al. 2016a, b). Crystal structure search has been performed for Co_nN compounds with n = 1, 2 ... 8, respectively. New lower-energy structures are obtained from the AGA searches and some are shown to have intriguing magnetic properties. In particular, the calculated MAE of the new structures can be as high as 195.4 μeV/Co atom ($\sim$2.45 MJ/m^3). For the purpose of comparison, typical room temperature MAE value is $\sim$0.5 MJ/m^3 for hexagonal Co, $\sim$5 MJ/m^3 for tetragonal $Nd_2Fe_{14}B$ and 17 MJ/m^3 for hexagonal $SmCo_5$ (Skomski and Sellmyer 2009). It is also found that substituting Co with Fe in the Co_nN structures results in larger magnetic moment and at the same time further increases the magnetic anisotropy. The newly discovered structures are promising prototypes to design strong permanent magnets without using RE elements.

In Fig. 6, several representative structures that have large MAE are plotted. For Co_3N, two structures obtained from the AGA search are plotted in Fig. 6a, b, respectively. Co_3N-1 structure (space group *R-3c*) is found to have slightly lower energy ($\sim$1 meV/atom) than the experimental phase (space group *P6$_3$22*) according to GGA calculations. The *R-3c* and the *P6$_3$22* structures share the same Co packing (ABAB ... type of stacking), but the arrangements of N atoms are different, resulting in different unit cell sizes. The orthorhombic Co_3N-2 structure (space group *Cmcm*) in Fig. 6b has an ABCABC ... Co stacking, i.e., Co atoms are in an fcc-like lattice. However, one of the six N-Co bonds (2.05 Å) in the Co_3N-2 structure is noticeably larger than the others (1.85 Å) as indicated in the figure, resulting in large distortions in the structure. The formation energy of the Co_3N-2 structure is higher than that of the ground-state structure by $\sim$45.5 meV/atom.

From the MAE calculations, the *P6$_3$22* structure observed in experiments has nearly zero anisotropy, while the Co_3N-1 (*R-3c*) structure obtained from the AGA search exhibits a uniaxial anisotropy of 1.39 MJ/m^3 with the easy axis along its **c** axis. This MAE value is comparable to what found in the high temperature
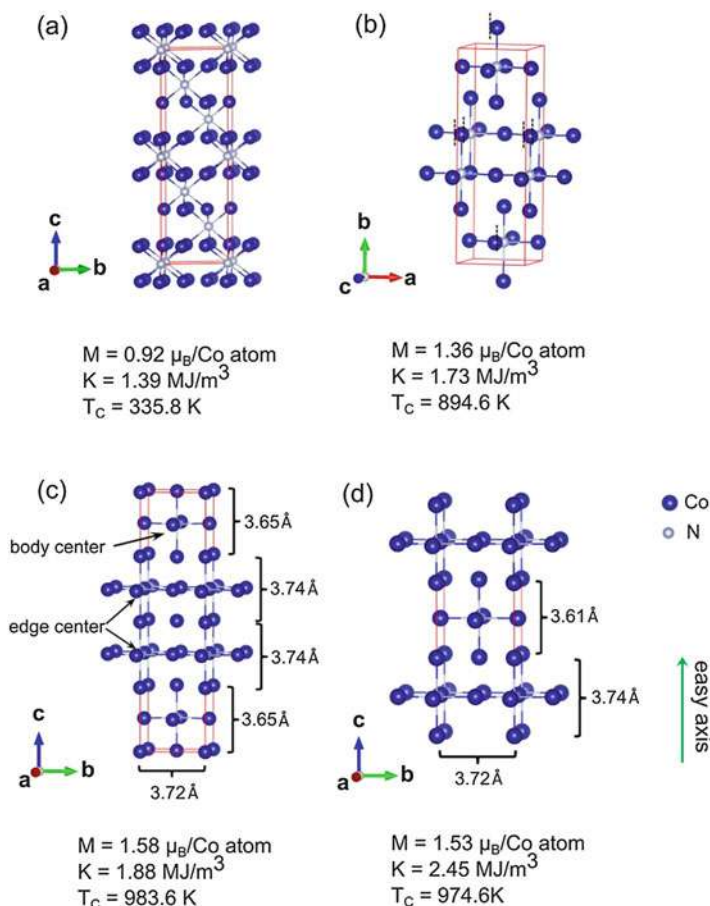
Fig. 6 Structures of Co_nN with large magnetic anisotropy energies: (**a**) Co_3N-1 (*R-3c*); (**b**) Co_3N-2 (*Cmcm*); (**c**) Co_4N-1 (*I4/mmm*); (**d**) Co_4N-2 (*I4/mmm*). Red box indicates the unit cell. The easy axis in the MAE calculations is represented by the green arrow

"Zr_2Co_{11}" phases discussed above (Zhao et al. 2014c). The *Cmcm* structure has a bigger volume, thus much larger magnetic moments than that of Co_3N-1 *R-3c* structure. Calculation also shows a larger MAE value of 1.73 MJ/m^3 with the easy axis along its lattice vector **b** and a large enough Curie temperature ($\sim$895 K) for the *Cmcm* structure. The prediction of large MAE in the Co_3N -1 (*R-3c*) structure has inspired experimental studies and the Co_3N -1 (*R-3c*) structure has been recently synthesized and characterized by experiment (Balasubramanian et al. 2018). The results from the experimental measurement agree well with the theoretical predictions as one can see from the comparison of XRD in Fig. 7.

For Co_4N, experiments have shown that the synthesized *Pm-3m* phase has the same structure as other tetra transition metal nitrides, in which metal atoms are
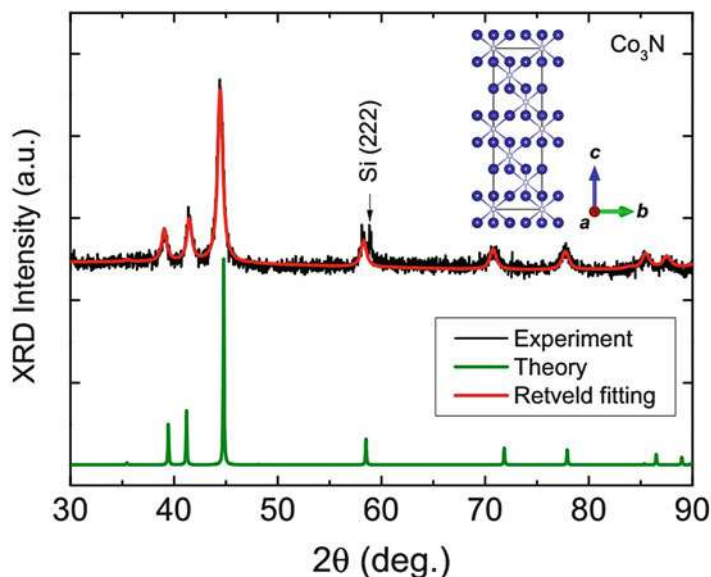
Fig. 7 XRD comparison between the experiment measurements and simulation using our Co_3N structure

arranged in fcc lattice and N atoms occupy the body centered positions. Several structures obtained from the AGA search are found to have lower energies than the experimental *Pm*-3*m* structure. Several low-energy structures are found to have large MAE. Two Co_4N structures that have large magnetic anisotropy energies are plotted in Fig. 6c, d, respectively. Both structures are tetragonal with a space group of *I4/mmm*. The Co_4N-1 structure has a uniaxial anisotropy of 1.88 MJ/m^3, while the MAE for Co_4N-2 is even larger, $\sim$2.45 MJ/m^3. The Curie temperatures for the two Co_4N structures are close to 1000 K, which is higher than that of Co_3N. These two new structures of Co_4N predicted by the AGA would be useful for magnetic materials applications.

Since Fe is much more abundant and cheaper than Co, it would be interesting to see if some of the Co in the Co-N compounds can be replaced by Fe. We have also studied the Fe substitution effects on the stability and magnetic properties of cobalt nitrides based on the structures obtained from the AGA search for cobalt nitrides, since the structures of iron nitrides and cobalt nitrides are very similar due to the similarities between Fe and Co elements. We found that many structures obtained by Fe substitution have negative formation energies meaning the substitution of Co with Fe in cobalt nitrides is energetically favorable. The effect of Fe substitution on the magnetic moment of $Co_{n-x}Fe_xN$ is plotted in Fig. 8 which clearly shows that increasing Fe composition significantly increases the magnetic moment of the system. However, it is also noticed that the largest average values of magnetic moment are not found in Fe_nN, but around $Co_{0.15}Fe_{0.85}N$, i.e., before all the Co
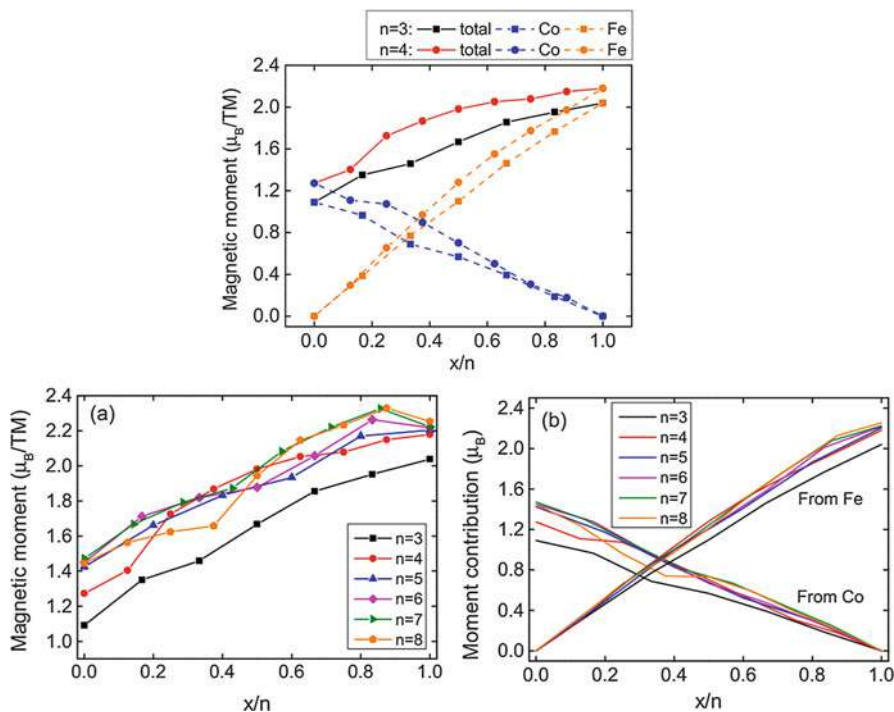
Fig. 8 (**a**) Effect of Fe substitution on the magnetic moments in $Co_{n-x}Fe_xN$. TM represents transition metal, i.e., Fe and Co atoms. (**b**) Magnetic moment contributions from Fe and Co separately, summation of which equals to the total moment of the system plotted in (**a**)

atoms are substituted by Fe. For magnetic anisotropy, many structures with Fe substitutions are found to have MAE values larger than 1 MJ/m³. The largest MAE number is obtained in one of the Co_2Fe_6N structures, which reaches ∼3.18 MJ/m³ (∼245.6 μeV per TM atom) (Zhao et al. 2016b). In Fig. 9, we plotted several structures of $Co_{n-x}Fe_xN$ that have large MAE and at the same time belong to a different type of structures compared with the well-known tetragonal Fe_8N phase. From Fig. 9, we see that the $Co_{n-x}Fe_xN$ structures with relatively large MAEs are all in ABCABC ... type of stacking, i.e., Co/Fe atoms are in fcc lattice. In fact, all the structures in Fig. 9 belong to the same type as those plotted in Fig. 6b–d, and we believe the high MAE comes from the structural anisotropy introduced by N atoms. Similarly, N atom sitting at the edge center makes the length of the fcc units different along **c** axis. The structures in Fig. 9a–d all have MAE larger than 2 MJ/m³. The structure plotted in Fig. 9e is the lowest-energy structure for Co_4Fe_4N. Different from other $Co_{8-x}Fe_xN$ structures adopted from *I4/mmm* Fe_8N (Zhao et al. 2016b), all the nitrogen atoms in the lowest-energy structure of Co_4Fe_4N are located at the edge centers, introducing the structural anisotropy and achieving an MAE value of 1.65 MJ/m³ from first principles calculation. Our calculations also show the
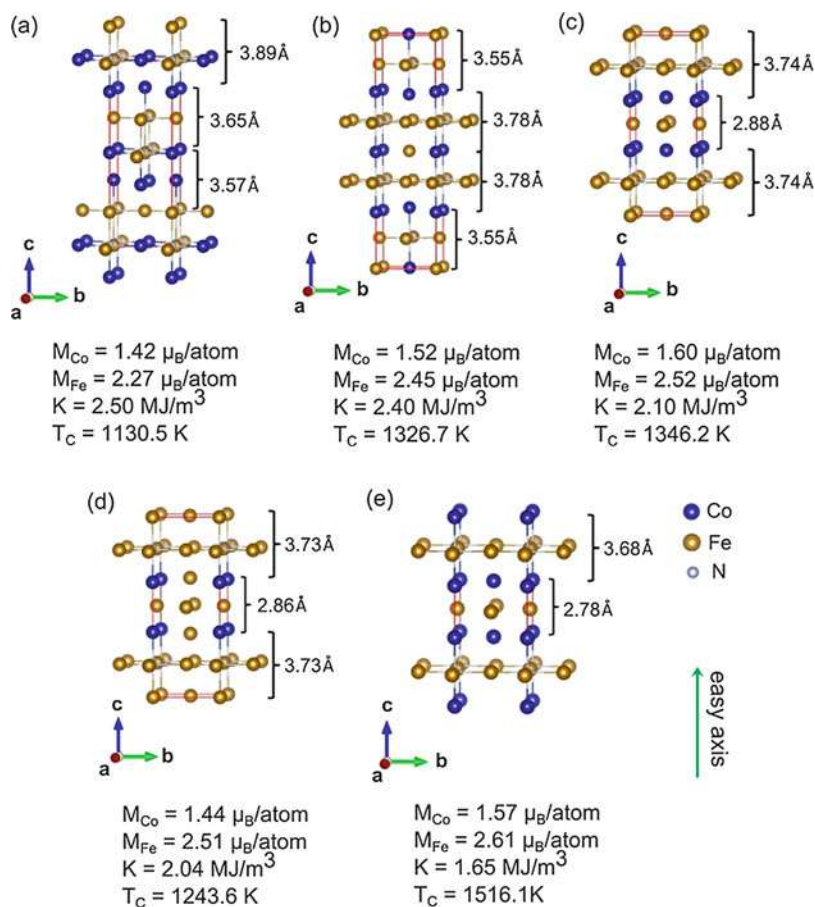
Fig. 9 Structures of $Co_{n-x}Fe_xN$ with high magnetic anisotropy energies: (**a**) $Co_3Fe_3N_2$ (*Amm2*); (**b**) $Co_3Fe_5N_2$ (*I4/mmm*); (**c**) Co_2Fe_4N (*P4/mmm*); (**d**) $CoFe_5N$ (*P4/mmm*); (**e**) Co_4Fe_4N (*P4/mmm*). The easy axes of all above structures are along the direction of their lattice vector **c**

Fe-substituted structures being discussed here have much larger Curie temperatures than binary cobalt nitrides.

Combining the results of Co_nN and Fe-substituted Co_nN, we see that mixing Fe and Co brings more structure anisotropy thus higher magnetic anisotropy energies. Moreover, magnetic moment increases significantly with Fe substitution and in some cases reaches values even higher than that of Fe_nN. Among the newly discovered prototype structures, those structures in which Co/Fe atoms are in fcc stacking and part of the N atoms locate at the edge centers are of particular interest and could be further explored for better magnets, because these structures possess both lower formation energies and higher magnetic anisotropy.

4   Concluding Remarks

In summary, we showed that the AGA is a very efficient and powerful method for predicting the atomistic structures of materials based on minimal information such as chemical compositions. Although we only gave some examples in application to the prediction and discovery the crystal structures of magnetic materials to demonstrate the usefulness of the method, the method can be applied to any other materials including 2D materials, surfaces and interfaces, 1D and 0D material such as nanowires and nanoclusters. The concept of using the adaptively adjusted auxiliary empirical potentials for the labor intensive local structure exploration can also be easily adopted by other structure prediction methods such as the minima hopping and particle swarm optimization methods described in this section.

While the AGA has been demonstrated to be powerful, there are still a lot of rooms for further improvement of the method. Currently, the adaptive interatomic potentials adjustment is done through the force-match fitting within a given mathematical form using the *potfit* code. In general, more reliable and robust adaptive empirical potentials without preassumption of mathematical form can be obtained by deep machine learning (e.g., neural network) process (Behler and Parrinello 2007; Behler 2017; Artrith et al. 2017; Onat et al. 2018; Zhang et al. 2018). Another key step that can greatly improve the efficiency of the GA loop is machine learning -guided offspring structure generation. The currently adopted random cut-and-paste operation for offspring structure generation may not be the most efficient one, particularly for structures with certain motif preference (e.g., tetrahedral or octahedral motif in many oxide materials). Therefore, identifying the preferring motifs by machine learning (Zhao et al. 2015b; Lv et al. 2017) during the GA search and use the knowledge in situ to guide the offspring structure generation will also greatly enhance the efficiency of the AGA scheme. Some more details of the machine learning methods can also be found in the section edited by N. Marzari in the Vol. 1 of this Handbook.

References

Artrith N, Urban A, Ceder G (2017) Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. Phys Rev B 96:014112

Balasubramanian B, Zhao X, Valloppilly SR, Beniwal S, Skomski R, Sarella A, Jin Y, Li X, Xu X, Cao H, Wang H, Enders A, Wang C-Z, Ho K-M, Sellmyer DJ (2018) Magnetism of new metastable magnetic nitride compounds. Nanoscale 10:13011

Bauer D, Diamond D, Li J, Sandalow D, Telleen P, Wanner B (2011) Critical materials strategy. US Department of Energy. http://energy.gov/sites/prod/files/DOE_CMS2011_FINAL_Full.pdf

Behler J (2017) First principles neural network potentials for reactive simulations of large molecular and condensed systems. Angew Chem Int Ed 56:2–15

Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 98:146401

Blochl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953

Brommer P, Gahler F (2006) Effective potentials for quasicrystals from ab-initio data. Philos Mag 86:753

Brommer P, Gahler F (2007) Potfit: effective potentials from ab initio data. Model Simul Mater Sci Eng 15:295

Buschow KHJ, Wernick JH, Chin GY (1978) Note on the Hf-Co phase diagram. J Less-Common Met 59:61–67

Ci PH et al (2017) Quantifying van der Waals interactions in layered transition metal dichalcogenides from pressure-enhanced valence band splitting. Nano Lett 17:4982

Daw MS, Baskes MI (1984) Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B 29:6443

Deaven DM, Ho KM (1995) Molecular geometry optimization with a genetic algorithm. Phys Rev Lett 75:288–291

Demczyk BG, Cheng SF (1991) Structures of Zr_2Co_{11} and $HfCo_7$ intermetallic compounds. J Appl Crystallogr 24:1023

Doll K, Schön JC, Jansen M (2007) Global exploration of the energy landscape of solids on the ab initio level. Phys Chem Chem Phys 9:6128–6133

Dong YH, Lu WC, Xu X, Zhao X, Ho KM, Wang CZ (2017) Theoretical search for possible Au-Si crystal structures using a genetic algorithm. Phys Rev B 95:134109

Eberhart ME, Clougherty DP (2004) Looking for design in materials design. Nat Mater 3:659–661

Gabay AM, Zhang Y, Hadjipanayis GC (2001) Cobalt-rich magnetic phases in Zr–Co alloys. J Magn Magn Mater 236:37

Goedecker S (2004) Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. J Chem Phys 120:9911–9917

Gupta R, Pandey N, Tayal A, Gupta M (2015) Phase formation, thermal stability and magnetic moment of cobalt nitride thin films. AIP Adv 5:097131

Harris KDM, Johnston RL, Kariuki BM (1998) Acta Crystallogr. The genetic algorithm: foundations and applications in structure solution from powder diffraction data. Acta Crystallogr A54:632–645

Ivanova GV, Shchegoleva NN (2009) The microstructure of a magnetically hard Zr_2Co_{11} alloy. Phys Met Metallogr 107:270

Ivanova GV, Shchegoleva NN, Gabay AM (2007) Crystal structure of Zr_2Co_{11} hard magnetic compound. J Alloys Compd 432:135

Ji M, Umemoto K, Wang CZ, Ho KM, Wentzcovitch RM (2011) Ultrahigh-pressure phases of H2O ice predicted using an adaptive genetic algorithm. Phys Rev B 220105(R):84

Kim TK, Takahashi M (1972) New magnetic material having ultrahigh magnetic moment. Appl Phys Lett 20:492

Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Kresse G, Furthmüller J (1996a) Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15

Kresse G, Furthmüller J (1996b) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169

Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. Phys Rev B 59:1758

Lv XB, Zhao X, Wu SQ, Wu P, Sun Y, Nguyen MC, Shi YL, Lin ZJ, Wang CZ, Ho KM (2017) A scheme for the generation of Fe-P networks to search for low-energy $LiFePO_4$ crystal structures. J Mater Chem A 5:14611–14618

Lyakhov AO, Oganov AR, Stokes H, Zhu Q (2013) New developments in evolutionary structure prediction algorithm USPEX. Comp Phys Comm 184:1172–1182

Maddox J (1988) Crystals from first principles. Nature 335:201

Nguyen MC, Choi JH, Zhao X, Wang CZ, Zhang Z, Ho KM (2013) New layered structures of cuprous chalcogenides as thin film solar cell materials: Cu_2Te and Cu_2Se. Phys Rev Lett 111:165502

Nguyen MC, Chen C, Zhao X, Liu J, Wang CZ, Ho KM (2018) Prediction of novel stable Fe-V-Si ternary phase. J Alloys Comp 732:567–572

Oganov AR, Glass CW (2008) Evolutionary crystal structure prediction as a tool in materials design. J Phys Condens Matter 20:064210

Onat B, Cubuk ED, Malone BD, Kaxiras E (2018) Implanted neural network potentials: application to Li-Si alloys. Phys Rev B 97:094106

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865 Phys. Rev. Lett. 78, 1396 (1997)(E)

Pickard CJ, Needs RJ (2011) Ab initio random structure searching. J Phys Condens Matter 23:053201

Skomski R, Sellmyer DJ (2009) Anisotropy of rare-earth magnets. J Rare Earth 27:675

Sugita Y, Mitsuoka K, Komuro M, Hoshiya H, Kozono Y, Hanazono M (1991) Giant magnetic moment and other magnetic properties of epitaxially grown $Fe_{16}N_2$ single-crystal films. J Appl Phys 70:5977

Voroshilov YV, Krypyakevych PI, Kuz'ma YB (1971) Crystal structures of $ZrCo_3B_2$ and $HfCo_3B_2$. Sov Phys–Crystallogr 15:813–816

Wales D, Doye J (1997) Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J Phys Chem A 101:5111–5116

Wang Y, Lv J, Zhu L, Ma Y (2010) Crystal structure prediction via particle-swarm optimization. Phys Rev B 82:094116

Wang Y, Lv J, Zhu L, Ma Y (2012) CALYPSO: a method for crystal structure prediction. Comp Phys Commun 183:2063–2070

Wille LT (1986) Searching potential energy surfaces by simulated annealing. Nature 324:46–48

Woodley SM (2004) Prediction of crystal structures using evolutionary algorithms and related techniques. Struct Bond 110:95–132

Woodley SM, Catlow R (2008) Crystal structure prediction from first principles. Nature Mater 7:937–946

Woodley SM, Battle PD, Gale JD, Catlow CRA (1999) The prediction of inorganic crystal structures using a genetic algorithm and energy minimization. Phys Chem Chem Phys 1: 2535–2342

Wu SQ, Umemoto K, Ji M, Wang CZ, Ho KM, Wentzcovitch RM (2011) Identification of post-pyrite phase transitions in SiO_2 by a genetic algorithm. Phys Rev B 83:184102

Wu SQ, Ji M, Wang CZ, Nguyen MC, Zhao X, Umemoto K, Wentzcovitch RM, Ho KM (2014) Adaptive genetic algorithm for crystal structure prediction. J Phys Condens Matter 26:035402

Wu S, Balamurugan B, Zhao X, Yu S, Nguyen MC, Sun Y, Valloppilly SR, Sellmyer DJ, Ho KM, Wang CZ (2017) Exploring new phases of $Fe_{3-x}Co_xC$ for rare-earth free magnets. J Phys D Appl Phys 50:215005

Zhang L, Han J, Wang H, Car R, E W (2018) Deep potential molecular dynamics: a scalable model with accuracy of quantum mechanics. Phys Rev Lett 120:143001

Zhang WY, Li XZ, Valloppilly S, Skomski R, Shield JE, Sellmyer DJ (2013) Magnetism of rapidly quenched rhombohedral Zr_2Co_{11}-based nanocomposites. J Phys D 46:135004

Zhao X, Shu Q, Nguyen MC, Wang Y, Ji M, Xiang H, Ho KM, Gong X, Wang CZ (2014a) Interface structure prediction from first-principles. J Phys Chem C 118:9524–9530

Zhao X, Nguyen MC, Wang CZ, Ho KM (2014b) New stable Re-B phases for ultra-hard materials. J Phys Condens Matter 26:455401

Zhao X, Nguyen MC, Zhang WY, Wang CZ, Kramer MJ, Sellmyer DJ, Li XZ, Zhang F, Ke LQ, Antropov VP, Ho KM (2014c) Exploring the structural complexity of intermetallic compounds by an adaptive genetic algorithm. Phys Rev Lett 112:045502

Zhao X, Ke L, Nguyen MC, Wang CZ, Ho KM (2015a) Structures and magnetic properties of Co-Zr-B magnets studied by first-principles calculations. J Appl Phys 117:243902

Zhao X, Wu SQ, Lv XB, Nguyen MC, Wang CZ, Lin ZJ, Zhu ZZ, Ho KM (2015b) Exploration of tetrahedral structures in silicate cathodes using a motif-network scheme. Sci Rep 5:15555

Zhao X, Ke L, Wang CZ, Ho KM (2016a) Metastable cobalt nitride structures with high magnetic anisotropy for rare-earth free magnets. Phys Chem Chem Phys 18:31680–31690

Zhao X, Wang CZ, Yao Y, Ho KM (2016b) Large magnetic anisotropy predicted for rare-earth-free $Fe_{16-x}Co_xN_2$ alloys. Phys Rev B 94:224424

Zhao X, Yu S, Wu S, Nguyen MC, Wang CZ, Ho KM (2017a) Structures, phase transitions, and magnetic properties of Co_3Si from first-principles calculations. Phys Rev B 96:024422

Zhao X, Wang CZ, Kim M, Ho KM (2017b) Fe-cluster compounds of chalcogenides: candidates for rare-earth-free permanent magnet and magnetic nodal-line topological material. Inorg Chem 56:14577–14583

Multi-objective Optimization as a Tool for Material Design

116

Zahed Allahyari and Artem R. Oganov

Contents

Z. Allahyari (✉)
Materials science and Engineering, Skolkovo Institute of Science and Technology,
Moscow, Russia
Moscow Institute of Physics and Technology, Moscow, Russia
e-mail: zahed.allahyari@gmail.com

A. R. Oganov
Materials science and Engineering, Skolkovo Institute of Science and Technology,
Moscow, Russia
Moscow Institute of Physics and Technology, Moscow, Russia
International Center for Materials Design, Northwestern Polytechnical University, Xi'an, China
e-mail: a.oganov@skoltech.ru

Abstract

In this chapter, we explain the concept of Pareto optimality and Pareto dominance and use these concepts in solving multi-objective (MO) optimization problems. Then, we discuss a few different MO optimization methods and show how MO optimization can be used as a tool for designing new materials. A simple Pareto-based MO optimization method is examined on a few practical case studies to assess how efficient is this method in optimizing double-objective problems.

1   Introduction

In today's world, computational optimization is essential for solving the problems in a wide area of scientific, technological, and engineering applications. Generally, these problems can be classified into single-objective (SO) and multi-objective (MO) problems. SO problems are problems with only one objective to be optimized. But MO problems are problems with more than one objectives which have to be optimized simultaneously. MO optimization is usually more complex than SO, since instead of a single objective, several, often conflicting objectives need to be optimized. This means that improving one objective usually ends up in degrading the other objectives. In MO problems there is a set of optimal solutions which is called Pareto front. The task of MO optimization methods is to find this optimal set of solutions, so that a decision-maker can then select solutions preferable from various considerations (e.g., cost, or expected ease of synthesis, etc.). Optimizing energy alone gives one the stable phase, which usually (but not always) is easy to synthesize, but does not necessarily possess the best properties. Optimizing physical properties alone gives the theoretical limit for the property of interest – but the predicted phase may be so high in energy that its chances to be synthesizable are negligible. MO optimization, simultaneously optimizing energy and the property of interest, is the best way known to us to find solutions that are practically interesting, possessing attractive properties and with a high chance (albeit not a guarantee) of being synthesizable. In this chapter, it is shown that how MO optimization methods can be useful in solving the central problem of computational material design, which is the discovery of new materials optimal in multiple properties; therefore, at first, several different MO optimization methods are discussed and it is shown how MO optimization can be used as a tool for designing new materials. Then, a simple Pareto-based MO optimization method is applied to a few practical cases, demonstrating the efficiency of this method in optimizing double-objective problems.

2   What Is the Pareto Front?

The goal of MO optimization methods is to find the set of optimal solutions with a trade-off between the objectives. This set of solutions is called Pareto front. To understand how the Pareto front is obtained, the concept of dominance is useful.

Fig. 1 Illustration of dominance concept
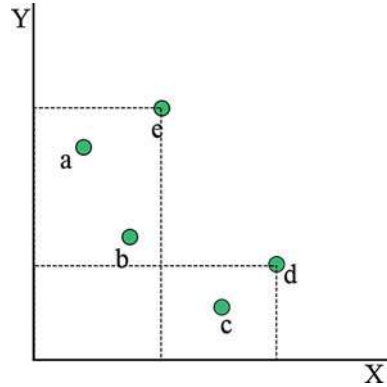
Pareto dominance, Pareto optimality, and Pareto front are the concepts that are commonly used for comparison of candidate solutions in a population. A solution *a* is said to dominate solution *b*, if *a* is at least as good as *b* in every objective and better than *b* in at least one objective. Figure 1 is a schematic representation of solutions on the Ashby plot (Ashby 2011) (Ashby plot is a two-dimensional plot of any two objectives where each objective is plotted along one of the axes of the plot). The goal in this illustration is to find the set of solutions that minimize X and Y objectives in the best way. In this figure, solution *c* dominates solution *d*, because *c* has lower (or better) value of both objectives than *d*. *a* and *b* dominate *e* for the same reason. But *a*, *b*, and *c* do not dominate each other. For example, *a* is better than *b* and *c* in one objective, while it is worse than them in the other objective. Solutions *a*, *b*, and *c* that are not dominated by any other solutions are said to be non-dominated solutions. In fact, Pareto front is a set of non-dominated solutions.

In mathematical terms, for a set of solutions $\{S_n\}$ $(n = 1, \ldots, N)$, each solution with F_m $(m = 1, \ldots, M)$ objectives, MO optimization can be formulated as follows:

```
S = [S₁, S₂, ... Sₙ];              #  The set of solutions,
F = [F₁, F₂, ... Fₘ];              #  The set of objectives,
counter = 0;
Pcounter= Ø;
while length({S}) ≠ 0              # While there are solutions in
                                     the population.
    S = (S ∪ Pcounter) - (S ∩ Pcounter)      # Removing the non-
                                                dominated solutions
                                             # from population
    Sᵢ dominates Sⱼ,  (i,j = 1,...,N and i≠j)
    If     Fₖ(Sᵢ) ≤ Fₖ(Sⱼ)  (k = 1,...,M),   for all objectives
           and
           Fₖ(Sᵢ) < Fₖ(Sⱼ),    At least in one objective
    S*= Sᵢ if   Sᵢ is not dominated by any other solutions
    counter = counter + 1;
    {Pcounter} = S*          # copying non-dominated solutions to {P}
end
```

A solution S^* is not dominated by any other and is, therefore, Pareto optimal. $\{P_t\}$ ($t = 1, \ldots, T$ and $T \leq N$) is a set of non-dominated solutions. After finding the set of non-dominated solutions, $\{P_1\}$ or first Pareto front, these solutions are removed from the main population, and again the non-dominated solutions of the rest of the population are found – these form the second Pareto front, $\{P_2\}$. This process is continued until all solutions are classified in different Pareto fronts. In the worst situation, that each Pareto front only includes one solution, T can be equal to N. This usually happens when objectives are not conflicting, i.e., a linear relation exists between all objectives.

In MO optimization, if one solution can be found which is more optimal than all other solutions in every objective, it is called Utopian solution. For single objectives, there is always a utopian solution, while generally there is no utopian solution for MO problems.

3 Different MO Methods

For designing an efficient Pareto-based MO optimization method, generally, two important factors need to be optimized. The distance to the optimal front is to be minimized, and the diversity of the generated solutions is to be maximized. The latter is useful when more than two objectives are optimized. In the following, we discuss simple Pareto ranking method which is already sufficient for double-objective optimization as well as a few more complex methods in MO optimization.

3.1 Layer Classification (A Simple Pareto Ranking)

The set of non-dominated solutions (solutions forming the Pareto front) are determined using dominance concept. These are the most optimal solutions in the current population, thus the best rank (highest probability of selection) is assigned to this set of solutions. Then these solutions are eliminated from the population temporarily, and again, Pareto front for the rest of population is found. This procedure is repeated until all solutions are classified into different Pareto fronts (Fig. 2). To all the solutions of the same Pareto front, the same fitness is assigned, and there is no priority in selecting the structures of the same Pareto front. Obviously, the highest probability of selection is given to the solution of the first Pareto front, the second highest probability to the second Pareto front, and so on. Then the selection can be performed using any binary selection method (like tournament selection). We call this method "layer classification Pareto technique." Some of the methods that are introduced in this chapter are using this layer classification technique. However, since this method fails in some cases particularly when the number of objectives increases, all of the following methods use some additional technique to make sure that successful results are accessible even if the number of objectives increases.

Fig. 2 Layer classification of solutions in different Pareto fronts for minimizing both objectives
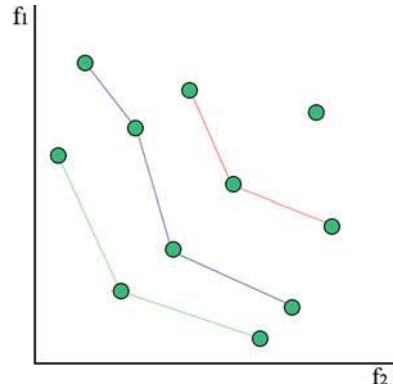
3.2   Vector Evaluated Genetic Algorithm (VEGA) (Schaffer 1985)

This method is not based on Pareto optimality techniques. This method works as a typical genetic algorithm with one difference, which applies in the way of parent selection method. Let us assume an *m-objective* problem with *N*, number of best fraction of solutions in each generation that are allowed to be selected as a parent. At the end of each generation, the population is ranked for each single objective, and the *N/m* of best ranked solutions is added to a hypothetical archive. In this way, if a solution is optimal in *n* objectives *(n ≤ m)*, it is *n* times more likely to be selected than a solution optimal in only one objective. However, this method is susceptible to the shape of the Pareto front, and it fails to generate solutions which do not necessarily excel in one objective but optimal in a sense of Pareto front (Ngatchou et al. 2005).

3.3   Non-dominated Sorting Genetic Algorithm (NSGA)

In the proposed NSGA (Srinivas and Deb 1994), a layer classification method is used, and a sharing function is then applied to the solutions of each Pareto front in the way that some priorities are given to the solutions in the less crowded regions. This sharing function is totally dependent on the choice of sharing parameter, δ_{share}, which is a user-predefined parameter. This parameter denotes the largest value of distance metric within which two solutions share each other's fitness (Deb et al. 2002). But because of the user dependencies and other failures of this fitness sharing method (for more details please refer to reference (Srinivas and Deb 1994)), it was replaced with "crowded-comparison approach" in NSGA-II (Deb et al. 2002) to overcome the failures of fitness sharing method. This approach again gives some priorities to the solutions placed in the less crowded regions of each Pareto front, while it has no user-dependent parameter.

Fig. 3 Illustration of
selection method in PESA
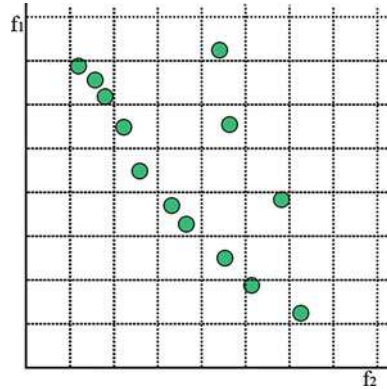
3.4 Pareto Envelope-Based Selection Algorithm (PESA)

In this method, only the non-dominated solutions (Pareto front solutions) are permitted to be ranked and selected. Here, after determining the Pareto front, a hypervolume is defined as a result of finding the closest distance of two neighbor solutions for each objective. This hypervolume estimates the isolation factor for solutions in the way that each solution in a hypervolume with the maximum number of solutions gets the minimum isolation value and vice versa (Fig. 3). The probability of selection of these solutions is proportion to their isolation value. The goal is to give a higher probability to the solutions in the isolated regions than those in the crowded regions. This way of selecting individuals – based on the probability which is given to each of them – is called "individual-based selection." But this method was replaced with "region-based selection" in PESA2. In the region-based selection, instead of giving an isolation value to each individual, the isolation value is given to hypervolumes, and as a result, selection is applied on hypervolumes instead of solutions. This way, the hypervolume which contains less solutions is considered to be more isolated than the one with more solutions (hypervolumes with higher isolation values are preferred). After selection of a hypervolume, one of its solutions is chosen randomly. This method is shown to be more efficient and preferred by the author (Corne et al. 2001) than individual-based selection. For more details on distribution of probabilities and selection mechanism, we refer to the original paper (Corne et al. 2001).

3.5 Strength Pareto Evolutionary Algorithm (SPEA)

Like NSGA and PESA, there are two versions for SPEA, the first version was published on 1999 (Zitzler and Thiele 1999), but later it turned out that this method has serious problems in special situations and was claimed in some works (Corne et al. 2001; Zitzler et al. 2001; Deb et al. 2002) to be unsuccessful in finding the
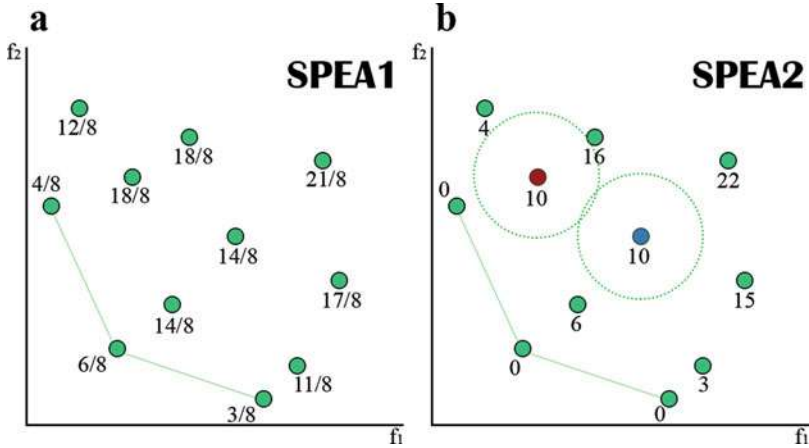
Fig. 4 Different fitness assignment in (**a**) SPEA1 and (**b**) SPEA2 for the same distribution of solutions. According to the density estimation technique which is used in SPEA2, the blue solution has the higher chance of being selected than the red one, since it is placed in the less crowded region

optimal solution. However, in SPEA2 (2001) (Zitzler et al. 2001), the selective fitness assignment method was totally changed, and a more solid algorithm was proposed. Here, we discuss the way of fitness assignment in both methods and briefly mention the main failures of SPEA1.

SPEA1: In this method (Zitzler and Thiele 1999), the population is divided into two groups, *external* (non-dominated solutions) and *internal* (dominated solutions). In SPEA1 optimal solutions can be selected from both external and internal population; hence, fitness is assigned to all population members. This fitness is evaluated using a predefined strength function. For a non-dominated solution (in external population), strength is equal to the number of other solutions it dominates, divided by N, in which N is the size of internal population. For a dominated solution (in the internal population), strength is obtained by summing up the strength of non-dominated individuals that dominate it, divided by N. The selective fitness is equal to the strength of the solution if the solution is a member of external population. Otherwise, selective fitness is equal to the strength of solution plus one (to make sure non-dominated solutions are always preferred to dominated solutions) (Fig. 4a). However, the method fails when many solutions are dominated by the same non-dominated individuals (and consequently, they all have the same fitness). For example, in the particular case that only one non-dominated solution exists, SPEA behaves like a random search algorithm. Moreover, SPEA is mostly unsuccessful in convergence/quick convergence to the optimal solutions.

SPEA2: For the sake of eliminating the deficiencies of SPEA, SPEA2 method was proposed (Zitzler et al. 2001). In fact, the fitness assignment in this method totally differs from the previous version. In SPEA2, for improving the convergence

rate to the best optimal solutions, an archive with a fixed size is defined in which only the solutions of this archive are allowed to be selected. If the number of non-dominated solutions fits the archive's size, then only these solutions are selected. But if the number of non-dominated solutions exceeds the archive's size, non-dominated solutions from the dense regions (regions containing many non-dominated solutions in each other's vicinity) are eliminated from the archive. For more details on density estimation technique which is used in SPEA2, we refer to the original paper (Zitzler et al. 2001). In the case that the size of archive is larger than the number of non-dominated solutions, free places will be filled by dominated solutions according to their fitness (in the case of same fitness value for many solutions, density estimation technique is used).

Fitness assignment: At first, strength is computed for each solution (both dominated and non-dominated), which is equal to the number of other solutions it dominates. Then, a raw fitness is defined which is obtained by summing up the strength of solutions (both dominated and non-dominated) that are dominating the target individual (therefore, raw fitness is zero for all non-dominated individuals). Then a fitness is assigned to each individual, which is equal to the sum of raw fitness and the density of the region (in the case of the same raw fitness for two or more solutions). It must be mentioned that in both SPEA1 and SPEA2, the fitness is to be minimized. Although SPEA2 was claimed to be very effective in MO optimization, it can be clearly seen from Fig. 4b that this algorithm fails in some cases to give a proper ranking to the non-dominated solutions, i.e., the blue solution has better ranking than the red one (because it is placed in a less crowded region), while the red solution belongs to the second Pareto front and blue solution to the third Pareto front.

4   Combining MO Optimization with USPEX for Material Design

We performed MO evolutionary search for materials optimal in two properties: hardness and stability (energy above convex hull), for a few binary systems, i.e., Fe_xB_y, Mo_xB_y, and Mo_xN_y, using the USPEX code (Oganov and Glass 2006; Lyakhov et al. 2013). In these calculations, the simple Pareto ranking method (layer classification Pareto technique) without any further procedures such as fitness sharing, crowded comparison, clustering or any other methods (to consider the density of solutions in the objective space) was used. The reason for this choice is these additional procedures strengthen the MO Pareto-based methods when three or more objectives are under study. Although they can be also useful when dealing with two objectives, it is better to keep it simple as long as it works very well (this method was efficiently used in the prediction of new hard (Kvashnin et al. 2017) and thermoelectric (Núñez-Valdez et al. 2016) materials).

In this method, at the end of each evolutionary generation, a fitness is assigned to each structure, and fitness is derived from the Pareto rank of the structure. Thus,

the highest rank (lowest fitness) will be assigned to the structures forming the first Pareto front, and these structures have the highest probability of being selected as parents. Then structures of the next generation will be created by applying variation operators (heredity and mutation) to these parents. One can expect that the combination of MO optimization and evolutionary search should be very effective. The main reason of this effectiveness is that the method always tries to fill the blank places in the Ashby plot in which structures with multi-optimal properties must be placed. These places are filled by creating new structures that were produced by optimum parent structures chosen from the same region of the Ashby plot. Since an unlimited number of structures can be created, and only optimum structures have a chance to be selected (bad structures are discarded from the population instantly), the blank places of Ashby plot will be filled as long as there are good structures in reality to be placed there (this method is also useful when a divergent Pareto front is obtained). Of course, finding all optimum structures of these regions may need long time and even then is not guaranteed. However, by taking advantage of this well-defined method which intelligently produces optimal solutions, even in a limited time and computations power, useful (even if approximate) Pareto fronts are obtained.

It is worth mentioning that the method of MO optimization discussed here is effective for minimum multidimensionality (two dimensions), and the efficiency decreases dramatically when the number of objectives (properties to be studied for each material) increases. This is obviously due to the increase of the number of optimal structures. The dimensionality of Pareto front is proportional to the number of properties that we wish to optimize. For example, when the dimensionality increases from two to three, the Pareto front is not a line (one-dimensional) anymore. It becomes a plane (two-dimensional). Similarly, for four properties, the Pareto front becomes a 3D hyperplane and so on. Thus, the number of optimal structures which are placed on the Pareto front increases (all having the same highest rank), and the intelligent evolutionary algorithm turns to a useless algorithm with a random parent selection.

In the following, a few examples of crystal structure prediction using evolutionary methodology USPEX in combination with simple Pareto ranking (layer classification) method is given. The goal of these calculations is to find hard and stable structures in target systems using MO optimization method and assess the efficiency of this method along with evolutionary crystal structure prediction. In these examples, all the ab initio calculations were done using the PBE exchange-correlation functional (Perdew et al. 1996) and the projector-augmented wave (PAW) method (Blöchl 1994; Kresse and Joubert 1999), as implemented in the VASP code (Kresse and Furthmüller 1996) with the plane-wave basis set cutoff of 550 eV and Brillouin zone sampling using a grid of spacing $2\pi \times 0.06$ Å^{-1}. Evolutionary calculations for each binary system had initial population size = 120 and subsequent population size of 60 structures and were run for 50 generations. Hardnesses were computed using the Lyakhov-Oganov model (Lyakhov and Oganov 2011).

4.1   Example 1: Mo_xN_y

Since stable compounds of Mo_xN_y system were previously studied (Yu et al. 2016) using SO optimization mode of USPEX, this is a very good example which gives a chance to compare results and consequently realize how efficient is the simplest MO optimization method which is implemented in USPEX.

In the previous work (Yu et al. 2016), MoN, MoN_2, and Mo_4N_3 were the stable compounds under zero pressure. Also, two very low-energy metastable structures for Mo_2N and one metastable structure for Mo_3N_2 were reported (Yu et al. 2016). However, in our search, the results are a little bit different. We found four stable compounds (MoN, MoN_2, Mo_2N, and Mo_3N_2) for this system. The reason for this difference is that in our search, the stable structure of Mo_4N_3 has not been found, and as a consequence, metastable structures of Mo_2N and Mo_3N_2 appeared on the convex hull (Fig. 5a). Although Mo_4N_3 was not found in our search, all other stable and metastable compounds, which were reported in the previous work, were found. Moreover, a new metastable compound (Mo_5N_2) with energy of 2 meV/atom above the convex hull was discovered in our calculation, which has never been reported before. Also, we found another structure for Mo_2N with the *Cmcm* space group that is a little more stable than the computationally predicted $I4_1/amd$ structure (Yu et al. 2016) of this compound. It is worth mentioning that the stable $P6_3/mmc$-MoN_2 structure that was found in our work is in perfect agreement with the stable structure computationally reported for MoN_2 (Yu et al. 2016) and is lower in energy by 0.8 eV/atom (thus, thermodynamically more stable) than the experimentally reported (but later shown by Yu et al. (2016) to be incorrect) R-$3m$ structure.

Altogether, this example shows that MO optimization not only discovered a few unknown low-energy metastable structures but also successfully reported all already studied stable and metastable structures of Mo_xN_y system (except one, Mo_4N_3).
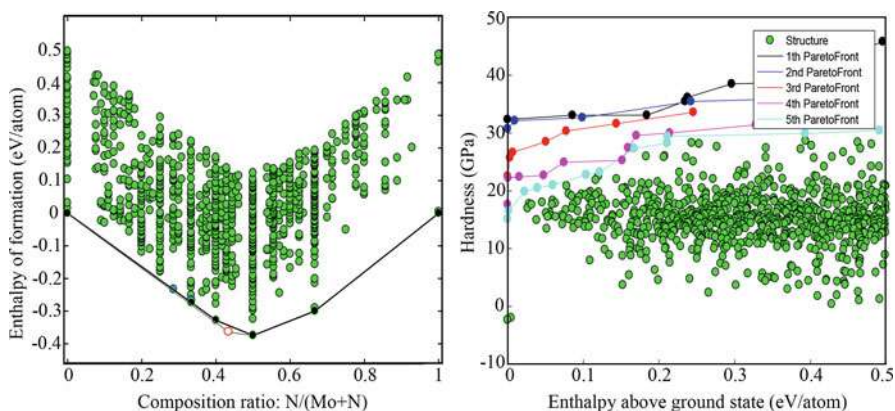
Fig. 5 Convex hull diagram (left) and Ashby plot of hardness and energy above the convex hull (right) for the Mo-N system. The red hollow circle shows the stable Mo_4N_3 compound reported in the literature. Low-energy metastable Mo_5N_2 and *Cmcm* structure of Mo_2N are colored in blue

This indicates that MO optimization method (in the case of optimization of two properties) is almost as efficient as SO optimization.

Figure 5b shows the Ashby plot of hardness and instability of Mo_xN_y system. It can be seen from this figure that several hard and stable/low-energy metastable structures were found in this system. Although most of the structures in this figure are in the hardness range of 10–20 GPa and energy range of 0.1–0.5 eV/atom above the convex hull, it can be clearly seen that the algorithm has tried to focus on the regions where optimal structures in both properties are located.

4.2   Example 2: Fe_xB_y

In the second example, we study the Fe_xB_y system using the same method. For this system two stable binary compounds (FeB and FeB_3) and several metastable compounds (FeB_2, FeB_3, FeB_4, Fe_2B, Fe_3B_5, and Fe_3B_{10}) were found in our work. Some of these compounds were reported to be stable or metastable in the literature (Kolmogorov et al. 2010; Van Der Geest and Kolmogorov 2014), while some are discovered in our work. Below we discuss the differences between our results and reported findings.

In the literature, FeB, FeB_2, and Fe_2B are discussed to be stable, while Fe_2B and FeB_2 turned out to be metastable in our calculation, by 3 meV/atom and 42 meV/atom, respectively. Even though both of these compounds were found in our search, the lowest-energy structures of these compounds still have not been found. The other difference is that two structures of FeB_3 were found in our calculation. Most likely, this compound would not appear on the convex hull if the stable structure of FeB_2 was found; nevertheless, FeB_3 is not reported neither as a stable nor as a metastable compound in the literature. FeB_4 was found to be metastable in our work, which is in agreement with the literature (Kolmogorov et al. 2010; Van Der Geest and Kolmogorov 2014). Fe_4B is one of the reported metastable compounds that have not been detected in our calculation. In this calculation, several simultaneously hard and stable/low-energy metastable phases were detected for this system, and some of these are superhard (Fig. 6b).

4.3   Example 3: Mo_xB_y

Mo_xB_y system is the third example in our list. For this system, MoB and MoB_2 were found to be stable, which is fully in agreement with the literature (Spear and Liao 1988; Liang et al. 2012). In addition to these stable compounds, many metastable compounds were reported both theoretically (Zhang et al. 2010) and experimentally (Spear and Liao 1988) in the literature, i.e., MoB_3, MoB_4, Mo_2B, Mo_2B_5, Mo_3B_2, and Mo_5B_3. In our search, we found eight low-energy metastable compounds (Mo_2B, MoB_3, MoB_5, Mo_3B_5, Mo_3B_2, Mo_4B, Mo_5B_2, and Mo_2B_3), three of them are already reported metastable compounds (Mo_2B, Mo_3B_2, and MoB_3), and five of them, MoB_5, Mo_2B_3, Mo_3B_5, Mo_4B, and Mo_5B_2, with the
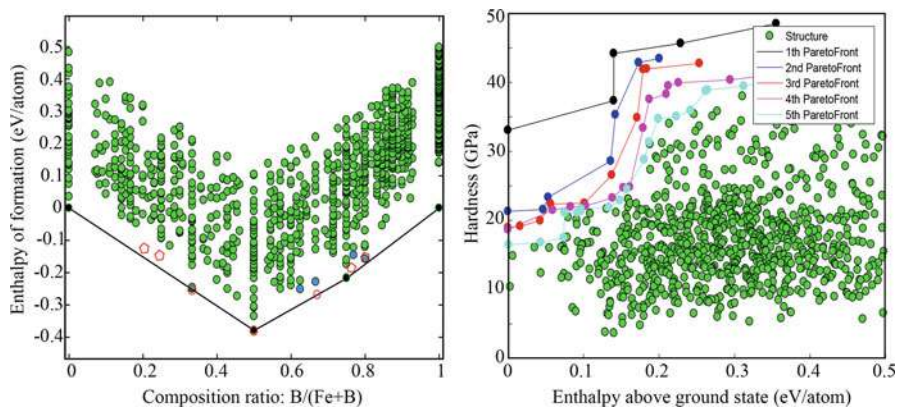
Fig. 6 Convex hull diagram (left) and Ashby plot of hardness and energy above the convex hull (right) for the Fe-B system. The red hollow circles and polygons show the stable and metastable compounds reported in the literature. Low-energy metastable structures found in our work are colored in blue
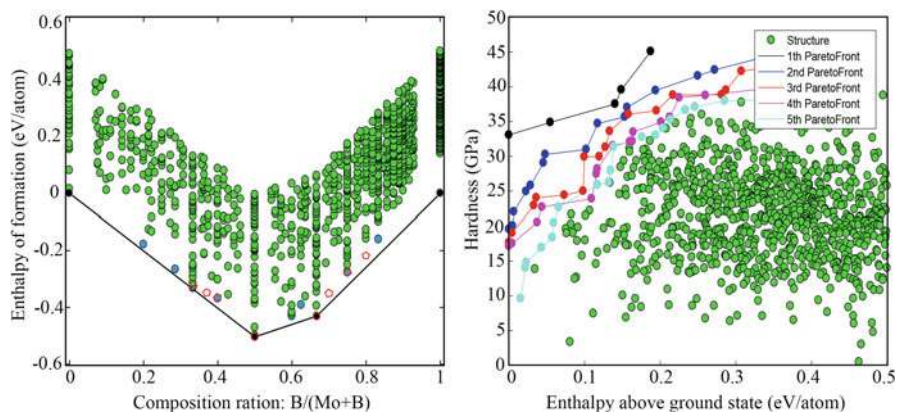
Fig. 7 Convex hull diagram (left) and Ashby plot of hardness and energy above the convex hull (right) of the Mo-B system. The red hollow polygons show the metastable compounds reported in the literature, while blue circles indicate the compounds discovered in our work

55 meV/atom, 29 meV/atom, 57 meV/atom, 23 meV/atom, and 21 meV/atom energy above the convex hull, respectively, are unknown compounds which are discovered in our work. Although three reported metastable compounds (MoB_4, Mo_2B_5, and Mo_5B_3) have not been found in our calculation, the MO optimization acted successfully by discovering five new low-energy metastable compounds of Mo_xB_y system, which indicates the efficiency of the method in crystal structure prediction with multi-optimal properties. In Figs. 6b and 7b, the hardest stable compound (with the hardness of 33.6 GPa) corresponds to pure α-boron.

Stable binary compounds with the highest Lyakhov-Oganov hardness for each of the above examples are MoN_2 (32.6 GPa), FeB_3 (20 GPa), and MoB_2 (19.6 GPa).
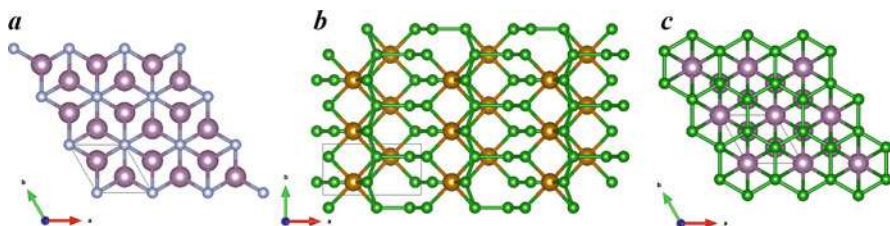
Fig. 8 Structures of stable binary compounds with the highest Lyakhov-Oganov hardness in the systems considered here: (**a**) *P6₃/mmc*-MoN2 (Chen's hardness 22.3 GPa), (**b**) *P2₁/m*-FeB3 (Chen's hardness 30.2 GPa), (**c**) *R-3 m*-MoB2 (Chen's hardness 28.5 GPa)

These phases have space groups *P6₃/mmc*, *P2₁/m*, and *R-3m*, respectively, and are shown in Fig. 8. While Lyakhov-Oganov model of hardness is convenient for screening (it is very cheap and numerically robust), more accurate values are predicted using Chen's model (Chen et al. 2011) – the latter, however, requires calculations of the elastic tensor and is therefore much more computationally expensive and is sensitive to numerics. Thus, after initial screening, the final hardnesses were computed using Chen's model.

5   Conclusion

In this chapter, some of the most prominent multi-objective optimization methods were discussed, and the "layer classification" Pareto technique in combination with USPEX was used in search for simultaneously hard and stable materials. The most realistic and favorable outcome of material design is the discovery of new materials with simultaneously optimal multiple properties; i.e., in our examples, unstable hard materials are as useless as soft stable materials. We need to have materials which are optimal in all required properties. Therefore, MO optimization method can be used as a verified tool for materials discovery. Searching for materials with optimal multi-properties is possible using the combination of evolutionary search and MO Pareto ranking. The combination of these two provides a powerful tool for the search and discovery of new materials.

References

Ashby MF (2011) Materials selection in mechanical design. Butterworth-Heinemann, Burlington
Blöchl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953–17979. https://doi.org/10.1103/PhysRevB.50.17953
Chen X-Q, Niu H, Li D, Li Y (2011) Modeling hardness of polycrystalline materials and bulk metallic glasses. Intermetallics 19:1275–1281. https://doi.org/10.1016/J.INTERMET.2011.03.026

Corne D, Jerram N, Knowles JD et al (2001) PESA-II: region-based selection in evolutionary multiobjective optimization. In: Proceedings of the genetic and evolutionary computation conference, pp 283–290

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6:182–197. https://doi.org/10.1109/4235.996017

Kolmogorov AN, Shah S, Margine ER et al (2010) New superconducting and semiconducting Fe-B compounds predicted with an *Ab Initio* evolutionary search. Phys Rev Lett 105:217003. https://doi.org/10.1103/PhysRevLett.105.217003

Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186. https://doi.org/10.1103/PhysRevB.54.11169

Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. Phys Rev B 59:1758–1775. https://doi.org/10.1103/PhysRevB.59.1758

Kvashnin AG, Oganov AR, Samtsevich AI, Allahyari Z (2017) Computational search for novel hard chromium-based materials. J Phys Chem Lett 8:755–764. https://doi.org/10.1021/acs.jpclett.6b02821

Liang Y, Yuan X, Fu Z et al (2012) An unusual variation of stability and hardness in molybdenum borides. Appl Phys Lett 101:1–6. https://doi.org/10.1063/1.4764547

Lyakhov AO, Oganov AR (2011) Evolutionary search for superhard materials: methodology and applications to forms of carbon and TiO 2. Phys Rev B 84:92103. https://doi.org/10.1103/PhysRevB.84.092103

Lyakhov AO, Oganov AR, Stokes HT, Zhu Q (2013) New developments in evolutionary structure prediction algorithm USPEX. Comput Phys Commun 184:1172–1182. https://doi.org/10.1016/j.cpc.2012.12.009

Ngatchou P, Zarei A, El-Sharkawi A (2005) Pareto multi objective optimization. In: Proceeding of the 13th international conference on, intelligent systems application to power systems, pp 84–91. https://doi.org/10.1109/ISAP.2005.1599245

Núñez-Valdez M, Allahyari Z, Fan T, Oganov AR (2016) Efficient technique for computational design of thermoelectric materials. Comput Phys Commun 222:152–157. https://doi.org/10.1016/j.cpc.2017.10.001

Oganov AR, Glass CW (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. J Chem Phys 124:244704. https://doi.org/10.1063/1.2210932

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865

Schaffer JD (1985) Multiple objective optimization with vector evaluated genetic algorithms. In: 1st International conference on genetic algorithms, pp 93–100

Spear KE, Liao PK (1988) The B−Mo (Boron-Molybdenum) system. Bull Alloy Phase Diagr 9:457–466. https://doi.org/10.1007/BF02881867

Srinivas N, Deb K (1994) Multiobjective optimization using nondominated sorting in genetic algorithms. Evol Comput 2:221–248. https://doi.org/10.1017/CBO9781107415324.004

Van Der Geest AG, Kolmogorov AN (2014) CALPHAD: computer coupling of phase diagrams and thermochemistry stability of 41 metal − boron systems at 0 GPa and 30 GPa from fi rst principles. 46:184–204. https://doi.org/10.1016/j.calphad.2014.03.005

Yu S, Huang B, Jia X et al (2016) Exploring the real ground-state structures of molybdenum dinitride. J Phys Chem C 120:11060–11067. https://doi.org/10.1021/acs.jpcc.6b00665

Zhang M, Wang HH, Wang HH et al (2010) Structural modifications and mechanical properties of molybdenum borides from first principles. J Phys Chem C 114:6722–6725. https://doi.org/10.1021/jp100225c

Zitzler E, Thiele L (1999) Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans Evol Comput 3:257–271. https://doi.org/10.1109/4235.797969

Zitzler E, Laumanns M, Thiele L (2001) SPEA2: improving the strength pareto evolutionary algorithm. Evolutionary methods for design optimization and control with applications to industrial problem, pp 95–100. https://doi.org/10.3929/ethz-a-004284029

Minima Hopping Method for Predicting Complex Structures and Chemical Reaction Pathways

117

Maximilian Amsler

Contents

Abstract

The arrangement of atoms in a material fundamentally defines its properties. Predicting the crystal structure of a solid or the most stable isomer of a molecule or cluster from computation given solely the chemical composition remains a challenging task, which requires solving a complex optimization problem and identifying the global minimum on a high-dimensional potential energy surface. The minima hopping method uses a walker to explore this energy landscape by performing hops from one local minimum to the next. These hops are based on molecular dynamics trajectories followed by local relaxations, which are tuned to exploit the Bell-Evans-Polanyi principle, and drive the search toward low-energy

M. Amsler (✉)
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY, USA
e-mail: amsler.max@gmail.com

solutions. The details of the algorithm are discussed, and its application in materials discovery challenges is illustrated based on two examples.

1   Introduction

Knowing the atomic structure of a condensed matter system is of fundamental importance in materials science and an essential prerequisite for atomistic materials modeling. Structure prediction encompasses the task of finding the arrangement of atoms in a cluster or a crystal that minimizes a specific target thermodynamic potential. For many practical applications, one would like to know the most stable phase given only the chemical composition at a specific pressure and temperature, which corresponds to the minimum of the Gibbs free energy $G = E - TS + pV$. A method which allows to accurately determine this phase prior to synthesis in turn allows the prediction of its properties and, therefore, provides a powerful tool in theoretical materials research.

For a system with N_{at} atoms at zero pressure and temperature, a structure is considered kinetically stable if it corresponds to a local minimum on the potential energy surface (PES) (Wales 2003). For molecules and clusters, the PES is a function of the atomic coordinates $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{Nat}$, whereas for periodic systems, additional degrees of freedom describe the cell vectors, commonly summarized by a matrix $h = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. Every local minimum M is located at the bottom of its catchment basin, within which small perturbations of the coordinates will lead to forces that will drive the system back to the structure M. At finite temperature, the displacements due to thermal excitations can be large enough that the system jumps from one catchment basin to the next if the barriers separating them are sufficiently low. In this way, the system can move toward local minima which are lower in energy. Every local minimum which is not the lowest in energy is called metastable and will eventually decay to a lower energy state. Depending on the detailed features of the PES, metastable phases can however persist over a long time scale. The global minimum is considered the thermodynamic ground state, for which no other lower energy local minima exist.

Structure prediction algorithms aim at identifying this ground state structure by globally optimizing the PES. Since the number of possible atomic arrangements in a system scales exponentially with the dimensionality of the search space, an exhaustive sampling of the PES for any system of realistic size is computationally prohibitive. Fortunately, the PES can be significantly simplified by eliminating irrelevant features, thereby reducing its complexity. In particular, one can subdivide the PES into disjoint regions which uniquely correspond to catchment basins of local minima: every point on the PES can be mapped to its local minimum by

simply performing a local geometry relaxation using a steepest descent method with a sufficiently small step size. In this way, all unreasonable structures, e.g., with unphysically short atomic bond lengths, are removed and not considered in the global optimization process.

Despite the reduced complexity of this discretized PES, the number of local minima for any realistic system can be large. In fact, Stillinger (1999) showed that the number of distinguishable local minima increases exponentially with the number of atoms in a system. Hence, mapping out all minima exhaustively and ranking them according to their energies is simply not feasible. Therefore, any optimization algorithm must rely in one way or another on an efficient sampling of only the low-lying local minima of the PES. However, since the sampling is now not exhaustive, there is no way of determining if the lowest energy structure encountered during the search is indeed the ground state.

The reliable representation of a PES poses a further computational challenge. Density functional theory (Parr and Weitao 1994) (DFT) calculations provide accurate results but are computationally demanding and currently only allow reliable structural searches in systems with at most 100 atoms. Classical force fields (Buckingham 1938; Dick and Overhauser 1958; Tosi and Fumi 1964; Daw and Baskes 1984) or semiempirical methods (Goringe et al. 1997; Papaconstantopoulos and Mehl 2003) on the other hand can treat hundreds or thousands of atoms by neglecting or approximating quantum mechanical atomic interactions, but are less accurate and can introduce spurious local minima on the PES which increase the search space (Ghasemi et al. 2010). Very recently, a new class of interatomic potentials have emerged based on machine learning techniques using artificial neural networks (ANN) (Behler and Parrinello 2007). Trained on ab initio date, such ANNs provide high accuracies at computational costs of force fields and therefore represent a promising prospective alternative. A detailed discussion of these methods is beyond the scope of this chapter and can be found elsewhere.

2   The Minima Hopping Method

2.1   Principles of the Minima Hopping Method

The minima hopping method (MHM) (Goedecker 2004; Amsler and Goedecker 2010) aims at mapping out the low-lying local minima of any complex potential energy or enthalpy landscape, to ultimately find its global minimum. For this purpose, the MHM uses a walker to explore the PES by hopping from one local minimum to the next. These hops are performed based on a set of strategic rules that efficiently drive the system toward low-energy structures. On one hand, the walker performs hops that, on average, lead to a decrease in energy. On the other hand, the walker avoids revisiting already known local minima on the PES, since it is inefficient to sample already explored regions on the PES. However, a revisiting of local minima is not completely prohibited. The latter rule is especially important for complex PESs that contain multiple funnels, which might be connected through one

or several low-energy funnels. Such hub-funnels must be permitted to be traversed to get from one portion of the PES to another. Otherwise, the walker could get trapped in a funnel that does not contain the global minimum, thereby failing to find the ground state.

There are many different ways how the hops on the PES could be performed, and several options have been proposed in the past (Sicher et al. 2011). Random displacement of the atomic positions (and the cell vectors for periodic systems) is a simple choice, but is rather inefficient. Better moves exploit the Bell-Evans-Polanyi (BEP) principle (Jensen 1998), which states that exothermic chemical reactions have low activation energies. Since transitions between neighboring local minima are essentially chemical reactions, the BEP principle implies that it is preferable to cross over low-energy barriers, behind which low-energy local minima are more likely to be found. Note that the BEP principle does not strictly apply to every transition between local minima. However, it has been shown that the BEP principle holds on average for pairs of neighboring local minima (Roy et al. 2008), which is sufficient to bias the moves toward a downhill direction in a global structural search.

A natural choice for a move that exploits the BEP principle is to explore all first-order transition states surrounding a local minimum and to propagate the system preferably across the lowest energy barriers toward the neighboring local minimum. Extensive tests have shown that such moves are indeed very efficient and lead to a high success rate in finding the ground state structure with a low number of hops. However, the computational cost to locate a transition state, using, for example, the dimer method (Henkelman and Jónsson 1999), is computationally demanding. On one hand, the number of force evaluations to converge the saddle point search is high compared to the force evaluations needed for local relaxations. On the other hand, new saddle points need to be found and crossed every time a local minimum is revisited to preserve the ergodicity of the system and to avoid the walker from "going in circles." Hence, the saddle point search method is efficient in terms of hops, but inefficient in terms of energy and force evaluations (Sicher et al. 2011).

The most efficient moves are based on short molecular dynamics (MD) trajectories. Using a constant energy (or enthalpy) ensemble and by starting the MD run from a local minimum, the height of the barrier that can be crossed is limited by the initial kinetic energy E_{kin} of the system due to energy conservation. Hence, the BEP principle can be exploited by keeping the value of E_{kin} as low as necessary to preferably cross over low-energy barriers. In practice, an MD trajectory connecting two basins of attraction rarely passes exactly through the saddle point. Nevertheless, the MD moves are on average more efficient than the saddle point search moves, both in terms of hops and the number of force evaluations. The superiority of MD moves can be explained by the fact that several barriers are frequently crossed during a single MD escape trial with less force evaluations than a saddle point search. Another advantage of the MD moves is that MD routines are readily available in many atomic simulation packages or are fairly easy to implement. For molecular systems, a velocity Verlet algorithm can be used (Verlet 1967). For periodic systems, the cell variables are also included in the dynamics to allow transitions between different crystal lattices. Several Lagrangians have

been proposed for variable volume or variable cell shape MD, e.g., the Parrinello-Rahman (1980) or Wentzcovitch (1991) method, which lead to equations of motion for both the atomic and cell coordinates.

2.2   The Minima Hopping Algorithm

The MHM algorithm consists of two parts, an inner loop to perform the moves on the energy landscape and an outer acceptance/rejection part. Fig. 1 shows a flowchart of the algorithm. Starting from an initial seed structure M_{cur} (representing the current local minimum), the inner loop attempts an escape trial by performing a short MD run. The atomic velocities are initialized randomly according to a Boltzmann distribution with an average kinetic energy E_{kin}. The potential energy during the MD simulation is monitored, keeping track of the number of energy minima and maxima encountered along the MD trajectory. As soon as the MD_{min}-th minimum is reached, the MD run is stopped, and the system is relaxed to a local minimum M. With reasonable values of MD_{min} between 2 and 6, the number of time steps is typically in the range of 50–150. If the kinetic energy is not sufficiently high, the system simply oscillates in the vicinity of M_{cur} without crossing a barrier to a neighboring basin and will relax back to the same local minimum $M = M_{\text{cur}}$. In this case, the kinetic energy of the system is slightly increased by a factor $\beta_s > 1$, and an escape trial is performed with a new random velocity distribution. This process is repeated until the escape is successful ($M \neq M_{\text{cur}}$), at which point the structure M is proposed to the outer part.

The outer part introduces a preference for low-energy structures based on a simple threshold approach: the proposed structure M is accepted only if the energy difference with respect to the current local minimum M_{cur} is less than E_{diff}, i.e., if $E(M) - E(M_{\text{cur}}) < E_{\text{diff}}$. This condition is obviously fulfilled in the case of a downhill energy step, which is the most desirable outcome. On the other hand, M is rejected if its energy is higher by more than E_{diff} compared to $E(M_{\text{cur}})$. To avoid the algorithm from getting trapped in any part of the PES, the threshold value E_{diff} is dynamically adjusted throughout the simulation: E_{diff} is increased by a factor $\alpha_r > 1$ whenever M is rejected and decreased by $\alpha_a < 1$ if accepted. Hence, if the inner loop only proposes high-energy structures, they are eventually accepted after a sufficient amount of rejections. The ratio between α_a and α_r determines the fraction of accepted/rejected steps. In practice, a ratio of $\frac{\alpha_a}{\alpha_r} = 1$ leads to good efficiency, where half of all proposed structures are accepted.

An essential component of the MHM algorithm is the second feedback mechanism on the kinetic energy, which is based on a history of visited local minima. After every successful escape step, the minimum M is compared against a list of all known structures visited up to this point. If M is new, it is added to the history and the value of E_{kin} is reduced by a factor $\beta_n < 1$. On the other hand, if M corresponds to a known, old minimum, E_{kin} is increased by $\beta_o > 1$ to drive the system away from portions of the PES that have already been explored. In other words, by keeping track of the visited local minima, the MHM continuously "learns" more about the

Fig. 1 A flowchart of the MHM algorithm. The inner loop is responsible for efficient MD moves on the PES, while the outer part introduces a preference for low-energy structures. A feedback mechanism on the kinetic energy of the MD escape trials drives the search toward new, unexplored regions of the PES
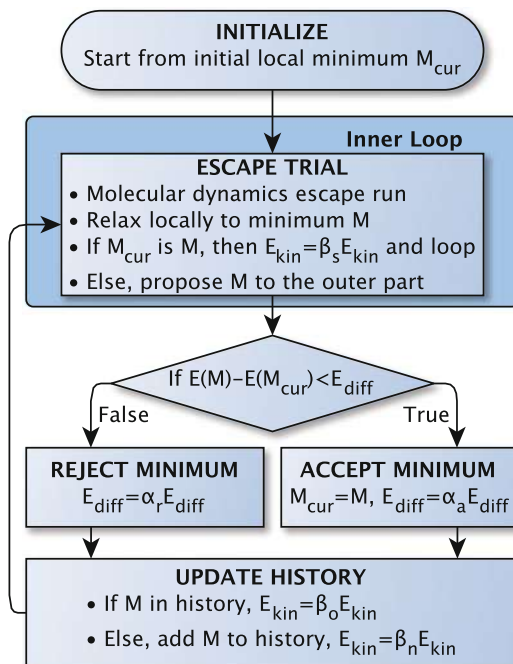
topology of the PES and avoids the least interesting regions. Note that this feature distinguishes the MHM significantly from any thermodynamic method based on a Metropolis algorithm with a Boltzmann factor, like simulated annealing (Salamon et al. 2002) or basin hopping (Doye et al. 1998; Doye and Wales 1998; Wales and Doye 1997), which aims at generating a thermodynamic distribution.

Fig. 2 shows schematically how the two dynamically adjusted parameters E_{diff} and E_{kin} influence the possible and allowed steps on a PES. Points A and B represent two distinct current states M_{cur} of a system at two different stages during a MHM run. Hops from A to A' are allowed but are rejected since the energy difference exceeds E_{diff}, whereas hops from A to A" are accepted. On the other hand, hops from B to B" are impossible due to the limited value of E_{kin}, and only an escape to B' is possible.

The interplay between the parameters α_i and β_j influences the performance of the MHM. Typical values in practice are $\alpha_r = \alpha_a^{-1} = 1.05$ and $\beta_s = \beta_o = \beta_n^{-1} = 1.05$, but slight deviations from these parameters do not significantly affect the efficiency. It is however important that any reasonable choice of the parameters satisfies the so-called explosion condition. It can be shown that the kinetic energy will increase exponentially in well explored regions of the PES by choosing $\frac{\log(\alpha_r)}{\log(1/\alpha_a)} \geq \frac{\log(1/\beta_n)}{\log(\beta_o)}$, while β_s must be greater than 1. This explosion condition not only ensures that the system will never get stuck in any part of the PES during a MHM simulation but more importantly that the walker will efficiently leave already explored funnels.
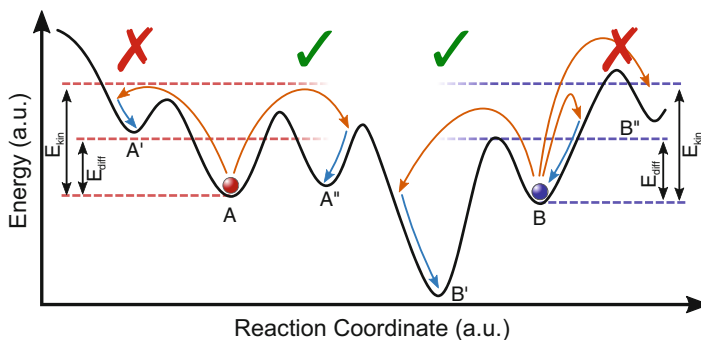
Fig. 2 A one-dimensional potential energy landscape with two structures at points A and B, corresponding to M_{cur} at two different stages of a MHM run. Orange arrows represent MD escape paths, while blue arrows are local relaxations

2.3   Optimizing Moves on the Potential Energy Surface

Chemical reactions are rare events, and observing a structural transitions can take a long time in conventional MD simulations during which the system mostly oscillates in the catchment basin surrounding a local minimum. To avoid such inefficient oscillations and to accelerate the transition across barriers, the MD velocities are biased prior to every MHM escape step. Instead of using a random direction for the initial velocity, it is approximately aligned along a soft-mode direction of the current local minimum, i.e., along a direction of low curvature. This procedure is called softening and leads to several major benefits for the MHM: first, it eliminates the high-frequency vibrations. Consequently, the integration of the equations of motion can be performed with larger and fewer time steps, thereby reducing the computational cost. Second, the dynamics is biased toward global movements of the whole system which cover a larger distance in the configurational space. Third, there is a higher probability of encountering a low-energy barrier along a soft-mode direction. Fig. 3 shows the strong correlation between the curvature of low-frequency eigenmodes and the barriers encountered along these directions.

Fig. 3 Correlation between the barrier heights and the corresponding curvatures of local minima for a LJ$_{19}$ cluster. The saddle points were found using a mode-following dimer method. The colors indicate the mode index, and the large circles represent the median curvature and barrier values
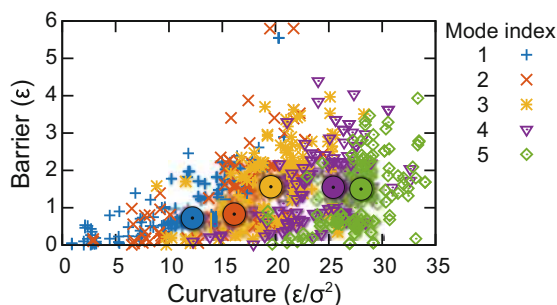
Hence, trajectories along low-curvature directions will lead to low-energy barriers, behind which one more likely finds low-energy structures according to the BEP.

One way to identify these soft-mode directions is to explicitly compute and diagonalize the Hessian matrix $H_{i,j} = \frac{\partial^2 E}{\partial x_i \partial x_j}$ at the local minimum, where x_i denote the components of a vector in configurational space, and to select the eigenvectors corresponding to the lowest eigenvalues. However, starting the MD trajectories every time along the same direction would lead to a loss of ergodicity. Hence, instead of aligning the velocity vector $\mathbf{V}$ directly along the lowest eigenmode, a superposition $\mathbf{V} = \sum_i \frac{\mathbf{e}_i}{\varepsilon_i}$ of a randomly selected subset of eigenvectors $\mathbf{e}_i$ weighted by the inverse of the corresponding eigenvalues ε_i is a better choice.

Methods to compute the complete Hessian $H_{i,j}$ explicitly are computationally demanding and cannot be employed in practical applications. Instead, if a quasi-Newton method is used to perform the local geometry optimizations, an approximate Hessian is naturally available for each local minimum. Geometry optimizers like the Broyden-Fletcher-Goldfarb-Shanno method use an iterative scheme to update the approximate Hessian based on the gradient vectors along the relaxation pathway (Press et al. 1992; Chong and Zak 2001). These Hessians are symmetric and positive definite by construction and can be readily diagonalized. In most cases, the resulting eigenvectors with low eigenvalues are sufficiently accurate and approximately point in the general direction of soft modes.

Another method is to use a dimer method to iteratively identify the eigenmodes using a minimization scheme. In practice, it is not even necessary to find the exact eigenvectors of the Hessian, but merely to eliminate any components along hard modes, a procedure called softening (Schönborn et al. 2009). Starting from the local minimum at $\mathbf{R}_m$ in the high-dimensional configurational space, an image is created at $\mathbf{R}_d^0 = \mathbf{R}_m + d \frac{\mathbf{V}}{|\mathbf{V}|}$ along the initial, randomly selected escape direction $\mathbf{V}$, where d is the fixed length of the dimer. The energy and forces are evaluated at the image, which is then iteratively moved according to $\mathbf{R}_d^{i+1} = \mathbf{R}_d^i + \alpha \mathbf{F}_{d,\perp}^i$, where $\mathbf{F}_{d,\perp}^i = \mathbf{F}_d^i - \frac{\mathbf{D}^i \mathbf{F}_d^i}{d^2} \mathbf{D}^i = \mathbf{F}_d^i - \mathbf{F}_{d,\parallel}^i$ is the force component perpendicular to the dimer $\mathbf{D}^i = \mathbf{R}_d^i - \mathbf{R}_m$ acting on the image and α is a small step size. This procedure corresponds to an iterative rotation of the dimer toward a direction pointing along a soft mode in the configurational space. The dimer length is rescaled after each iteration to keep it at the value d. Essentially, the energy of the image $\mathbf{R}_d$ is minimized subject to the constraint that it remains on a hypersphere with radius d, centered at $\mathbf{R}_m$.

The curvature along the dimer can be approximately evaluated either by using the second-order finite difference of the energies, $\kappa_e^i = 2 \frac{E(\mathbf{R}_d^i) - E(\mathbf{R}_m^i)}{d^2}$, or from the force acting on the image along the dimer $\kappa_f^i = \frac{|\mathbf{F}_{d,\parallel}^i|}{d}$. In the limit of a small dimer size and assuming that the PES close to the minimum is harmonic, both approximations give the same values of the curvature $\kappa_i := \lim_{d \to 0} \kappa_e^i = \lim_{d \to 0} \kappa_f^i$. Eventually, the dimer will converge to point exactly toward the lowest curvature mode, and one would again sacrifice ergodicity. In fact, there exists an optimal amount of softening steps after which the efficiency of the MHM starts to decrease. In practice, the softening procedure is stopped after a fixed amount of iterations, commonly in

the range of 10–40 steps. Alternatively, the softening is terminated as soon as the approximate curvature drops below a certain fraction of the initial value κ_0, e.g., if $\kappa_i < 0.1\kappa_0$.

2.4   Metrics for Measuring Structure Similarities

Assessing the similarities (or dissimilarities) between two structures is essential in two contexts in the MHM algorithm: to decide if an MD escape trial was successful and to determine if a given structure is present in the history of visited local minima. A proper measure of configurational distances $d(a, b)$ between structures a and b must satisfy the properties of a metric, namely, positiveness $(d(a, b) \geq 0)$, symmetry $(d(a, b) = d(b, a))$, the coincidence axiom $(d(a, b) = 0 \Leftrightarrow a = b)$, and triangular inequality $(d(a, b) + d(b, c) \geq d(a, c))$. Hence, instead of directly comparing the Cartesian coordinates of two configurations, the atomic arrangement in a system is commonly transformed to some structural fingerprint, which can be compared through a norm to measure their differences. For a molecular system, such a fingerprint representation must be invariant with respect to any translation and rotation of a structure. For a periodic system, it must additionally be invariant with respect to the arbitrary choice of the lattice vectors. Most fingerprints that satisfy these criteria rely on a descriptor derived from averaged local atomic environments. Radial distribution functions and approaches based on computed powder diffraction patterns are such examples (Oganov and Valle 2009), but they don't take explicitly into account any angular information. Hence, a large portion of the structural data is lost, and two distinct structures can have identical fingerprints. To alleviate this issue, some fingerprints, like the bond characterization matrix (BCM) (Wang et al. 2012, 2015), explicitly take into account angular information.

The ideal fingerprint metric is sensitive enough to distinguish two different structures, but at the same time tolerates numerical noise on the atomic and cell coordinates, which naturally arise from a limited number of geometry relaxation steps. The computational cost of both fingerprint evaluation and comparison must be low compared to the cost of an energy and force evaluation. Such efficiency considerations might be irrelevant when performing a computationally demanding structural search directly at the ab initio level but can become a bottleneck issue when using classical force fields to approximate the PES.

The MHM relies on fingerprints that are derived from an overlap of Gaussian-type orbitals (GTO) (Sadeghi et al. 2013; Zhu et al. 2016). One or several normalized GTOs $G_i(\mathbf{r})$ with index i are placed at the Cartesian coordinate of each atom $\mathbf{r}_{w(i)}$, where $w(i)$ denotes the index of an atom. To make the fingerprint sensitive toward distinct atomic species, the width of each GTO is given by the covalent radius of the corresponding atom $w(i)$. For a molecular structure with free boundary conditions, the Gaussian overlap matrix $S_{i,j}$ is then the integral over all pairs of GTOs,

$$S_{i,j} = \int G_i(\mathbf{r}) G_j(\mathbf{r}) d\mathbf{r} \tag{1}$$

This matrix is symmetric and positive definite and can be therefore diagonalized. The structural fingerprint is given by the corresponding set of eigenvalues, sorted in descending order in a vector $\mathbf{V}$. For a molecule with N_{at} atoms, the dimension of this vector is $n_{orb}N_{at}$, where n_{orb} is the number of GTO per atom ($n_{orb} = 1$ for s-orbitals, $n_{orb} = 4$ for $s + p$-orbitals, etc.). Hence, the distance between the fingerprints $\mathbf{V}^a$ and $\mathbf{V}^b$ of two structures a and b can be computed simply by the Euclidean distance

$$d(a, b) = \sqrt{\sum_i (\mathbf{V}_i^a - \mathbf{V}_i^b)^2} \tag{2}$$

In a crystal, fingerprints obtained according to Eq. (1) are not invariant with respect to the choice of the periodic cell. This issue is addressed by using atomic fingerprints instead of a global one, arising from atomic overlap matrices $S_{i,j}^k$ which capture the local environment around each atom k individually. For every atom in the unit cell, a large sphere with radius r_c is cut out from a supercell, centered around $\mathbf{r}_k$. To avoid discontinuities in the eigenvalues arising from atoms crossing the border of the sphere, the GTOs are scaled with a radial weight that goes smoothly to zero at the cutoff radius, $f_i^k = \left(1 - \frac{|\mathbf{r}_k - \mathbf{r}_{w(i)}|^2}{r_c^2}\right)^n$. The value of n determines the order of continuous derivatives of the smooth cutoff function at the surface of the sphere, and a value of $n = 3$ is commonly used. The cutoff radius is chosen according to $r_c = \sqrt{2n}\sigma_c$, where the characteristic length scale σ_c is typically twice the largest covalent radius in the system. For this choice of parameters, the selected pseudo-molecule contains in the range of 40–80 atoms, depending on the density of the material.

Within this pseudo-molecule, the environment-dependent atomic overlap matrix is computed according to

$$S_{i,j}^k = \int f_i^k G_i(\mathbf{r} - \mathbf{r}_{w(i)}) f_j^k G_j(\mathbf{r} - \mathbf{r}_{w(j)}) d\mathbf{r} \tag{3}$$

The set of eigenvalues of this atomic overlap matrix in descending order represents the atomic fingerprint vector $\mathbf{V}_k$. Since the number of atoms within the cutoff sphere can vary for different atoms k, the dimensions of $\mathbf{V}_k$ are not uniquely defined. In practice, additional zeros can be appended to $\mathbf{V}_k$ if required for a vector operation. In fact, any atom entering the pseudo-molecule across the surface of the cutoff sphere gives rise to an additional entry in $\mathbf{V}_k$, which continuously transforms from zero to a finite value. Hence, the atomic fingerprint vector changes continuously with respect to any changes in the atomic coordinates.

A metric $d(a, b)$ to define uniquely a configurational distance between two crystal structures a and b must be invariant under the exchange of any two atoms in a given structure. For an atomic fingerprint, this is achieved by minimizing the Euclidean distance over all possible permutations P that matches atom k in a with atom $P(k)$ in b:

$$d(p, q) = \min_P \sqrt{\sum_k^{N_{\text{at}}} |\mathbf{V}_k^a - \mathbf{V}_{P(k)}^b|^2}, \tag{4}$$

The Hungarian algorithm (Kuhn 1955) finds the optimal permutation within an $\mathcal{O}(n^3)$ running time, which needs as its input the cost matrix $C_{k,l}$, given by

$$C_{k,l} = |\mathbf{V}_k^a - \mathbf{V}_l^b|^2$$

2.5   Influence of Seed Structures

Since the MHM walker hops between local minima on the PES sequentially, the funnel in the vicinity of the initial seed structure is explored first. If this funnel contains the ground state, or in the case of a single-funnel system, the global minimum is found quickly within a few MHM steps, independent of the choice of the initial structure. Most realistic systems are however more complex and consist of multiple funnels separated by high barriers, which might take many MHM steps to overcome. A MHM simulation is least efficient if the initial structure is in a funnel which does not contain the global minimum, and the walker spends a long time exploring the wrong region on the PES. Since it is a priori neither known which structure belongs to which funnel nor which funnel contains the global minimum, the efficiency of the MHM strongly relies on the choice of the starting point of the walker.

The efficiency and success rate of the MHM for a multifunnel system increases if multiple MHM walkers are spawned in parallel, which simultaneously sample distinct regions of the PES. The set of these structural seeds must be carefully selected to be as diverse as possible. Using the metric defined in Sect. 2.4, the diversity in a set with N structures can be measured by the sum of inter-structural distances $D = \sum_{i,j}^N d(i, j)$. By selecting only seeds from a pool of structures that give high values of D will ensure that the PES is sampled as uniformly as possible, thereby increasing the chance of quickly finding the global minimum.

The most straightforward approach of generating initial structures, namely, randomly distributing the atoms into the simulation cell, often produces input structures far away from the ground state. Taking into account chemical intuition and considering simple geometrical constraints can significantly improve the quality of the structures (Pickard and Needs 2011). Imposing minimal pairwise distances between atoms avoids strong repulsive forces during the initial geometry relaxation. It is also beneficial to homogeneously distribute the atomic species in a multicomponent system or to use predefined building blocks if the chemistry and geometry of preferred bonding is already a priori known. On top of these considerations, symmetry constraints further improve the diversity of the structures. For molecules and clusters, the structures are randomly generated in one of the axial or polyhedral point groups. Periodic systems on the other hand adopt structures

Table 1 The impact of seeding the MHM simulations with symmetrized input structure is shown for a $A_{16}B_8$ Lennard-Jones mixture, parametrized with $\sigma_{AA} = 1.50$, $\sigma_{AB} = 2.25$, $\sigma_{AA} = 3.00$, and $\varepsilon_{AA} = 1.00$, $\varepsilon_{AB} = 1.25$, $\varepsilon_{AA} = 1.00$. All values are statistical averages gathered from 100 distinct MHM runs for each setting. For every method used to generate the input seed structures, column 2 shows the average number of visited minima $\bar{n}_{\mathrm{min}}$ before reaching the global minimum. The third and fourth columns contain the average number of MHM steps before success if $m = 4$ and $m = 10$ MHM walkers are launched in parallel, respectively.

Input seed	$\bar{n}$	$\bar{n}_{\mathrm{par}}^4$	$\bar{n}_{\mathrm{par}}^{10}$
Random	1293.84	1053.85	932.92
Symmetrized, 1 subcell	632.03	176.30	12.54
Symmetrized, 2 subcells	302.10	21.33	11.75
Symmetrized, 4 subcells	427.26	37.47	12.34

according to one of the 230 space groups. For both clusters and crystals, atoms can be either placed in a general position, where it will spawn equivalent atoms for every symmetry operation of the given point or space group, or on high-symmetry points, which remain invariant under at least one symmetry operation (aside from identity). Particularly for large periodic systems, it is beneficial to first randomly generate subcells of smaller size and use them as building blocks to generate a supercell with the target number of atoms (Lyakhov et al. 2010).

Table 1 shows how the choice of the initial seed structures influences the efficiency of the MHM runs for a binary $A_{16}B_8$ Lennard-Jones benchmark system in a periodic cell. The first column indicates the method used to generate 100 different input structures each, and the performance data in the remaining columns represent corresponding statistical averages. In some cases, the initial randomly generated structure relaxes directly into the ground state. In fact, the number of all such "successful random guesses" corresponds to the success rate of random search algorithms (Pickard and Needs 2011), which is here zero for completely random input structures, but increases to slightly above 20% for symmetrized cells. This improvement in the quality of the input structures is also reflected in the efficiency of the MHM, which increases by a factor of 2–4 as shown in the second column of Table 1. The columns three and four show the number of MHM steps required if m individual MHM runs had been blocked together and were run in parallel. All processes i in each block had been stopped as soon as one of the processes found the ground state structure after l_i steps, hence requiring a total of $n_{\mathrm{par}}^m = l_i m$ MHM steps. Here, the efficiency $\bar{n}_{\mathrm{par}}^m$ for such a parallel executions was computed by blocking the MHM runs randomly and averaging over all blocks. The data shows that combining symmetrized input structures with parallelizing the MHM runs over $m = 4$ and $m = 10$ walkers significantly improves the efficiency. Essentially, using parallel runs is equivalent to the approach used in evolutionary algorithms, where the number m corresponds to the population size in every generation l (Oganov 2010). For example, $\bar{n}_{\mathrm{par}}^{10} = 11.75$ means that the ground state structure is found on average within 1.175 generations with a population size of 10. Using a combination of symmetric input structures and parallelization, a speedup of roughly two orders of magnitude is achieved compared to a serial run with completely random input seeds.

2.6   Finding Chemical Reaction Pathways

The sole knowledge of the local minima on the PES only allows to assess the thermodynamics of a system, but at finite temperatures, kinetic effects govern processes in protein folding, catalysis, nucleation, phase transition, and many more. The study of such chemical reactions and transformations between stable and metastable states requires the knowledge of additional features of the PES (Wales 2003; Henkelman 2017). According to the transition state theory (TST), the properties of the transition state surface between pairs of local minima fundamentally determine the reaction rates. Within the harmonic TST, the rate-determining terms are approximated through the free energies of the first-order saddle points between neighboring local minima, which correspond to the reaction channels. Once these data are available in a so-called stationary point database, methods like the master equation approach or kinetic Monte Carlo are available to map out the dynamic properties of a system.

The physical moves used in the MHM are well suited to construct such stationary point databases (Schaefer et al. 2014a). Within every MHM step out of a local minimum, the system crosses over one or multiple energy barriers which essentially connect the initial and final local minima. Since the MHM trajectory is usually short, the structures of the two minima are geometrically similar and well aligned with each other. Hence, in the minima hopping guided path search approach, these structures are used as end-points for a two-ended or a recursive single-ended saddle point search, directly after each MHM step. By keeping track of both the minima and the saddle points during an MHM run, a stationary point database is constructed on the fly.

To illustrate the efficiency of this method, the minima hopping guided path search was recently applied to the 75-atom and 102-atom Lennard-Jones clusters (Schaefer et al. 2014a). Both of them are multifunnel systems, which are commonly used to benchmark structure search algorithms. In the case of the 75-atom system, new reaction pathways between the two lowest energy local minima were identified with lower rate-determining barriers and shorter reaction path lengths compared to previous reports. For the 102-atom system, shorter path lengths and a new, low-energy funnel were found. The minima hopping guided path search is a promising method to study complex atomistic reaction mechanisms in condensed matter systems and catalytic reactions (Schaefer et al. 2014b).

3   Applications in Materials Discovery

One of the major advantages of the MHM is that it can be easily adapted to study many different systems. Very early MHM structural searches were conducted on silicon clusters, using DFT to model the PES (Goedecker et al. 2005). It was shown that many local minima exist in a very small energy window above the lowest energy structure, making it difficult to conclusively identify the ground state structure at finite temperatures. A similar behavior was observed for neutral and

charged Au26$^{(-)}$ nano clusters, where several competing structural motifs were found with an energetic ordering that strongly depends on the employed DFT exchange-correlation functional (Schaefer et al. 2014b). Photoelectron spectroscopy measurements confirm the theoretical prediction and the presence of multiple isomers with low-symmetry core-shell-type structures.

The MHM has also been successfully employed to study bulk structures in various materials. One of the earliest applications in a periodic system was the prediction of a novel high-pressure phase of carbon, which could readily explain the experimentally observed phase transition in cold-compressed graphite (Amsler et al. 2012b). Our MHM simulations also revealed that a low-energy phase emerges in the LiAlH$_4$ hydrogen storage material with a polymeric structural motif (Amsler et al. 2012a; Huan et al. 2013), a prediction that was subsequently confirmed in experiments (Huang et al. 2015). Other systems that have been studied with the MHM include materials for energy harvesting and conversion, e.g., in thermoelectric (Amsler et al. 2016a, b) and photovoltaic (Botti et al. 2012; Amsler et al. 2015) applications, and in superconducting materials (Flores-Livas et al. 2012, 2016; Amsler et al. 2017a; Amsler and Wolverton 2017). Modified versions of the algorithm have also been applied to investigate the reconstructions of surfaces and the atomic structures at the solid-liquid interface of materials in an aqueous environment (Amsler et al. 2013; Fisicaro et al. 2017).

Since an extensive discussion of all relevant MHM applications is beyond the scope of this chapter, two specific materials discovery efforts are highlighted in detail below.

3.1 High-Pressure Materials Discovery

With the recent advances in high-pressure techniques, static pressures up to ≈ 20 GPa in large-volume and multi-anvil presses, and up to ≈ 400 GPa in diamond anvil cells (DAC) can be meanwhile readily reached. At such pressures, exotic chemical bonds between various elements emerge and new materials are formed, some of which can be recovered and survive as metastable phases at ambient conditions. Many of these high-pressure compounds exhibit exciting physical properties and have therefore attracted significant attention in the search for new superconducting, magnetic, superhard, and other exotic materials (Zhang et al. 2017). A major challenge in high-pressure synthesis remains the accurate in situ characterizations of a material. Although the diamonds in a DAC allow the transmission of X-rays and other radiation to probe the sample, the direct measurement of many physical properties is difficult. This poses significant limitations especially for samples that cannot be recovered for detailed examination outside of the compression chamber. Structure prediction together with quantum mechanical calculations is therefore especially valuable to assist experimental efforts both in the search and characterization of high-pressure materials.

The MHM determines the optimal structure at a fixed composition, and modifications to the algorithm are required to investigate multicomponent systems. In principle, the ground state configuration at every composition must be identified, and only phases with formation energies lower than those of any combination of competing phases are considered thermodynamically stable and constitute the so-called convex hull of stability. Such multicomponent searches are especially important for high-pressure systems, where new compounds often emerge with unexpected stoichiometries.

Adding dimensions in the compositional space however rapidly increases the complexity of the search problem. For a binary system of size N, the number of possible compositions is roughly $\frac{1}{2}\frac{N!}{2!}$. Furthermore, at a fixed composition in a binary system of $n_A + n_B = N$ atoms, there are $\frac{N!}{n_A! n_B!}$ arrangement of atoms in a single structural geometry, so-called homotops (ignoring symmetry-equivalent structures) (Ferrando et al. 2008). Hence, an accurate prediction of the ground state structures as a function of composition and pressure even for binary or ternary systems from ab initio calculations is challenging. In practice, the number of explored compositions is limited by cells containing a few formula units, and system sizes below a threshold value, commonly in the range of a few dozens of atoms. The compositions are either selected based on chemical intuition or on the fly using a genetic algorithm (Cerqueira et al. 2015).

A good example for a high-pressure materials discovery effort with the MHM is the search for novel intermetallics in binary systems. Commonly, high-pressure experiments and structural searches are conducted with constituent elements that are known to form some compound at ambient pressures or at least exhibit known chemical bonds in the solid state. These constraints limit the risk of elemental decomposition and increase the chance of forming new phases upon compression. Materials discovery however becomes significantly more challenging when dealing with elements that do not form any compounds at ambient pressures. An example of such an ambient-immiscible system is Fe–Bi, where the zero-pressure phase diagram shows essentially no solubility of either solid-state element in the other. The successful prediction and synthesis of a Fe–Bi compound is especially attractive since many bismuthides exhibit compelling physical properties, including superconductivity, strong magnetism, and promising thermoelectric performance.

We recently performed a multicomponent MHM search for the Fe–Bi system with up to 4 f.u. per cell at the compositions shown by the gray dashed lines in Fig. 4 (Amsler et al. 2017b). An initial pre-screening of a few compositions had shown that Bi-rich compounds are overall preferred, which were subsequently more extensively sampled. The colored solid bars represent stable phases that lie on the convex hull of stability, which includes various elemental allotropes of Bi and Fe. Two binary phases at the compositions $FeBi_3$ and $FeBi_2$ with *Cmcm* and *I4/mcm* symmetry, respectively, become stable in a pressure range above 35 GPa. While $FeBi_3$ only exhibits a very small stability window between 36.1 and 39.2 GPa, the $FeBi_2$ phase is thermodynamically stable from 37.5 up to at least 100 GPa. Based on this extended pressure range of stability, it is not surprising that so
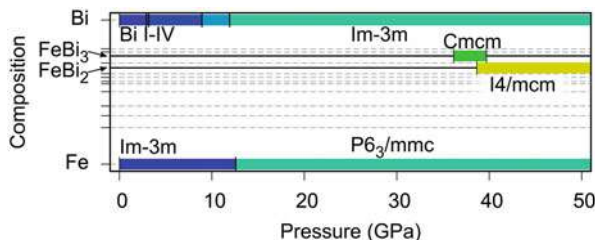
Fig. 4 Stability ranges of the predicted high-pressure phases in the Fe–Bi system. The gray dashed lines indicate the compositions which were sampled during the structural search, while all thermodynamically stable phases are represented by colored bars (the colors reflect the onset pressure of stability). The four elemental high-pressure phases of Bi between 0 and ≈ 14 GPa are collectively labeled with Bi I-IV

far only the latter phase has been confirmed through high-pressure synthesis in a DAC (Walsh et al. 2016). According to DFT calculations, the ferromagnetic ordering in $FeBi_2$ is suppressed with increasing pressure, which could potentially lead to a (unconventional) superconducting behavior.

3.2 Constrained Structural Searches

By modifying the target optimization function through imposing constraints, one can bias the structural search toward a specific region of the configurational space. Since the hops on the PES are performed based on physical MD moves and local geometry relaxations, such constraints can be easily implemented into the MHM if the derivatives of the penalty function with respect to the atomic and cell variables are available. For example, the bias toward structures around a given target volume V_0 can be included by simply adding a quadratic term $E_{V_0} = k(V - V_0)^2$ to the potential energy E. The additional forces on the cell vectors are given by the derivatives of this penalty term with respect to the cell components, $f_{\alpha\beta} = -\frac{\partial E_{V_0}}{\partial h_{\alpha\beta}} = 2kV(V_0 - V)h_{\beta\alpha}^{-1}$. Such volume constraints can be used to search specifically for low-density phases or strained materials (Amsler and Goedecker 2010).

With a similar approach, the target function can be modified to search for two-dimensional (2D) structures. A 2D confining potential centered at e^z along a nonperiodic direction (here, z-direction) is imposed, given by a sum of atomic contributions, $C = \sum_i c(e, r_i^z)$. Here, every atom i at the Cartesian coordinate $\mathbf{r}_i$ experiences a force $\mathbf{f}_i = -\frac{\partial c_i}{\partial \mathbf{r}_i}$, where $c(e^z, \mathbf{r}_i^z) = A(|e^z - \mathbf{r}_i^z| - r_c)^n$ is a polynomial function of order n beyond a cutoff distance r_c around the equilibrium position e^z. Hence, all atoms are confined within a layer of roughly the width $2r_c$, allowing the search for low-energy 2D structures. Furthermore, a search for multilayer materials or 2D heterostructures is possible by including several potentials centered at different equilibrium coordinates.
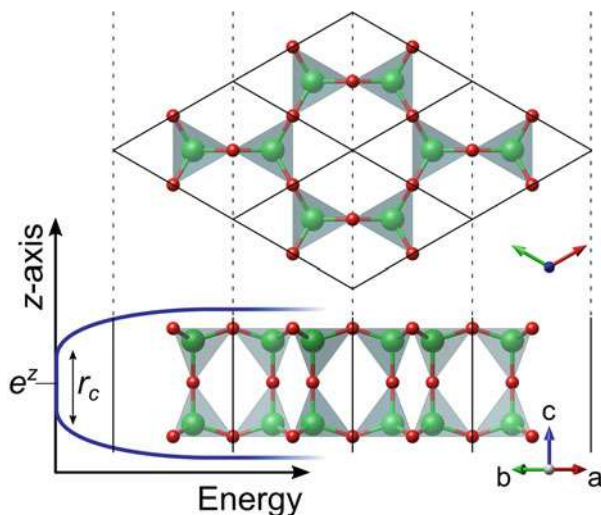
Fig. 5 The structure of the TiO_2 hexagonal nanosheet along the top and side view. The red small and green large spheres denote the O and Ti atoms, respectively. The schematic illustration at the bottom left shows how the confinement potential biases the MHM search for 2D structures

As an example of such a constrained search, we recently performed predictive MHM simulations to identify low-energy structures of layered TiO_2 (Eivari et al. 2017). The PES was pre-screened with an ANN potential based on a charge equilibration process, which was initially trained on reference DFT data (Faraji et al. 2017; Ghasemi et al. 2015). After refining the most promising candidates at the ab initio level, we identified a new putative ground state for quasi two-dimensional TiO_2 with a honeycomb lattice. Fig. 5 shows the structure of this hexagonal nanosheet (HNS), together with a schematic illustration of the confinement potential along the out-of-plane direction of the HNS. According to DFT predictions, the HNS is mechanically stable, chemically inert, and exhibits an exceptionally large band gap that can be tuned through strain and defect engineering.

4   Conclusions

Advanced global optimization algorithms like the MHM meanwhile allow the reliable prediction of structures in condensed matter system, given only the chemical composition and the thermodynamic conditions. With increasing computational resources, systems with larger number of degrees of freedom and with higher complexity become predictable. The MHM is a versatile tool that can be adapted and extended to tackle different materials prediction challenges. Using the MHM to sample transition states on the PES in addition to local minima, both the thermodynamic and kinetic properties of a system can be studied in detail.

References

Amsler M, Goedecker S (2010) Crystal structure prediction using the minima hopping method. J Chem Phys 133(22):224104

Amsler M, Wolverton C (2017) Dense superconducting phases of copper-bismuth at high pressure. Phys Rev Mater 1(3):031801

Amsler M, Flores-Livas JA, Huan TD, Botti S, Marques MAL, Goedecker S (2012a) Novel structural motifs in low energy phases of LiAlH$_4$. Phys Rev Lett 108(20):205505

Amsler M, Flores-Livas JA, Lehtovaara L, Balima F, Ghasemi SA, Machon D, Pailhs S, Willand A, Caliste D, Botti S, San Miguel A, Goedecker S, Marques MAL (2012b) Crystal structure of cold compressed graphite. Phys Rev Lett 108(6):065501

Amsler M, Botti S, Marques MAL, Goedecker S (2013) Conducting boron sheets formed by the reconstruction of the α-boron (111) surface. Phys Rev Lett 111(13):136101

Amsler M, Botti S, Marques MAL, Lenosky TJ, Goedecker S (2015) Low-density silicon allotropes for photovoltaic applications. Phys Rev B 92(1):014101

Amsler M, Goedecker S, Zeier WG, Snyder GJ, Wolverton C, Chaput L (2016a) ZnSb polymorphs with improved thermoelectric properties. Chem Mater 28(9):2912–2920

Amsler M, He J, Xia Y, Naghavi SS, Hegde VI, Hao S, Goedecker S, Ozolinš V, Wolverton C (2016b) Ultralow thermal conductivity in full Heusler semiconductors. Phys Rev Lett 117(4):046602

Amsler M, Clarke SM, Walsh JPS, Yu T, Wang Y, Meng Y, Jacobsen SD, Wolverton C, Freedman DE (2017a) Creating binary Cu-Bi compounds via high-pressure synthesis: a combined experimental and theoretical study. Chem Mater 29(12):5276–5285

Amsler M, Naghavi SS, Wolverton C (2017b) Prediction of superconducting iron–bismuth intermetallic compounds at high pressure. Chem Sci 8(3):2226–2234

Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 98(14):146401

Botti S, Flores-Livas JA, Amsler M, Goedecker S, Marques MAL (2012) Low-energy silicon allotropes with strong absorption in the visible for photovoltaic applications. Phys Rev B 86(12):121204

Buckingham RA (1938) The classical equation of state of gaseous helium, neon and argon. Proc R Soc Lond A 168(933):264–283

Cerqueira TFT, Sarmiento-Prez R, Amsler M, Nogueira F, Botti S, Marques MAL (2015) Materials design on-the-fly. J Chem Theory Comput 11(8):3955–3960

Chong EKP, Zak SH (2001) An introduction to optimization, 2nd edn. Wiley, New York

Daw MS, Baskes MI (1984) Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B 29(12):6443–6453

Dick BG, Overhauser AW (1958) Theory of the dielectric constants of alkali halide crystals. Phys Rev 112(1):90–103

Doye JPK, Wales DJ (1998) Thermodynamics of global optimization. Phys Rev Lett 80(7):1357–1360

Doye JPK, Wales DJ, Miller MA (1998) Thermodynamics and the global optimization of Lennard-Jones clusters. J Chem Phys 109(19):8143–8153

Eivari HA, Ghasemi SA, Tahmasbi H, Rostami S, Faraji S, Rasoulkhani R, Goedecker S, Amsler M (2017) Two-dimensional hexagonal sheet of TiO$_2$. Chem Mater 29(20):8594–8603

Faraji S, Ghasemi SA, Rostami S, Rasoulkhani R, Schaefer B, Goedecker S, Amsler M (2017) High accuracy and transferability of a neural network potential through charge equilibration for calcium fluoride. Phys Rev B 95(10):104105

Ferrando R, Fortunelli A, Johnston RL (2008) Searching for the optimum structures of alloy nanoclusters. Phys Chem Chem Phys 10(5):640–649

Fisicaro G, Sicher M, Amsler M, Saha S, Genovese L, Goedecker S (2017) Surface reconstruction of fluorites in vacuum and aqueous environment. Phys Rev Mater 1(3):033609

Flores-Livas JA, Amsler M, Lenosky TJ, Lehtovaara L, Botti S, Marques MAL, Goedecker S (2012) High-pressure structures of disilane and their superconducting properties. Phys Rev Lett 108(11):117004

Flores-Livas JA, Amsler M, Heil C, Sanna A, Boeri L, Profeta G, Wolverton C, Goedecker S, Gross EKU (2016) Superconductivity in metastable phases of phosphorus-hydride compounds under high pressure. Phys Rev B 93(2):020508

Ghasemi SA, Amsler M, Hennig RG, Roy S, Goedecker S, Lenosky TJ, Umrigar CJ, Genovese L, Morishita T, Nishio K (2010) Energy landscape of silicon systems and its description by force fields, tight binding schemes, density functional methods, and quantum Monte Carlo methods. Phys Rev B 81(21):214107

Ghasemi SA, Hofstetter A, Saha S, Goedecker S (2015) Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. Phys Rev B 92(4):045131

Goedecker S (2004) Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. J Chem Phys 120(21):9911

Goedecker S, Hellmann W, Lenosky T (2005) Global minimum determination of the Born-Oppenheimer surface within density functional theory. Phys Rev Lett 95(5): 055501

Goringe CM, Bowler DR, Hernández E (1997) Tight-binding modelling of materials. Rep Prog Phys 60(12):1447

Henkelman G (2017) Atomistic simulations of activated processes in materials. Annu Rev Mater Res 47(1):199–216

Henkelman G, Jónsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111(15):7010–7022

Huan TD, Amsler M, Marques MAL, Botti S, Willand A, Goedecker S (2013) Low-energy polymeric phases of alanates. Phys Rev Lett 110(13):135502

Huang X, Duan D, Li X, Li F, Huang Y, Wu G, Liu Y, Zhou Q, Liu B, Cui T (2015) High-pressure polymorphism as a step towards high density structures of $LiAlH_4$. Appl Phys Lett 107(4):041906

Jensen F (1998) Introduction to computational chemistry. Wiley, Chichester

Kuhn HW (1955) The hungarian method for the assignment problem. Naval Res Logist Quart 2(1–2):83–97

Lyakhov AO, Oganov AR, Valle M (2010) How to predict very large and complex crystal structures. Comput Phys Commun 181(9):1623–1632

Oganov AR (ed) (2010) Modern methods of crystal structure prediction, 1st edn. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim/Chichester

Oganov AR, Valle M (2009) How to quantify energy landscapes of solids. J Chem Phys 130(10):104504

Papaconstantopoulos DA, Mehl MJ (2003) The Slater-Koster tight-binding method: a computationally efficient and accurate approach. J Phys Condens Matter 15(10):R413

Parr RG, Weitao Y (1994) Density-functional theory of atoms and molecules. Oxford University Press, New York

Parrinello M, Rahman A (1980) Crystal structure and pair potentials: a molecular-dynamics study. Phys Rev Lett 45(14):1196

Pickard CJ, Needs RJ (2011) Ab initio random structure searching. J Phys Condens Matter 23(5):053201

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1992) Numerical recipes in FORTRAN: the art of scientific computing, 2nd edn. Cambridge University Press, New York

Roy S, Goedecker S, Hellmann V (2008) Bell-Evans-Polanyi principle for molecular dynamics trajectories and its implications for global optimization. Phys Rev E 77(5):056707

Sadeghi A, Ghasemi SA, Schaefer B, Mohr S, Lill MA, Goedecker S (2013) Metrics for measuring distances in configuration spaces. J Chem Phys 139(18):184118

Salamon P, Sibani P, Frost R (2002) Facts, conjectures, and improvements for simulated annealing. Mathematical modeling and computation. Society for Industrial and Applied Mathematics, Philadelphia

Schaefer B, Mohr S, Amsler M, Goedecker S (2014a) Minima hopping guided path search: an efficient method for finding complex chemical reaction pathways. J Chem Phys 140(21):214102

Schaefer B, Pal R, Khetrapal NS, Amsler M, Sadeghi A, Blum V, Zeng XC, Goedecker S, Wang LS (2014b) Isomerism and structural fluxionality in the Au_{26} and Au_{26}^- nanoclusters. ACS Nano 8(7):7413–7422

Schönborn SE, Goedecker S, Roy S, Oganov AR (2009) The performance of minima hopping and evolutionary algorithms for cluster structure prediction. J Chem Phys 130(14):144108–144109

Sicher M, Mohr S, Goedecker S (2011) Efficient moves for global geometry optimization methods and their application to binary systems. J Chem Phys 134(4):044106

Stillinger FH (1999) Exponential multiplicity of inherent structures. Phys Rev E 59(1):48–51

Tosi MP, Fumi FG (1964) Ionic sizes and Born repulsive parameters in the NaCl-type alkali halides-II: the generalized Huggins-Mayer form. J Phys Chem Solids 25(1):45–52

Verlet L (1967) Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev 159(1):98–103

Wales D (2003) Energy landscapes: applications to clusters, biomolecules and glasses. Cambridge University Press, Cambridge

Wales DJ, Doye JPK (1997) Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J Phys Chem A 101(28):5111–5116

Walsh JPS, Clarke SM, Meng Y, Jacobsen SD, Freedman DE (2016) Discovery of $FeBi_2$. ACS Cent Sci 2(11):867–871

Wang Y, Lv J, Zhu L, Ma Y (2012) CALYPSO: a method for crystal structure prediction. Comput Phys Commun 183(1):2063–2070

Wang Y, Lv J, Zhu L, Lu S, Yin K, Li Q, Wang H, Zhang L, Ma Y (2015) Materials discovery via calypso methodology. J Phys Condens Matter 27(20):203203

Wentzcovitch RM (1991) Invariant molecular-dynamics approach to structural phase transitions. Phys Rev B 44(5):2358–2361

Zhang L, Wang Y, Lv J, Ma Y (2017) Materials discovery at high pressures. Nat Rev Mater 2:17005

Zhu L, Amsler M, Fuhrer T, Schaefer B, Faraji S, Rostami S, Ghasemi SA, Sadeghi A, Grauzinyte M, Wolverton C, Goedecker S (2016) A fingerprint based metric for measuring similarities of crystalline structures. J Chem Phys 144(3):034203

Stochastic Surface Walking Method and Applications to Real Materials

Cheng Shang and Zhi-Pan Liu

Contents

Abstract

The macroscopic properties of a material are generally governed by its potential energy surface (PES) that determines not only thermodynamics but also kinetics. A thoroughly search of the global PES of material, however, has been a great challenge in theory. Three major hurdles, namely, the high energy barrier in material transformation, the large entropy due to huge structural configurations, and the large atomic degrees of freedom, are often encountered simultaneously in computational simulation of material. Owing to these, the limitation in the timescale of current simulations restricts heavily theoreticians to address many important questions in material science. In this chapter, we introduce a

C. Shang · Z.-P. Liu (✉)
Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science (Ministry of Education), Department of Chemistry, Fudan University, Shanghai, China
e-mail: zpliu@fudan.edu.cn

newly developed theoretical method, stochastic surface walking (SSW) method targeting for both global PES exploration and reaction pathway sampling. The SSW PES sampling is automated, unbiased, and taking into account the second derivative information. The algorithm of SSW is summarized here in detail, focusing on its mechanism to follow low energy pathways while being able to overcome high barriers. SSW simulation has recently been applied to different areas in material and reaction systems. Several typical examples of PES exploration by combining SSW with first principles and neural network potential calculations are presented to illustrate the power of SSW for unbiased PES exploration and pathway searching.

1   Introduction

The structure prediction and the reaction pathway search are two important themes in modern computational simulation. While both of them are closely related to the exploration of the potential energy surface (PES), the former focuses on the thermodynamics properties and the latter aims to provide kinetics quantities on material transformation. Since 1990s, there have been a rapid development on the methodology for structure search and a group of successful methods are now available for identifying the most stable minimum on PES, the global minimum (GM) (Wales and Doye 1997; Martonak et al. 2006; Oganov and Glass 2006; Amsler and Goedecker 2010; Wang et al. 2012) of different systems, including molecules, clusters, and solid crystals. On the other hand, the pathway sampling remains challenging in computational approach even for systems with only a few atoms (Wales 2002; Carr et al. 2005), because of the lack of efficient methods to identify the reaction coordinate (Oganov et al. 2011). To find the minimum energy pathway between minima on PES must rank one of the top challenges in PES exploration.

To search the GM on PES, a typical measure for the efficiency of global optimization method, the current algorithms for structure search generally involve aggressive structure changes, which are required to bypass the transition region of PES and prevent long-time trapping in local minimum. For example, the basin hopping (Wales and Doye 1997) method developed by Wales in 1990s starts from one single entry (one structure) on PES and the structure is perturbed by a random hopping with a magnitude of ~ 0.4 Å for each degree of freedom. The PES seen by basin hopping is therefore a transformed PES without saddle regions. Similar to basin hopping, simulate annealing (Kirkpatrick et al. 1983; Pannetier et al. 1990; Schön and Jansen 1996, 2001) and minimum hopping (Amsler and Goedecker 2010) also start from one single entry but utilize molecular dynamics (via heating to high temperature) to overcome the barrier on PES. On the other hand, the methods starting from multiple entries (a group of different initial structures on PES) developed in 2000s, such as evolutionary algorithm (EA) as represented by USPEX (Oganov and Glass 2006), XtalOpt (Lonie and Zurek 2011) and GSGO (Trimarchi and Zunger 2007), and the particle-swarm-optimization (PSO) algorithm

as represented by CALYPSO (Wang et al. 2012), are more popular in predicting unknown crystal structures of material. The symmetry adaptation in structure generation is one of the key techniques to identify rapidly the stable, ordered structures (Wang et al. 2012). Overall, the structure search methods do not yield a continuous trajectory due to the large change between structures, and thus, it is generally not feasible to obtain the low energy pathways.

The enhanced sampling methods (Siepmann and Frenkel 1992; Voter 1997; Rosso et al. 2002; Woods et al. 2003; Maragliano and Vanden-Eijnden 2006), such as metadynamics using the bias potential constraints (Huber et al. 1994; Laio and Parrinello 2002; Raiteri et al. 2005; Martonak et al. 2006) and adiabatic free energy dynamics (AFED) with temperature acceleration (Yu and Tuckerman 2011), are able to generate continuous trajectory during PES search as intrinsic to MD simulation. Thus, this class of methods has been utilized to identify low energy pathways in crystal phase transition. For example, by imposing the collective variable on the lattice degrees of freedom, metadynamics has been utilized to simulate the phase transition from graphite to diamond (Khaliullin et al. 2011). In general, it is a general concern for MD-based techniques to capture the rare event. Since the occurrence of a reaction event depends exponentially on the reaction barrier, the correct description of the reaction coordinate (e.g., collective variable in metadynamics) and the choice of the simulation temperature are key parameters for producing reactions in simulation. Once the reaction occurs in MD, it is then possible to glean the initial state and the final state from MD trajectory and use the double ended transition state (TS) search method to search for the saddle point. Typical choice for finding the TS in crystal phase transition is the generalized solid-state nudged elastic band method (G-SSNEB) (Sheppard et al. 2012) and variable-cell double-ended surface walking method (VC-DESW) (Zhang and Liu 2015a).

In 2013, Liu's group developed a new PES searching method, namely, the stochastic surface walking (SSW) (Shang and Liu 2013) method. It was initially designed for aperiodic systems, i.e., clusters and molecules, and then extended to periodic solid systems (Shang et al. 2014) in 2015. Unlike the traditional methods for the global PES exploration, the searching in SSW is guided by soft mode directions as computed from the second derivatives and utilizes gentle structure perturbation followed by local optimization to follow low energy pathways. The method can thus be applied to both structure prediction and reaction pathway search (Zhang and Liu 2015b). In this chapter, the algorithm of SSW will be described in detail and some recent applications in material discovery and solid phase transition are then presented.

2   Stochastic Surface Walking (SSW) Method

The central idea of SSW method comes from the bias-potential-driven dynamics (Iannuzzi et al. 2003) and Metropolis Monte Carlo (MC) (Metropolis et al. 1953). It manipulates smoothly the structural configuration from one minimum to another on PES and relies on Metropolis MC at a given temperature to decide the acceptance

of the move. A series of consecutive minimum structures, both local and global, is generated in SSW simulation forming a continuous trajectory, which can provide the key information on the reaction pathways between minima. By applying SSW to a number of systems, Shang and Liu (2013) have shown that the SSW can explore blindly complex PES with high efficiency as demonstrated in Lennard-Jones (e.g., LJ_{75}) and short-range Morse particles (M_{80}).

Each step in SSW, also termed as a MC step, comprises three independent parts, namely, (i) the climbing, (ii) the relaxation, and (iii) the Metropolis MC. As schematically illustrated in a one-dimensional PES in Fig. 1a, such a MC step utilizes the climbing module to move up hill and the relaxation module to locate minimum. Once a minimum is reached, the Metropolis MC is used to judge whether the structure will be accepted or refused. The climbing procedure lies at the heart of the SSW method and is elaborated below in detail.
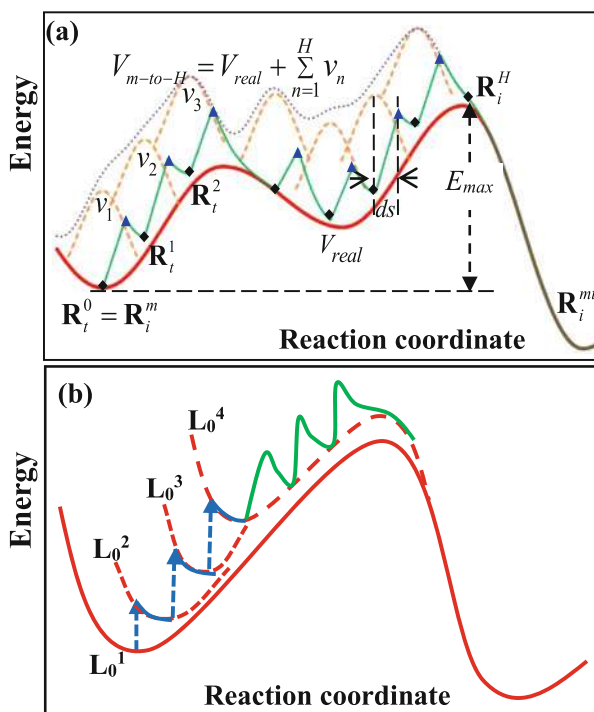
Fig. 1 The illustration of (**a**) the SSW method and (**b**) the SSW-crystal method. The real PES is described by the red curve showing separated minima. (**a**): one SSW step from one minimum to another, where the bias Gaussian potentials are imposed only on the atom degrees of freedom; (**b**): one combined lattice and atom displacement step of the SSW-crystal method. The blue arrows and curves represent the lattice displacement and the local relaxation, respectively, as governed by the CBD-cell module, which gradually lifts up the potential energy of the crystal. The green curve represents the consequent atom displacement (at the fixed lattice) as governed by the SSW-module, being a simplified view of (**a**). (Reproduced from Ref. (Shang et al. 2014) with permission from the PCCP Owner Societies)

The climbing module of the SSW involves repeated bias-potential-driven structure extrapolation and local geometry optimization, which drags gradually $\mathbf{R}_i^0$ to a high energy configuration $\mathbf{R}_i^H$, where "i" is the index of the current MC step (Fig. 1a). Starting from the current minimum $\mathbf{R}_i^0$, SSW first generates a random direction $\mathbf{N}_i^0$, a normalized vector pointing to a direction to change the current geometry.

To enable an unbiased exploration of PES, the initial direction $\mathbf{N}_i^0$ by design combines two randomly generated vectors, the so-called global random mode $\mathbf{N}_i^g$ and the local random mode $\mathbf{N}_i^l$, as Eq. (1).

$$\mathbf{N}_i^0 = \left(\mathbf{N}_i^g + \lambda \mathbf{N}_i^l \right) / \left| \mathbf{N}_i^g + \lambda \mathbf{N}_i^l \right| \tag{1}$$

where the mixing parameter λ controls the relative portion of these two displacement directions. Specifically, $\mathbf{N}_i^g$ in our implementation is set as a randomly generated normalized vector that obeys the Maxwell-Boltzmann velocity distribution at 300 K, as utilized in standard molecular dynamics method for generating initial random velocity. As $\mathbf{N}_i^g$ distributes over a group of atoms, it represents a gentle, global displacement pattern of atoms. By contrast, $\mathbf{N}_i^l$ describes a stiff, local atomic move, and in our implementation it is set as a collision movement between two distant atoms. For example, the $\mathbf{N}_i^l$ associated with an atom A (e.g., the first atom in system) and an atom B (the second atom in system) can be derived as Eq. (2), using their coordinates q_A and q_B. The atom pair A and B can be either chosen randomly or learned from the previous trajectories, as long as the two atoms are not in the close contact (i.e., their distance >3 Å).

$$\mathbf{N}_i^l = \begin{pmatrix} q_B \\ q_A \\ 0 \\ \vdots \end{pmatrix} - \begin{pmatrix} q_A \\ q_B \\ 0 \\ \vdots \end{pmatrix} \tag{2}$$

$$\mathbf{R}_1 = \mathbf{R}_0 + \mathbf{N_t} \cdot \Delta R \tag{3}$$

$$C = \frac{(\mathbf{F}_0 - \mathbf{F}_1) \cdot \mathbf{N_t}}{\Delta R} \tag{4}$$

$$V_{R1} = V_{real} + V_N \tag{5}$$

$$V_N = -\frac{a}{2} \cdot \left[(\mathbf{R}_1 - \mathbf{R}_1) \cdot \mathbf{N}_i^0 \right]^2 = -\frac{a}{2} \cdot \left(\Delta R \cdot \mathbf{N_t} \cdot \mathbf{N_i^0} \right)^2 \tag{6}$$

Because reactions with low barriers generally involve soft normal mode directions, it would be ideal to soften the randomly generated $\mathbf{N}_i^0$ towards one eigenvector of Hessian matrix with small eigenvalues (not necessarily the lowest one). However, the direct computation of Hessian is expensive and not affordable in PES search. To solve this, Shang et al. develop a numerical approach, the bias-potential constrained Broyden dimer (BP-CBD) method (Shang and Liu 2012), to soften $\mathbf{N}_i^0$ using Eqs. (3), (4), (5), and (6). Following the unbiased dimer rotation (Henkelman and Jonsson 1999), Shang et al. define two images separated by a fixed distance of ΔR on PES, namely, $\mathbf{R_0}$ and $\mathbf{R_1}$ [Eq. (3)]. The rotation of the dimer by their force ($\mathbf{F_0}$ and $\mathbf{F_1}$) perpendicular to the dimer vector $\mathbf{N}$ using constrained Broyden optimization (CBD) (Shang and Liu 2010) will converge to the softest eigenvector of Hessian, and the local curvature [C in Eq. (4)] can be computed according to the finite difference equation. The softest eigenvector is, however, often not the desired one for reaction since it corresponds to the translational and rotational modes (C = 0) in the basin region. Therefore, a biased rotation scheme as implemented in BP-CBD is developed, in which the potential of $\mathbf{R_1}$ is modified as Eq. (5), where V_N is the bias potential added to the real PES V_{real} of $\mathbf{R_1}$ that is a quadratic function of coordinates $\mathbf{R_1}$ along $\mathbf{N}_i^0$ [Eq. (6)]. As long as the parameter a [Eq. (6)] is large enough, the biased rotation can guarantee the rotation of dimer will not deviate far away from $\mathbf{N}_i^0$. The force due to the bias potential in constraining the dimer rotation can be evaluated straightforwardly.

$$
V_{m-to-H} = V_{real} + \sum_{n=1}^{H} v_n = V_{real} + \sum_{n=1}^{H} w_n \times \exp\left[-\frac{\left(\mathbf{R}^t - \mathbf{R}_t^{n-1}\right)^2 \cdot \mathbf{N}_i^{n\,2}}{2 \times ds^2} \right]
$$
(7)

$$
\mathbf{F}_{tot} = \mathbf{F}_{real} + \sum_{n} w_n \cdot \exp\left[-\frac{\left(\mathbf{R}^t - \mathbf{R}_t^{n-1}\right)^2 \cdot \mathbf{N}_i^{n\,2}}{2 \times ds^2} \right] \cdot \frac{\left(\mathbf{R}^t - \mathbf{R}_t^{n-1}\right) \cdot \mathbf{N}_i^n}{2 \times ds^2} \cdot \mathbf{N}_i^n
$$
(8)

On moving from $\mathbf{R}_i^m$ to a high energy configuration $\mathbf{R}_i^H$, a modified PES V_{m-to-H} $\left(= V_{real} + \sum_{k=1}^{n} v_k\right)$ is utilized, as shown in Eq. (7), in which a series of bias Gaussian potential v_n (n is the index of the bias potential, $n = 1,2 \cdots H$) is added one by one consecutively along the direction $\mathbf{N}_i^n$ and thus creates a series of local minima $\mathbf{R}_t^n$ along the moving trajectory on the modified PES. Similar techniques have been used in Metadynamics (Iannuzzi et al. 2003). The local minima $\mathbf{R}_t^n$ is identified using local geometry optimization, where the force can be evaluated according to Eq. (8). The w and ds in Eq. (7) control the height and the width of Gaussian function v_n, respectively. While w can be computed on-the-fly to guarantee the success of the uphill, ds is left as an adjustable parameter in SSW simulation.

It should be mentioned that $\mathbf{N}_i^n$ will always be updated from the initial random direction $\mathbf{N}_i^0$ at each $\mathbf{R}_t^n$ and is then refined using BP-CBD rotation.

To sum, the uphill movement from $\mathbf{R}_t^m$ to $\mathbf{R}_t^H$ is a repeated procedure containing (i) updating the direction $\mathbf{N}_i^n$ at $\mathbf{R}_t^n$; (ii) adding a new Gaussian function v_n and displacing $\mathbf{R}_t^n$ along the direction $\mathbf{N}_i^n$ by a magnitude of ds ($\mathbf{R}_t^n + \mathbf{N}_i^n \cdot ds$); and (iii) relaxing to $\mathbf{R}_t^{n+1}$ on the modified PES.

The overall efficiency of SSW method depends on the choice of the step length of surface walking ds and the max number of Gaussian potentials H. A typical value of ds ranges from 0.2 to 0.6 Å, being 10% to 40% bond length of chemical bonding. The max number of Gaussian potentials (H) is also system-dependent and in general a number in the range of 6~15 is a sensible choice. With large ds and large H, a large scope of PES can be explored rapidly, but it is at the expense of the resolution on the reaction pathway between minima. Typically, for complex Lennard-Jones clusters (e.g., LJ$_{75}$ with multiple funnels and high barrier), the SSW simulation requires 300~400 energy/force evaluation per MC step with $ds = 0.6$ and $H = 14$. It takes typically 70~80% computational efforts (energy and gradient evaluation) on the climbing and the left 20~30% on the relaxation to a minimum.

3   SSW-Crystal Method

The SSW method can be extended to explore the PES of periodic crystal systems, named as the SSW-crystal method. Due to the additional degrees of freedom from the lattice, the key issue in the SSW-crystal method is to couple effectively the lattice and the atom movement in the solid structure displacement. In fact, there are a number of possible ways to couple the lattice and atom displacement. One simplest way is as introduced in Shang et al.'s work in 2014 (Shang et al. 2014), in which the cell movement and atom movement are operated alternatively in one MC step. A more sophisticated way is to combine the lattice parameters and atom coordinates into a generalized coordinate, based on which PES is explored. The generalized coordinate approach is similar to that implemented for TS search between two crystal phases in variable-cell double-ended surface walking method (Zhang and Liu 2015a). Below we mainly introduce the simplest way to couple the lattice and atom displacement in the original SSW-crystal method.

The SSW-crystal method is constituted by three parts, the atom displacement module using the SSW method (SSW module), the lattice displacement module using the CBD method to treat cell vector displacement (CBD-cell module), and the Metropolis MC module. The SSW module and the CBD-cell module are responsible for the displacement of atom and lattice coordinates, respectively. By coupling these two modules, the SSW-crystal can mimic the collective movement in crystal phase transition that involves generally both the atom and the lattice degrees of freedom. Shown schematically in Fig. 1b is such a combined CBD-cell (blue curve) and SSW (green curve) displacement to manipulate one crystal structure from one minimum

to another. Compared to the SSW method, only the CBD-cell module is new, and it is elaborated in the following.

The CBD-cell module to displace lattice mimics the way to displace atom in SSW method. In short, a few cycles are performed to displace smoothly the lattice along one soft lattice mode direction. It has been found that the displacement along the soft lattice modes, instead of a random direction or hard lattice modes, is critical for achieving the high efficiency of crystal PES exploration. Each CBD-cell cycle (blue arrows/curves in Fig. 1b) contains three steps, namely, the soft lattice mode identification, the lattice displacement, and the local relaxation of atoms in the fixed lattice. Specifically, the lattice displacement is achieved by displacing the current lattice $\mathbf{L}_0^n$ (n is the index of the current CBD-cell cycle) along a soft lattice mode direction $\mathbf{N_{cell}}$ to yield a new lattice $\mathbf{L}_0^{n+1}$, as described in Eq. (9). The lattice $\mathbf{L}$ is a nine dimensional (3 × 3) vector. After the lattice displacement, the local energy minimization is then performed on the fixed lattice.

$$\mathbf{L}_0^{n+1} = \mathbf{L}_0^n + \Delta L \cdot \mathbf{N_{cell}} \tag{9}$$

In the equation, ΔL represents a small displacement for the lattice, which is under the constraints for the maximum change in the lattice length (e.g., *a*, *b,* and *c* **axis**) and the volume. Typically, $\Delta L = 0.15 \times \sqrt{\sum l_{ij}^2}, l_{ij} \in L_0^n$, i.e., ΔL being about 15% change in the cell vector. It is obvious that the quality of $\mathbf{N_{cell}}$ determines largely how efficient the PES exploration of solid will be.

$$\mathbf{L}_1 = \mathbf{L}_0 + dL \cdot \mathbf{N}_{cell}^t \tag{10}$$

$$F_{rot}^L = 2\left(\dot{\mathbf{L}}_0 - \dot{\mathbf{L}}_1\right) - 2\left[\left(\dot{\mathbf{L}}_0 - \dot{\mathbf{L}}_1\right) \cdot \mathbf{N}_{cell}^t\right] \cdot \mathbf{N}_{cell}^t \tag{11}$$

$$\dot{L}_{ij} = \partial H / \partial L_{ij} = -\Omega \sum_k (\sigma_{ik} - p_{\text{ext}}) \left[\left(\mathbf{L}^t\right)^{-1}\right]_{kj} \tag{12}$$

To generate a physically meaningful $\mathbf{N_{cell}}$, Shang et al. also use CBD rotation method (Shang and Liu 2010) to identify one soft mode of the lattice Hessian, starting from a random vector $\mathbf{N}$ (a nine-component vector, three rotational modes are projected out in CBD rotation). The CBD rotation algorithm for lattice is summarized in Eqs. (10) and (11), where $\dot{L}$ denotes the first derivative of the enthalpy (H) with respect to the lattice; σ is the stress tensor, Ω is the volume, and p_{ext} is the external pressure; $\mathbf{L_1}$ and $\mathbf{L_0}$ are two lattice configurations, namely, a dimer, separated by a fixed Euclidean distance of dL along $\mathbf{N}_{cell}^t$. In CBD rotation, the dimer will rotate to minimize the rotation force ($\mathbf{F}_{rot}^L$) until a preset criterion is reached. $\mathbf{N_{cell}}$ will in principle converge to the softest eigenstate of the lattice Hessian matrix. The curvature (eigenvalue) of the $\mathbf{N_{cell}}$ can then be derived using the numerical finite difference as:

$$C_{cell} = \left(\dot{\mathbf{L}}_0 - \dot{\mathbf{L}}_1\right) \cdot \mathbf{N}_{cell}^t / dL \tag{13}$$

The CBD rotation method is able to fast identify one soft lattice mode starting from a random vector, typically within a few steps of energy/force evaluations. Taking α-quartz $(SiO_2)_{12}$ crystal as an example where the SiO_2 PES is described by the BKS (Beest, Kramer and Santen) potential (van Beest et al. 1990), Shang et al. have performed the numerical experiment to examine the efficiency of CBD for the softest lattice mode finding. The six lattice modes of quartz can be numerically calculated and the curvature C_{cell} of them are 4.98, 5.81, 8.47, 20.1, 21.7, 35.0 ($\times 10^{-2}$ a.u.). As the representatives, the calculated softest ($\mathbf{N}_{cell}^1$) and hardest ($\mathbf{N}_{cell}^6$) lattice modes are shown in Eqs. (14) and (15). Obviously, the softest lattice mode represents the expansion along c axis and the compression along a and b axes, while the hardest lattice mode represents the compression along all three directions. Averaged over 5000 independent runs initialized from random vectors, the CBD rotation prefers clearly the soft lattice modes: the probability for converging to three soft modes and to three hard modes are 57.8% and 4.7%, respectively. The rest 37.5% converges to the hybrid modes.

$$N_{cell}^1 = \begin{pmatrix} -0.16 & 0.08 & -0.03 \\ 0.52 & -0.74 & -0.02 \\ -0.04 & -0.03 & 0.38 \end{pmatrix} \tag{14}$$

$$N_{cell}^6 = \begin{pmatrix} 0.33 & 0.24 & 0.02 \\ 0.08 & 0.35 & 0.00 \\ 0.03 & 0.00 & 0.83 \end{pmatrix} \tag{15}$$

4   SSW Reaction Sampling (SSW-RS)

4.1   Reaction Sampling

Because the SSW simulation produces a continuous trajectory between minimum and each displacement follows the soft mode direction, the SSW simulation can also be utilized to identify low energy reaction pathways. The SSW reaction sampling (SSW-RS) developed in 2015 (Zhang and Liu 2015b), as illustrated in Fig. 2, aims to generate an ensemble of reactant-product (R/P) pairs between two connected minima on PES. The two structures of the R/P pair can be utilized later for TS location using the double-ended TS searching method.

As shown in Fig. 2, the SSW reaction sampling starts from one random configuration of reactant, which is the input structure and may well not be the most stable configuration of the reactant. During the simulation, the structures nearby this reactant will be visited, including the conformation isomers of the same reactant (the same phase region) and the likely products with different bonding patterns (the other phase regions). Different from the Metropolis MC scheme in the global structure search, a special structure selection module (to decide whether a new minimum is accepted or refused) is designed to judge whether a (desirable) reaction occurs

Fig. 2 The flow chart of SSW-RS method
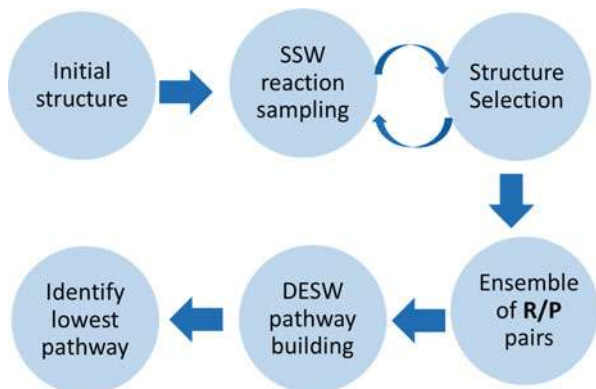
or not. If a new minimum in the other phase regions is identified, it is recorded as the product of the current SSW step. Then the program will reject the new minimum and continue the PES exploration from the reactant; on the other hand, if the new minimum identified is still at the same reactant phase, the simulation will accept the new isomeric structure and continue the structure exploration. The whole procedure will be repeated until a large number of R/P pairs are collected. Typically, a few hundreds of R/P pairs are required to reveal the lowest energy pathway of an elementary reaction step (see Fig. 2).

The probability for accepting a new minimum in the structure selection module is thus described by P_r in Eq. (16)

$$P_r = \begin{cases} 0, \text{reaction occurs} \\ 1, \text{otherwise} \end{cases} \tag{16}$$

where the occurrence of reaction is characterized by a group of preset conditions, e.g., a dramatic change of system energy, the making/breaking of chemical bonds and the change of the local chirality of atom for molecular reactions, the change of crystal symmetry, and the change of structure order parameter for crystal phase transition. These conditions need to constitute a sensitive measure for detecting the occurrence of a desirable reaction from SSW trajectory. The probability P for recording an R/P pair is given in Eqs. (17) and (18).

$$P = (1 - P_r) \cdot P_{mc} \tag{17}$$

$$P_{mc} = \begin{cases} \exp\left[\frac{E(\mathbf{R}^m) - E(\mathbf{R}^{mt})}{kT}\right], when\ E\left(\mathbf{R}^m\right) < E\left(\mathbf{R}^{mt}\right) \\ 1, otherwise \end{cases} \tag{18}$$

where P_{mc} is the Metropolis MC probability for accepting the new minimum $\mathbf{R}^{mt}$ or reverting to the starting minimum $\mathbf{R}^m$. The Metropolis NC scheme in reaction

sampling helps to screen the highly endothermic reactions, which will not be considered for subsequent pathway building.

4.2   Pathway Building

Once enough R/P pairs are collected, the low energy pathways connecting these R/P pairs can then be determined by building the pathway between them. The pathway building tool utilized in our work is the double-ended surface walking (DESW) method (Zhang et al. 2013), which can treat both aperiodic and periodic systems (Zhang and Liu 2015a). The DESW method follows the same uphill mechanism in SSW method, i.e., repeated bias potential addition and local relaxation. It can fast connect a pseudopathway between an R/P pair, which has been shown to mimic the true MEP at the TS region. The DESW pseudopathway building helps to screen unlikely pathways in an efficient way: only those with relatively low energy pseudopathway will be considered further to locate the TS explicitly. In DESW, because the reactant and the product obtained from the SSW reaction sampling may be separated by more than one minimum, the highest energy TS along the pathway will be first identified. The barrier calculated with respect to the GM energy of reactant will be utilized a quantitative measure to compare different pathways and to select finally the low energy pathways. The exact knowledge on the minimum energy pathway (MEP) at this stage is in fact not essential (MEP may be calculated only for a few lowest energy pathways that are important for kinetics).

The pathway sampling in the multiple-step reaction systems is even more challenging for the presence of multiple intermediates. Liu group (Wei et al. 2017; Zhang et al. 2017a, b) have shown recently that by carrying out the SSW-RS sampling from every likely intermediates, the whole reaction network can be established eventually. The methods, such as Dijkstra's algorithm (Dijkstra 1959), can be utilized to analyze the network and identify the reaction mechanism. If the overall rate of the reaction system is the concern, one can construct the transition rate matrix and utilize the kinetics methods (such as kinetic MC) to simulate the reaction kinetics.

While the key elements of the SSW-RS for reactivity prediction share common features with the approaches proposed in recent years (Maeda et al. 2009; Maeda et al. 2014; Schaefer et al. 2014), the SSW-RS has its own advantages in both the reaction sampling and the pathway building. Thanks to the generality of the SSW method, the SSW-RS is also a general purpose tool for sampling different type of reactions, ranging from the association and dissociation molecular reactions (Zhang and Liu 2015b), to the surface catalytic reaction (Zhang et al. 2017a) and to the solid phase transition reactions (Zhu et al. 2015). This leads to the easy implementation and high transferability of the SSW-RS method for different applications. Taking molecular reactions as example, the SSW-RS can sample simultaneously both the reactant conformations, a very flat PES with many likely conformations, and the bond making/breaking reactions that have high reaction barriers. These features of the SSW-RS method will be demonstrated in the following section.

5   Application Examples

For the distinct features of SSW method (and the derived methods), such as the evaluation of second derivatives of the PES and the gentle structure perturbation, the SSW provides a unique tool to explore unbiasedly the global PES of a material and in the meantime to identify the reaction pathways between phases of concern. Besides, the gentle movement helps greatly the electronic structure calculation, allowing one to use the one-electron wavefunctions and the charge density of the earlier configuration, enabling a long-time stable PES sampling.

In this section, we describe a few typical applications of the SSW methods, including the global PES and phase transition searching of TiO_2, ZrO_2, and carbon. All these simulations are based on density-functional theory (DFT), where the exchange-correlation functional used is the generalized gradient approximation as parameterized by Perdew, Burke, and Ernzerhof (GGA-PBE) (Perdew et al. 1996). Corrections for van-der Waals dispersion are added in the case of carbon crystal systems using the Grimme-D3 approximation (Grimme et al. 2010). The interaction between atomic cores and valence electrons is described using the projector augmented wave method (Kresse and Joubert 1999). All calculations use plane-waves basis sets. For more details, the reader is referred to the original articles cited below.

5.1   Global PES and New Crystal Phases of TiO_2

The SSW method can be utilized to establish the global PES of material and thus identify new crystal phases. This is exemplified in TiO_2 system, an important functional material widely utilized as coating materials and catalysts in many fields. In particular, the photoactivity of TiO_2 was hotly investigated for water splitting (Li et al. 2010; Li and Liu 2011; Zhao and Liu 2014; Zhao et al. 2015). The PES of TiO_2 is highly complex as evident by the fact that many possible polymorphs are present in nature (the most common phases of TiO_2 at the ambient conditions are rutile, anatase, brookite, and TiO_2-B (Penn and Banfield 1999; Bakardjieva et al. 2006; Cheng et al. 2015)). New stable TiO_2 phases have long been searched for in material science.

In 2017, Huang et al. (2017) constructed TiO_2 neural network (NN) potential using SSW global PES data generated from first principles density functional theory (DFT) calculations in small unit cells (below 12-atom). Six independent SSW global search using NN potential (SSW-NN) were then carried out in a larger supercell (48-atom), 10,000 minima visited in each run. In total, 6151 distinct minima were obtained after removing duplicated structures. Figure 3a shows the global PES contour plot projected onto the structural order parameter (OP_2) and the total energy (E). The corresponding density of states (DOS) is plotted in Fig. 3b.

The global PES shows that the crystalline region of TiO_2 is below ~0.2 eV per formula unit (eV/f.u.) with respect to the GM (TiO_2-B phase), and contains 224
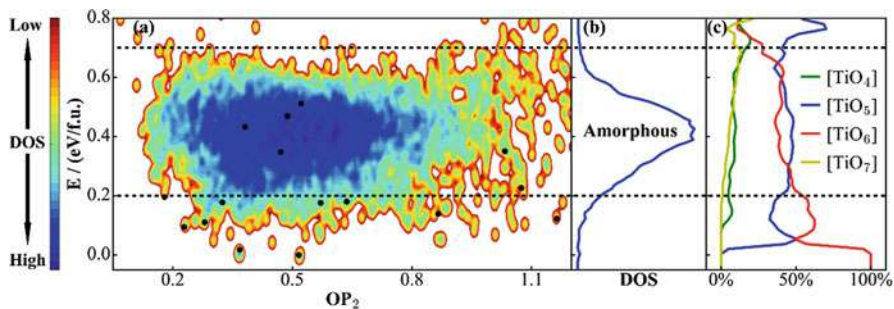
Fig. 3 (**a**) PES contour plot for TiO$_2$ distinct minima sampled from SSW-NN global search. (**b**) Density of states (DOS) plot for TiO$_2$ phases in (**a**), showing that amorphous structures appear in the 0.2~0.7 eV/f.u. window. (**c**) The percentage of differently coordinated Ti for TiO$_2$ structures. [TiO$_5$] and [TiO$_6$] are main structural features for amorphous structures. (Reproduced from Ref. (Zhang et al. 2017b) with permission from the PCCP Owner Societies)

distinct crystal phases (180 of them are newly found from SSW-NN simulation). It is noted that the number of distinct crystalline structure is substantially large considering that there are only ∼10 TiO$_2$ phases reported from experiment (Cheng et al. 2015). From 0.2 to 0.7 eV/f.u., one can see a salient deep-blue zone due to the energy degeneracy of structures, featuring a high, continuous, and broad peak in the DOS plot (Fig. 3b). This zone is dominated by amorphous structures.

The Ti coordination number x ([TiO$_x$]) for all structures is analyzed to understand the difference between amorphous and crystalline phases. The evolution of the Ti coordination in solids with the increase of energy is plotted in Fig. 3c. This is done by counting and averaging the Ti coordination for structures in the same energy interval, E to E + dE (5 meV/f.u.). Figure 3c shows that [TiO$_4$], [TiO$_5$], [TiO$_6$], and [TiO$_7$] constitute the major coordination environment for Ti. For the low energy crystal phases (e.g., rutile, anatase, TiO$_2$-B), they generally contain only TiO$_6$ octahedra (100% [TiO$_6$], energy lower than 25 meV/f.u.). The [TiO$_5$] starts to appear above 30 meV/f.u. This and [TiO$_6$] correspond to the dominant coordination for amorphous structures. By contrast, [TiO$_4$] and [TiO$_7$] are rare coordination in TiO$_2$ solids: they appear at high energies and are found either in high pressure phases ([TiO$_7$]) or in open structures ([TiO$_4$], e.g., cage, 2-D layer).

Interestingly, the SSW-NN PES exploration discovers stable porous structures, which could be potentially useful as molecular sieves and catalysts. Two such phases with open cages labeled as phase-87 (I4/m, #87) and phase-139 (I4/mmm, #139) rank 10 and 23 in the minimum energy spectrum respectively, having the similar stability as the common rutile phase that ranks 30 in the spectrum. The [TiO$_5$] coordination is the key structure feature in these porous TiO$_2$ structures, and they are found to be good candidates for anode material in Li ion battery (Ma et al. 2017). Importantly, the calculated barrier for the phase-87 to anatase transition is 0.22 eV/f.u., which is quite high considering that the barrier from TiO$_2$-B to anatase is only 0.12 eV/f.u. (Zhu et al. 2014) and the experimental temperature for TiO$_2$-B to

anatase conversion is ∼673 K (Zhang and Banfield 1999). Our results indicate that the porous phase-87 is kinetically stable. One can argue then that, once synthesized, the new phase could be stabilized at ambient conditions.

5.2　Zirconia Phase Transitions

The crystal-to-crystal phase transition has long been a great challenge for computer simulations. Recently, the SSW-RS method has been utilized to provide important insights into the phase transition mechanism and kinetics. As a prototype system for crystal phase transition, the tetragonal-to-monoclinic (t-m) phase transformation in ZrO_2 (Kelly and Rose 2002) is widely exploited in mechanical ceramic toughening (Jin 2005) and in heterogeneous catalysis for tuning surface activity (Wang, et al. 2007). This phase transition is rapid and reversible above ∼1000 °C and incurs an appreciable volume increase (by 4.2%) that is in accompany with the decrease in the coordination number of Zr from 8 to 7. Although it was long regarded as a diffusionless Martensitic phase transition (Wolten 1963; Bhattacharya et al. 2004), the transformation kinetics at the atomic level has been debated for over 60 years.

Using SSW-RS method, Guan et al. (2015) in 2015 have determined the lowest energy transition pathways and resolve the physical origin of two types of orientation relationship (OR) in ZrO_2 t-m phase transition. From 269 pathways of t-m phase transition obtained from SSW-RS, they identified three distinct lowest energy pathways. The overall potential energy profiles of these pathways are shown in Fig. 4a. Path I and Path II are pathways for t-m phase transition with the same apparent OR, i.e., $(100)_m//(110)_t$; $[001]_m//[001]_t$, being the Type-C OR in literatures. Unlike Path I, Path II is an indirect channel involving an intermediate orthorhombic phase, o-phase (Pbc2/1, #29), with OR as $(100)_m//(001)_o$; $[001]_m//[100]_o$ and $(001)_o//(110)_t$; $[100]_o//[001]_t$. The 3-D contour plot of Path II is illustrated in Fig. 4b. Path III is a branch channel involving the ferroelastic transformation of the t-phase itself by switching the **c** and **b** axis of crystal. The OR is $(110)_t//(110)_{t'}$; $[001]_t//[1\bar{1}0]_{t'}$. When Path III (t-t') is followed by Path I or II (t-m), the apparent OR of t-m phase transition is switched to $(100)_m//(110)_t$; $[010]_m//[001]_t$, which turns out to be Type-B OR in literature.

The calculated barriers of Path I and II leaving t-phase are extremely low, 2.6 and 3.5 meV/f.u. relative to t-phase, respectively, while that for Path III is 29.2 meV/f.u., being much higher than the previous two pathways. Compared to the t-phase to o-phase (t-o), the o-phase to m-phase (o-m) in Path II needs to overcome much higher barrier (32.1 meV/f.u.), which indicates the o-phase is a trapping state at low temperatures. Because o-phase is only 0.60% larger than t-phase in volume, the transformation toughening is largely prevented as the o-phase forms at low temperatures. Although the volume of o-phase is close to t-phase, the Zr in o-phase is 7 coordinated; half of the O atoms are 4 coordinated and the others are 3 coordinated, which are the same as those in m-phase. In short, in Path II, the t-o phase transition breaks four Zr-O bonds per cell (12-atom) but has little change in
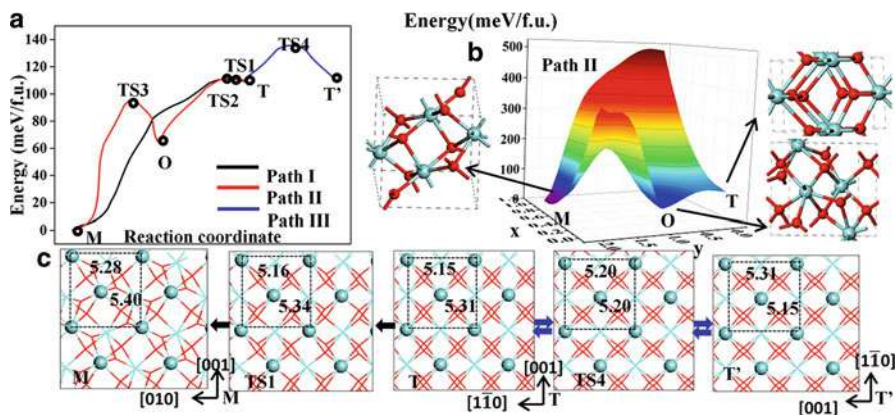
Fig. 4 (**a**) Potential energy profile for the lowest energy pathways from tetragonal phase (T or T')
to monoclinic (M) phase. (**b**) The 3-D PES of Path II obtained by the linear interpolation of lattice
parameters (x axis) and atomic fractional coordinates (y axis) from structure snapshots in Path II.
(**c**) The local structure change during phase transition following Path I and III viewed down from
the determined atomic habit plane, (110)t or (100)m surface. The Zr-Zr distances labeled are in Å.
Zr: cyan; O: red. (Reprinted with permission from (Guan et al. 2015). Copyright 2015 American
Chemical Society)

lattice; the o-m phase transition involves a large lattice distortion while keeping the
overall coordination.

The reaction pathway obtained from SSW-RS also reveals in detail how atoms
displace during the phase transition. The atomic movement on $(110)_t$ and $(100)_m$
in Path I and Path III is illustrated in Fig. 4c, where both two types of OR can be
generated. The $(110)_t$ and $(100)_m$ surfaces are determined to be the atomic habit
planes (interface) in the phase transition, which fulfill the following criteria: (i)
having minimum strain as evidenced by the close lattice parameters, e.g., $(110)_t$:
5.15×5.31 Å; $(100)_m$: 5.28×5.40 Å (Fig. 4c); (ii) having minimum atomic
movement in phase transition; (iii) acting as a coherent interface with close atomic
match between two phases, forming a heterophase junction.

Unlike the two lowest energy pathways, Path III is not directly leading to m-phase
but a ferroelastic transformation of t-phase. Such a reorientation of the individual
tetragonal domain is enabled by a permutation of the **b** and **c** axis in $(110)_t$ habit
plane. The t-t' phase transition goes through TS_4 that has a cubic-phase-like lattice
(P4/nbm, #125). The diagonal Zr-Zr distances in the habit plane are all 5.20 Å
(Fig. 4c), giving rise to a volume drop of 1.3% from t-phase to TS_4.

The results suggest that the OR of t-m phase transition should be dominated by
Type-C OR (Bansal and Heuer 1974). However, Type-B OR, although less often
observed, was reported experimentally in the phase transition of high temperature
zirconia single crystals (Wolten 1964) and in 3YSZ ceramics (Wu and Chiang
2011). To resolve this puzzle, the phase transition pathways under three different
external pressures (hydrostatic) at 5, 7, and 10 GPa are computed to mimic

the large stress-field that is developed at grain boundaries owing to the volume expansion during t-m phase transition. Among three pathways, only Path III is facilitated at high pressures, which turns out to be the lowest energy channel (ΔHa III $= 17.5$ meV) at 5 GPa compared to other two pathways (18.6 and 60.2 meV). This is in fact not surprising because the cubic-phase-like TS_4 has the smallest volume in all the located TSs and phases, and thus a compressive field helps to reduce the barrier greatly (the volumes of TS_1 and TS_2 are larger than t-phase, and that of TS_3 is also larger than o-phase). The result indicates that under high pressure condition the Type-B OR would be the more favorable OR for the t-m phase transition. Actually there are some experimental evidences that are consistent with this finding. For example, Chien et al. (1998) utilize Vickers microindentation experiment to induce the phase transition of Y-stabilized ZrO_2 from t-phase to m-phase. They observe the ferroelastic domain switching of t-phase together with the t-m Martensitic transition and importantly, only Type-B OR is found in the experiment. We emphasize that the presence of the o-phase does not add new OR. An important implication from current theoretical sampling on PES of ZrO_2 is that multiple energy-nearly-degenerate reaction channels can coexist in solid-to-solid phase transition and share the same OR. The knowledge of OR that is possible to obtain from experiment is in fact not enough to unravel the mechanism and kinetics of the solid-to-solid phase transition.

5.3   Graphite to Diamond Phase Transition

The graphite (G) to diamond transformation investigated by Xie et al. (2017) in 2017 is a good example to illustrate how to combine SSW and SSW-RS to resolve the mechanism of complex material structural transformation. It has long been a dream in human history (DeCarli and Jamieson 1961; Bundy and Kasper 1967; Hanneman et al. 1967; Naka et al. 1976) to convert graphite to diamond at mild reaction conditions. Modern research shows that the growth of large diamond crystals is kinetically hindered. The cubic diamond (CD), despite thermodynamically most stable among carbon allotropes above $\sim$5 GPa, is not the main product under mild static compression of graphite, $\sim$ 15 GPa and $\sim$1000 °C (Bundy and Kasper 1967; Yagi et al. 1992; Britun et al. 2004; Németh et al. 2014). In fact, the product is the dominated by hexagonal diamond (HD) and heterophase junctions. While great efforts have been devoted in past decades to understand the diamond phase growth from graphite matrix, the kinetics data computed from theory are still at odds with general experimental findings (Fahy et al. 1987; Scandolo et al. 1995; Tateyama et al. 1996; Mujica et al. 2003; Khaliullin et al. 2011; Xiao and Henkelman 2012).

Xie et al. (2017) first used the SSW method to scan the phase space of carbon allotropes. To identify the likely intermediates during diamond growth, they focused on those less stable, less ordered heterophase junctions that are transient in kinetics. SSW simulations with the environment dependent interatomic potential (EDIP) (Marks 2000) for carbon were carried out using different supercells up to 32-atom
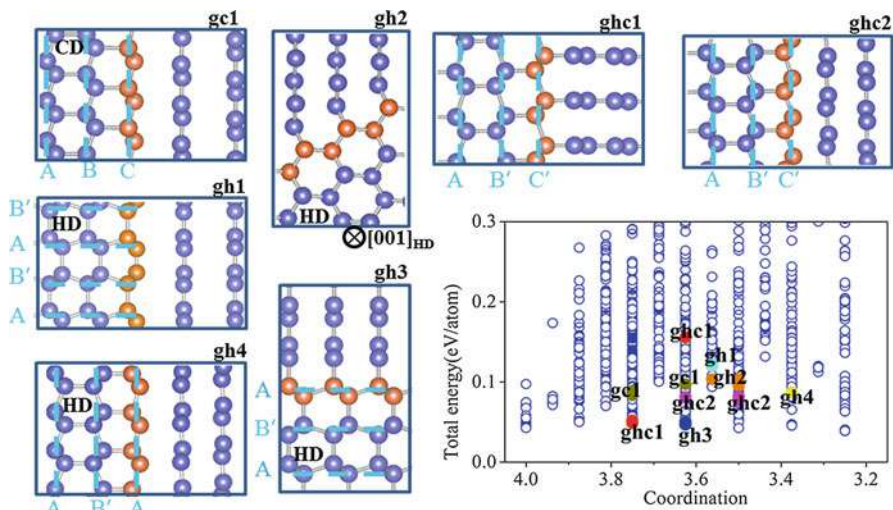
Fig. 5 Finding metastable heterophase junctions from SSW global PES exploration under 15 GPa. The right-bottom figure plots energy against average coordination number of 4000 structures. Different structures could have the same interface. (gh1)-to-(ghc2) show the structure of seven most stable interfaces from DFT. The dashed lines mark the close-packed planes, i.e., (111) for CD and (001) for HD. For gh2 interface, the interface is normal with $(001)_{HD}$. The interface atoms are orange colored for clarity. (Reprinted with permission from (Xie et al. 2017). Copyright 2017 American Chemical Society)

per cell at 15 GPa. The simulations were terminated until no more new types of junctions were revealed from PES exploration ($>10^6$ minima collected).

In total, more than 4000 minima were identified that are distinct from graphite and have relative low energetics (0.3 eV per atom above graphite). The energetics of the phases against the average carbon coordination number (N_c) is plotted in Fig. 5 (N_c is calculated by counting the first neighbors, where the C-C distance below 1.70 Å is regarded as bonding; N_c for diamond and graphite is 4 and 3, respectively). It shows that a large number of mixed phase intermediates (N_c in between 3 and 4) are obtained from SSW global structure search. By inspecting the geometry of all these structures closely, the metastable heterophase junctions that contain graphite component were screened out and distinguished according to the interface type and crystallographic orientation relationships (OR). There are seven types low energy heterophase junctions as indexed from gc1 to ghc2 in Fig. 5 and their associated interfaces are listed in Table 1.

For these heterophase junctions, gc1 is the only G/CD biphase junction; gh1 to gh4 are G/HD biphase junctions differing by the interfaces; and ghc1 and ghc2 are G/HD/CD three-phase junctions. HD/CD structure can be considered as a diamond structure with stacking faults on the close-packed plane. They belong to four distinct ORs as follows:

Table 1 Structure and energetics of metastable heterophase junctions in graphite-to-diamond solid phase transition. (Reprinted with permission from (Xie et al. 2017). Copyright 2017 American Chemical Society)

Name	Interface	OR_a	γ	ΔS	E_a
gc1	**$(001)_G//(111)_{CD}$**	**1**	**0.41**	**−3.7%**	**0.79**
gh1	**$(001)_G//(100)_{HD}$**	**3**	**0.23**	**0.7%**	**0.47**
gh2	$(110)_G//(1\bar{1}0)_{HD}$	3	0.24	22.6%	0.55
gh3	$(100)_G//(001)_{HD}$	3	0.35	25.2%	0.87
gh4	$(001)_G//(001)_{HD}$	4	0.42	−2.6%	0.85
ghc1	$(100)_G//(111)_{CD}$ $(100)_G//(001)_{HD}$	2,3	–	23.8%	0.84
ghc2	$(001)_G//(111)_{CD}$ $(001)_G//(001)_{HD}$	1,4	–	−3.2%	0.80

[a]Listed data include the interface planes, the crystallographic OR, the interfacial energy (γ, eV/Å^2), the reaction barrier (E_a, eV per interface atom) for the heterogeneous pathway mediated by the corresponding junction, the area misfit between connecting phases at the interface (ΔS, with respect to graphite)

OR1: (001)G//(111)CD, [100]G//[110]CD;
OR2: (001)G//(211)CD; [100]G//[110]CD;
OR3: (001)G//(100)HD, [1$\bar{1}$0]G//[001]HD;
OR4: (001)G//(001)HD, [010]G//[010]HD.

Among them, OR2 is 30 degrees off OR1, and OR4 is 90 degrees off OR3; OR2 does not occur independently but combines with OR3 to yield a C1-type three-phase junction (Scandolo et al. 1995). Only OR1 and OR3 were reported in experiment (Bundy and Kasper 1967; Yagi et al. 1992; Britun et al. 2004). For example, Britun et al. observed CD formation at 7 GPa and 1500 °C with OR1 (Britun et al. 2004).

The presence of multiple metastable heterophase junctions indicates that the kinetics of the pressure-induced graphite transformation is complex, governed by multiple anisotropic reaction channels. Xie et al. then utilized the SSW-RS simulation to determine the phase transition pathways that are mediated by these heterophase junctions.

The SSW-RS simulation starts from one heterophase junction (as listed in Table 1) and the pathways leading to G, HD, and CD phases are collected. After collecting a significant number of R/P pairs, i.e., >200 pairs for each heterophase junction, the TS was explicitly located using DFT and double-ended searching method, based on which the low energy pathways are determined by sorting the computed barriers. All the lowest energy pathways mediated by the seven different junctions were finally determined and the overall barriers of them (the highest energy TS along the pathway with respect to the graphite) are also listed in Table 1. The gh1-pathway (via gh1 junction) and gh2-pathway (via gh2 junctions) are the lowest barrier pathways, being 0.47 and 0.55 eV per interface atom (eV/ia), which is in line with the higher stability of their interfaces. The other pathways with less stable interfaces have much higher barriers, 0.75∼0.90 eV/ia.

Fig. 6 Reaction energy profile for three lowest energy pathways (top panel), and reaction snapshots along the lowest energy pathway from G to HD mediated by gh1-type interface (bottom panel). The relative energy of state (eV per interface atom) is indicated in the corner. (Reprinted with permission from (Xie et al. 2017). Copyright 2017 American Chemical Society)
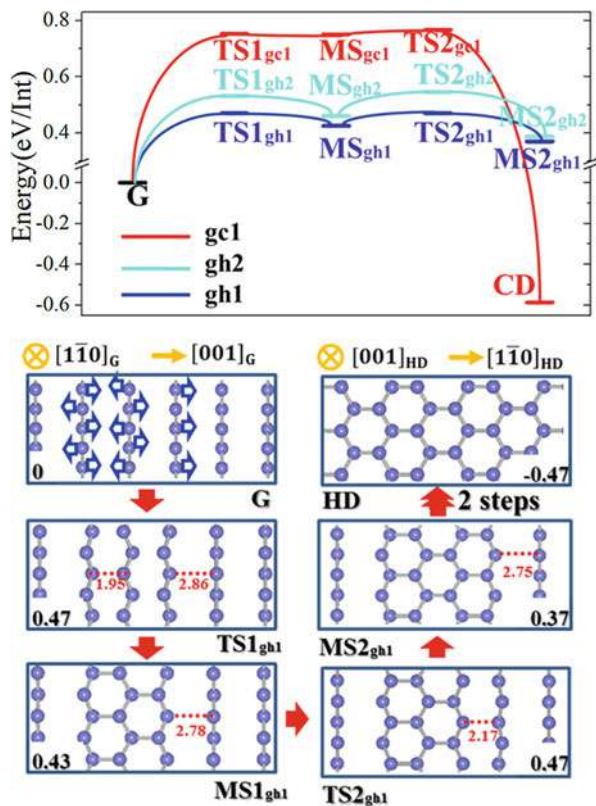
Figure 6 shows the reaction energy profiles of the lowest energy pathways mediated by gc1, gh1, and gh2 type interfaces. For the gh1 pathway from G to HD, it involves two intermediate states MS1 and MS2, both containing gh1 type interface. The lower overall barrier in gh1 compared to gc1 and gh2 can be attributed to the high stability of the reaction intermediates. The gh1 pathway in Fig. 6 is a representative to illustrate how atoms move during the phase transition.

The reaction initiates via the relative slipping of graphite basal plane along $[1\bar{1}0]_G$ by 1/2 C-C length (0.7 Å) followed by the compression at $[001]_G$. From G-to-MS1, the first step, the compression leads to the buckling of three neighboring graphite layers. The neighboring carbon rows along $[1\bar{1}0]_G$ in a graphite layer then displace collectively out of the basal plane at opposite directions and approach to the neighboring graphite layers. New C-C bonds are evolved at the TS ($TS1_{gh1}$) is 1.95 Å, after which a three-atomic-layer HD nucleus is created at MS1, showing the characteristic sp^3 bonding of diamond. The remaining steps, MS1-to-MS2 and MS2-to-HD, propagate the interface at $[001]_G$ with the same atom displacement pattern, leading to the growth of HD phase. The graphite basal plane evolves into $(100)_{HD}$, accompanied by 25.6% compression at $[001]_G$. In the pathway, the stability of the nascent HD nucleus (MS1) in graphite matrix dictates largely the

barrier height (TS1) of transformation, demonstrating the critical role of nucleation in kinetics.

The rate for graphite phase transition following heterogeneous pathways can be estimated: the formation of CD phase is at least 40 times slower compared to that to form HD phase at 1000 °C from microkinetics (by assuming the same prefactor). On the other hand, the reactions to CD phase and other mixed phase microstructures occur simultaneously: the rate difference is less than 1.12 times at 1000 °C. These quantities confirm HD phase is kinetically much favored, which correct the wrong prediction from collective pathways and MD results (Fahy et al. 1987; Scandolo et al. 1995; Tateyama et al. 1996; Khaliullin et al. 2011). Indeed, the G/HD composite with OR3 was commonly observed at 1000 °C (Bundy and Kasper 1967; Yagi et al. 1992; Britun et al. 2004; Németh et al. 2014), where CD phase is not formed yet; Németh et al. observed the mixed phase (CD/HD) at 2200 °C and 19 GPa (Németh et al. 2014) using transmission electron microscopy, indicating that the high temperature is a must to overcome the barrier of CD nucleation. The high HD selectivity is attributed to the presence of three low energy G/HD junctions, i.e., gh1, gh2, and gh3, orthogonal with each other, and two low energy anisotropic pathways via gh1 and gh2 junctions. These maximally stabilize HD nuclei inside graphite and allow for HD propagation in three dimensions (3D). By contrast, there is only one and high barrier gc1-pathway to form CD. Its barrier is similar to the pathways leading to the mixed phases. This indicates that the nascent CD nuclei in 3D inevitably introduce incoherent interfaces inside graphite and thus the CD growth is not favored both thermodynamically and kinetically.

6   Concluding Remarks

This chapter serves as a short overview of the SSW global optimization method and its recent application in material simulation. The SSW method is designed to combine some key features of existing theoretical methods for global structure search and reaction pathway sampling. A major feature in SSW method is the smooth structure perturbation, which mimics that in conventional MD simulations. While MD simulations meet great difficulties in overcoming high barriers, the SSW can go uphill on PES facilely by adding consecutive bias potentials along a soft curvature direction that is generated randomly and softened subsequently. The explicit consideration of the second derivation information and the repeated local relaxation are key techniques to keep the SSW trajectory close to low energy pathways.

Three examples on TiO_2, ZrO_2, and carbon are presented to demonstrate the current ability of SSW for global optimization and reaction pathway sampling. The SSW method in combination with first principles calculations can be utilized to generate a rich data on the PES of a material. It can reveal unknown new phases for constructing the phase diagram and help to map out the lowest energy pathways for understanding reaction kinetics. Perhaps more interestingly, this big data may also be utilized for machine learning, for example, to train high dimensional NN

potential of material. Considering that the NN potential is computationally much less demanding than the electronic structure calculation methods, the combination of SSW with machine learning techniques may bring a new era for material simulation in the decade to come.

References

Amsler M, Goedecker S (2010) Crystal structure prediction using the minima hopping method. J Chem Phys 133:224104

Bakardjieva S, Stengl V, Szatmary L, Subrt J, Lukac J, Murafa N, Niznansky D, Cizek K, Jirkovsky J, Petrova N (2006) Transformation of brookite-type TiO2 nanocrystals to rutile: correlation between microstructure and photoactivity. J Mater Chem 16:1709–1716

Bansal GK, Heuer AH (1974) On a martensitic phase transformation in zirconia (ZrO_2)—II, crystallographic aspects. Acta Metall 22:409–417

Bhattacharya K, Conti S, Zanzotto G, Zimmer J (2004) Crystal symmetry and the reversibility of martensitic transformations. Nature 428:55–59

Britun VF, Kurdyumov AV, Petrusha IA (2004) Diffusionless nucleation of lonsdaleite and diamond in hexagonal graphite under static compression. Powder Metall Met Ceram 43:87–93

Bundy FP, Kasper JS (1967) Hexagonal diamond—a new form of carbon. J Chem Phys 46:3437–3446

Carr JM, Trygubenko SA, Wales DJ (2005) Finding pathways between distant local minima. J Chem Phys 122:234903

Cheng S, Wei-Na Z, Zhi-Pan L (2015) Searching for new TiO2 crystal phases with better photoactivity. J Phys Condens Matter 27:134203

Chien FR, Ubic F, Prakash V, Heuer AH (1998) Stress-induced martensitic transformation and ferroelastic deformation adjacent microhardness indents in the tetragonal zirconia single crystals. Acta Mater 46:2151–2171

DeCarli PS, Jamieson JC (1961) Formation of diamond by explosive shock. Science 133:1821–1822

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1:269–271

Fahy S, Louie SG, Cohen ML (1987) Theoretical total-energy study of the transformation of graphite into hexagonal diamond. Phys Rev B 35:7623–7626

Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132:154104

Guan S-H, Zhang X-J, Liu Z-P (2015) Energy landscape of zirconia phase transitions. J Am Chem Soc 137:8010–8013

Hanneman RE, Strong HM, Bundy FP (1967) Hexagonal diamonds in meteorites: implications. Science 155:995–997

Henkelman G, Jonsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111:7010–7022

Huang S-D, Shang C, Zhang X-J, Liu Z-P (2017) Material discovery by combining stochastic surface walking global optimization with a neural network. Chem Sci 8:6327–6337

Huber T, Torda A, Gunsteren W (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. J Comput Aided Mol Des 8:695–708

Iannuzzi M, Laio A, Parrinello M (2003) Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. Phys Rev Lett 90:238302

Jin X-J (2005) Martensitic transformation in zirconia containing ceramics and its applications. Curr Opin Solid State Mater Sci 9:313–318

Kelly PM, Rose LRF (2002) The martensitic transformation in ceramics-its role in transformation toughening. Prog Mater Sci 47:463–557

Khaliullin RZ, Eshet H, Kühne TD, Behler J, Parrinello M (2011) Nucleation mechanism for the direct graphite-to-diamond phase transition. Nat Mater 10:693–697

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. Phys Rev B 59:1758–1775

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci U S A 99:12562–12566

Li Y-F, Liu Z-P (2011) Particle size, shape and activity for photocatalysis on titania anatase nanoparticles in aqueous surroundings. J Am Chem Soc 133:15743–15752

Li Y-F, Liu Z-P, Liu L, Gao W (2010) Mechanism and activity of photocatalytic oxygen evolution on titania anatase in aqueous surroundings. J Am Chem Soc 132:13008–13015

Lonie DC, Zurek E (2011) XtalOpt: an open-source evolutionary algorithm for crystal structure prediction. Comput Phys Commun 182:372–387

Ma S, Huang S-D, Fang Y-H, Liu Z-P (2017) Microporous titania crystals with penta-oxygen coordination. ACS Appl Energy Mater 1:22–26

Maeda S, Ohno K, Morokuma K (2009) Automated global mapping of minimal energy points on seams of crossing by the anharmonic downward distortion following method: a case study of H_2CO. J Phys Chem A 113:1704–1710

Maeda S, Taketsugu T, Morokuma K (2014) Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. J Comput Chem 35:166–173

Maragliano L, Vanden-Eijnden E (2006) A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. Chem Phys Lett 426:168–175

Marks NA (2000) Generalizing the environment-dependent interaction potential for carbon. Phys Rev B 63:035401

Martonak R, Donadio D, Oganov AR, Parrinello M (2006) Crystal structure transformations in SiO_2 from classical and ab initio metadynamics. Nat Mater 5:623–626

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Mujica A, Rubio A, Muñoz A, Needs RJ (2003) High-pressure phases of group-IV, III-V, and II-VI compounds. Rev Mod Phys 75:863–912

Naka S, Horii K, Takeda Y, Hanawa T (1976) Direct conversion of graphite to diamond under static pressure. Nature 259:38–39

Németh P, Garvie LAJ, Aoki T, Dubrovinskaia N, Dubrovinsky L, Buseck PR (2014) Lonsdaleite is faulted and twinned cubic diamond and does not exist as a discrete material. Nat Commun 5:5447

Oganov AR, Glass CW (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. J Chem Phys 124:244704

Oganov AR, Lyakhov AO, Valle M (2011) How evolutionary crystal structure prediction works-and why. Acc Chem Res 44:227–237

Pannetier J, Bassas-Alsina J, Rodriguez-Carvajal J, Caignaert V (1990) Prediction of crystal structures from crystal chemistry rules by simulated annealing. Nature 346:343–345

Penn RL, Banfield JF (1999) Formation of rutile nuclei at anatase {112} twin interfaces and the phase transformation mechanism in nanocrystalline titania. Am Mineral 84:871–876

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Raiteri P, Martoňák R, Parrinello M (2005) Exploring polymorphism: the case of benzene. Angew Chem Int Ed 44:3769–3773

Rosso L, Minary P, Zhu ZW, Tuckerman ME (2002) On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. J Chem Phys 116:4389–4402

Scandolo S, Bernasconi M, Chiarotti GL, Focher P, Tosatti E (1995) Pressure-induced transformation path of graphite to diamond. Phys Rev Lett 74:4015–4018

Schaefer B, Mohr S, Amsler M, Goedecker S (2014) Minima hopping guided path search: an efficient method for finding complex chemical reaction pathways. J Chem Phys 140:214102–214113

Schön JC, Jansen M (1996) First step towards planning of syntheses in solid-state chemistry: determination of promising structure candidates by global optimization. Angew Chem Int Ed Eng 35:1286–1304

Schön JC, Jansen M (2001) Determination, prediction, and understanding of structures, using the energy landscapes of chemical systems – part III. Z Kristallogr 216:361–383

Shang C, Liu Z-P (2010) Constrained Broyden minimization combined with the dimer method for locating transition state of complex reactions. J Chem Theory Comput 6:1136–1144

Shang C, Liu ZP (2012) Constrained Broyden dimer method with bias potential for exploring potential energy surface of multistep reaction process. J Chem Theory Comput 8:2215–2222

Shang C, Liu ZP (2013) Stochastic surface walking method for structure prediction and pathway searching. J Chem Theory Comput 9:1838–1845

Shang C, Zhang X-J, Liu Z-P (2014) Stochastic surface walking method for crystal structure and phase transition pathway prediction. Phys Chem Chem Phys 16:17845–17856

Sheppard D, Xiao P, Chemelewski W, Johnson DD, Henkelman G (2012) A generalized solid-state nudged elastic band method. J Chem Phys 136:074103

Siepmann JI, Frenkel D (1992) Configurational bias Monte-Carlo – a new sampling scheme for flexible chains. Mol Phys 75:59–70

Tateyama Y, Ogitsu T, Kusakabe K, Tsuneyuki S (1996) Constant-pressure first-principles studies on the transition states of the graphite-diamond transformation. Phys Rev B 54:14994–15001

Trimarchi G, Zunger A (2007) Global space-group optimization problem: finding the stablest crystal structure without constraints. Phys Rev B 75:104113

van Beest BWH, Kramer GJ, van Santen RA (1990) Force fields for silicas and aluminophosphates based on ab initio calculations. Phys Rev Lett 64:1955–1958

Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett 78:3908–3911

Wales DJ (2002) Discrete path sampling. Mol Phys 100:3285–3305

Wales DJ, Doye JPK (1997) Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J Phys Chem A 101:5111–5116

Wang CM, Fan KN, Liu ZP (2007) Origin of oxide sensitivity in gold-based catalysts: a first principle study of CO oxidation over Au supported on monoclinic and tetragonal ZrO. J Am Ceram Soc 129:2642–2647

Wang Y, Lv J, Zhu L, Ma Y (2012) CALYPSO: a method for crystal structure prediction. Comput Phys Commun 183:2063–2070

Wei Z-Y, Shang C, Zhang X-J, Liu Z-P (2017) Glassy nature and glass-to-crystal transition in the binary metallic glass CuZr. Phys Rev B 95:214111

Wolten GM (1963) Diffusionless phase transformations in zirconia and hafnia. J Am Ceram Soc 46:418–422

Wolten GM (1964) Direct high-temperature single-crystal observation of orientation relationship in zirconia phase transformation. Acta Cryst 17:763–765

Woods CJ, Essex JW, King MA (2003) The development of replica-exchange-based free-energy methods. J Phys Chem B 107:13703–13710

Wu Y-C, Chiang Y-T (2011) The m-t transformation and twinning analysis of hot-pressed sintered 3YSZ ceramics. J Am Ceram Soc 94:2200–2212

Xiao P, Henkelman G (2012) Communication: from graphite to diamond: reaction pathways of the phase transition. J Chem Phys 137:101101

Xie Y-P, Zhang X-J, Liu Z-P (2017) Graphite to diamond: origin for kinetics selectivity. J Am Chem Soc 139:2545–2548

Yagi T, Utsumi W, M-a Y, Kikegawa T, Shimomura O (1992) High-pressure \textit{in situ} x-ray-diffraction study of the phase transformation from graphite to hexagonal diamond at room temperature. Phys Rev B 46:6031–6039

Yu T-Q, Tuckerman ME (2011) Temperature-accelerated method for exploring polymorphism in molecular crystals based on free energy. Phys Rev Lett 107:015701

Zhang H, Banfield JF (1999) New kinetic model for the nanocrystalline anatase-to-rutile transformation revealing rate dependence on number of particles. Am Mineral 84:528–535

Zhang X-J, Liu Z-P (2015a) Variable-cell double-ended surface walking method for fast transition state location of solid phase transitions. J Chem Theory Comput 11:4885–4894

Zhang X-J, Liu Z-P (2015b) Reaction sampling and reactivity prediction using the stochastic surface walking method. Phys Chem Chem Phys 17:2757–2769

Zhang X-J, Shang C, Liu Z-P (2013) Double-ended surface walking method for pathway building and transition state location of complex reactions. J Chem Theory Comput 9:5745–5753

Zhang X-J, Shang C, Liu Z-P (2017a) Stochastic surface walking reaction sampling for resolving heterogeneous catalytic reaction network: a revisit to the mechanism of water-gas shift reaction on Cu. J Chem Phys 147:152706

Zhang X-J, Shang C, Liu Z-P (2017b) Pressure-induced silica quartz amorphization studied by iterative stochastic surface walking reaction sampling. Phys Chem Chem Phys 19:4725–4733

Zhao W-N, Liu Z-P (2014) Mechanism and active site of photocatalytic water splitting on titania in aqueous surroundings. Chem Sci 5:2256–2264

Zhao W-N, Zhu S-C, Li Y-F, Liu Z-P (2015) Three-phase junction for modulating electron–hole migration in anatase–rutile photocatalysts. Chem Sci 6:3483–3494

Zhu S-C, Xie S-H, Liu Z-P (2014) Design and observation of biphase TiO_2 crystal with perfect junction. J Phys Chem Lett 5:3162–3168

Zhu S-C, Xie S-H, Liu Z-P (2015) Nature of rutile nuclei in anatase-to-rutile phase transition. J Am Chem Soc 137:11532–11539

First-Principles-Assisted Structure Solution: Leveraging Density Functional Theory to Solve Experimentally Observed Crystal Structures

119

Kyle Michel, Bryce Meredig, Logan Ward, and Christopher Wolverton

Contents

Abstract

FPASS (first-principles-assisted structure solution) is a global optimization method that can be used to determine the structure of a material from a powder X-ray diffraction pattern (Meredig and Wolverton, Nat Mater 12(2):123–127, 2013). Using a hybrid of Rietveld refinement and minimization of energies from density functional theory calculations, FPASS can be used to solve crystal structures with reduced guidance from experts typically required for Rietveld

K. Michel (✉) · B. Meredig
Citrine Informatics Inc., Redwood City, CA, USA
e-mail: kyle@citrine.io; bryce@citrine.io

L. Ward
Computation Institute, University of Chicago, Chicago, IL, USA
e-mail: loganw@uchicago.edu

C. Wolverton
Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA
e-mail: c-wolverton@northwestern.edu

refinement alone. This article presents the core concepts of the FPASS method and how it compares to other structure solution and prediction methods. A number of case studies are presented that show the breadth of structure types that have been solved using this method, including those run in a high-throughput mode. Finally, future directions and research opportunities around the FPASS method are discussed.

1   Introduction

Powder X-ray diffraction (XRD) is a common technique used to characterize the structure of a material (Cullity and Stock 2001). A sample of material is turned into a powder and then bombarded with X-rays. The diffracted intensities of those X-rays are measured in the area surrounding it. The resulting intensity pattern, which is a function of the scattering angle, will have peaks and shapes that are characteristic of the structure of the material. Once the diffraction pattern for a material has been obtained, it can be compared against libraries of patterns for known materials. This process allows for the identification of the material, or combination of materials, that are contained in the sample. However, this method for determining the phases present in a sample only works when reference patterns exist for all of those phases. When this method fails, the structure of the material must be solved in a different way.

Powder XRD patterns are a direct function of the structure of the material being characterized. In this way, they represent a reduction from a very high-dimensional space (the unit cell shape and atomic positions within it) to a much lower-dimensional space (the two-dimensional pattern of intensity vs. scattering angle). The challenge in solving structures is to reverse that reduction, that is, to determine the crystal unit cell shape and atomic coordinates from the two-dimensional, measured pattern.

The common workflow for solving a crystal structure from powder XRD data is composed of three main steps: unit cell determination, identification of atomic positions, and structure refinement (Pecharsky and Zavalij 2005). The first step, unit cell determination, is the process of inferring the parameters and symmetry of the unit cell from the relative diffraction angles, intensities, and systematic absences of diffraction peaks. The second step involves determining which arrangement of atoms within the unit cell best describes the observed diffraction pattern. The final step is to optimize the agreement between the simulated pattern of the proposed structure and the diffraction data by making fine adjustments to the proposed solution that maximize the agreement. A common, modern approach for the refinement step is Rietveld refinement, a process that not just optimizes the crystal structure parameters but also accounts for other factors (e.g., grain size) that affect the shapes of the individual diffraction peaks (Rietveld 1969).

The complexity of solution from Powder XRD lies in the often-large space of possible crystal structures and the low amount of information present in powder diffraction patterns. With an infinite amount of computing time, every possible

combination of atomic positions for a structure could be generated and then refined. Unfortunately, in reality, this process would be impossibly slow. Including the shape of the lattice, there are $3N + 3$ degrees of freedom in a structure with N atoms (if the symmetry of the structure is ignored). While this space of candidate structures is occasionally small enough for a human expert to find the solution after only a few guesses, there do exist methods to automate the solution processing for the many difficult cases. One such option is to employ global optimization algorithms.

Methods that employ global optimization to solve crystal structures are known as direct-space structure solution methods (Černý 2017). Originally developed in the early 1990s (Harris and Tremayne 1996), direct-space methods leverage the ability of global optimization algorithms to determine solutions that maximize an objective function (in this case the match to the diffraction pattern) with minimal numbers of guesses, making it possible to search intractably large solution spaces. In the following decades, scientists developed a variety of direct-space methods that employ many different global optimization algorithms, including simulated annealing (Deem and Newsam 1989), genetic algorithms (Kariuki et al. 1997), and, more recently, particle swarm optimization (Gao et al. 2017).

In addition to the size of the search space that must be explored, a second difficulty can arise during global optimization; two or more structures may be generated with very similar matches to the experimental diffraction pattern. There are several routes to resolving this ambiguity when it occurs. Of course, one can always apply domain knowledge to select which of the candidate solutions is more likely or perform additional experiments (e.g., NMR, neutron diffraction) that may differentiate structures with similar X-ray patterns. Another route is to integrate the energy of a structure into the solution process. Known as hybrid direct-space structure solution methods, they use the energy computed from empirical potentials to find which structures are most likely to be stable (Lanning et al. 2000; Putz et al. 1999). Adding structural energy as a metric during the structure solution guides the optimization algorithm away from unphysical solutions, which further speeds the solution process (Lanning et al. 2000).

These powerful concepts of efficient global optimization using the energy as a guiding function are combined in the FPASS method (First-Principles Assisted Structure Solution) (Meredig and Wolverton 2013; Ward et al. 2015, 2017). FPASS allows for structure solution from powder XRD patterns with limited human intervention. It combines global search methods, biasing based on energies obtained from first-principles calculations, and local optimization through Rietveld refinement. The details of the method are described in the next section and a number of applications in the section following that.

2 The FPASS Method

There are three main components to the FPASS method. First, a global optimization strategy is employed to search across all configurations of atoms in a crystal structure with a fixed composition in order to find structures in the neighborhood

of those likely to have been observed in experiments. Second, energies computed from density functional theory (DFT) are used to guide the selection of candidates so that structures with relatively low energies are favored to those with higher energies. Third, a local refinement of each candidate crystal structure minimizes the difference between observed and computed XRD patterns. In the end, the goal of the FPASS method is to find a crystal structure that minimizes the total energy of the structure while having a simulated diffraction pattern that most closely matches the experimentally observed one.

Density functional theory is used in FPASS due to its ability to produce the energy of a material given only its lattice parameters and positions of atoms in the unit cell (Hohenberg and Kohn 1964; Kohn and Sham 1965). This feature makes it general enough to apply to any problem for which the counts of each element in the structure are known. However, any energy function could be used in its place. For example, a classical potential, if one existed for a system, could be used to more quickly calculate energies but would be less accurate than DFT. The only requirement of the energy calculation is that it accurately ranks the relative energies of any set of structures in a single FPASS run.

A coarse search of structures with a fixed composition is performed using a genetic algorithm. Genetic algorithms mimic the evolution of a species by creating a series of generations, each a function of crossover (or mating) operations of the members of the current generation, and random mutations of the children of those operations. In the context of structure solution, each member of a generation represents a single set of atomic positions within a fixed unit cell on a lattice. A genetic algorithm is parameterized by some initial population, a set of allowed crossover operations, and a set of randomly selected mutations.

The symmetry, or space group, of a material puts constraints on the atomic positions of the atoms in the structure (Michel and Wolverton 2014). These constraints manifest in two ways. First, the space group of a material defines a set of Wyckoff sites. Each Wyckoff site contains one or more positions that are equivalent by symmetry and which must all be occupied if any one of them is; this set of positions is known as the orbit of the site. Some space groups may only have a few Wyckoff sites while others can have tens of them. Choosing which set of Wyckoff sites to occupy (some can be occupied more than once) for a structure amounts to searching over all combinations such that the number of atoms sums to the correct amount. Second, the continuous positions of the symmetrically unique atoms may be constrained by the symmetry of a Wyckoff site. In this way, an atom might be free to move in all directions on a single position of a Wyckoff site, or it may be restricted to a plane, line, or even to a specific point.

Since the number of possible combinations of Wyckoff sites can be extremely large, FPASS includes heuristics to select likely Wyckoff site occupations that are based on experimentally observed structures. In particular, a survey of the Inorganic Crystal Structure Database (ICSD) (Belsky et al. 2002) has shown that a majority of crystal structures have site occupancies such that the number of Wyckoff sites that are occupied is minimized for the number of atoms in the structure. In other words, nature tends to favor crystal structures in which the number of symmetrically

unique atoms is minimized. When a structure is generated in FPASS, the list of possible combinations of Wyckoff sites is enumerated such that atomic counts are preserved. Each of these combinations is assigned a probability that is proportional to the fraction of ICSD compounds with that number of Wyckoff sites relative to the minimum possible. For example, if the minimum number of Wyckoff sites that a structure could occupy is two, then a set of four Wyckoff sites would be assigned a probability equal to the fraction of structures in the ICSD that contained exactly twice the minimum possible number of sites. The degree to which combinations of sites are biased by this distribution is an input to each FPASS calculation. When the ICSD distribution is weighted heavily, the structures that are found in the global search are likely to be similar to those in the ICSD. Conversely, when the ICSD distribution has a small weight or is turned off entirely, the optimization is more likely to find structures that are very dissimilar to those in the ICSD.

Once a set of Wyckoff sites has been selected, atomic positions are generated so that they obey the symmetries of those sites. For example, an atom on the 2h Wyckoff site (1/3,2/3,+z) of the $P\bar{6}$ space group can occupy any position along the z axis but must have x and y coordinates equal to 1/3 and 2/3, respectively. When the coordinates of one atom on a Wyckoff site have been generated, all atoms in the orbit of that site are fixed by its symmetry.

Two values are calculated for every structure in a generation. First, the match between the experimental and simulated XRD patterns is optimized (see discussion below on Rietveld refinement). Each structure i is assigned a score S_i^{XRD} which is equal to its rank against all other structures when their matches to experiment are placed in ascending order. For example, the structure with the best agreement between experimental and simulated XRD patterns would have $S_i^{XRD} = 1$, while the structure with the worst agreement would have S_i^{XRD} equal to the number of structures in the population. Second, the total energy of each structure is calculated using DFT. When ranked from lowest energy to highest, each structure is assigned a score S_i^{DFT} equal to its position in the list. Finally, each structure is given an aggregate score:

$$S_i = S_i^{XRD} + S_i^{DFT}. \tag{1}$$

As an example, a structure with both the lowest total energy and best match to experimental XRD pattern would have $S_i = 2$. A fixed number of the best performing structures (those with the lowest values of S_i) are immediately carried over to the next generation. The remaining structures are replaced by structures generated in crossover operations where those structures with lower values of S_i are more likely to have their features carried over to the next generation as compared to structures with higher values.

The objective function described in Eq. 1 was selected because summing ranks eliminates the need to unify scales between XRD match and DFT energies. In order to avoid assigning a weight to bias the algorithm toward sampling structures with better or worse matches to the XRD pattern, the structure with the lowest energy is

selected as the candidate solution. In approximately 100 test cases, it was found that the lowest-energy structure within the constraints employed in the FPASS calculation (e.g., symmetry, unit cell parameters) and the experimentally determined crystal structure are always the same. It is possible to test whether the lowest-energy structure also has the best XRD match for new cases because the XRD match and energy are recorded for each structure evaluated by FPASS.

Crossover operations are defined in FPASS by choosing combinations of Wyckoff sites from the parent structures. Consider the case of the $P\bar{6}$ space group with four atoms. Assume that the first parent has two atoms on the 2i site and two atoms on the 2h site, while the second parent has one atom on each of the 1f, 1e, 1d, and 1c sites. During crossover, all combinations of sites that add up to four total atoms are enumerated. For example, a structure with atoms on the 2i, 1f, and 1e sites would be among the set of candidates. Each combination is assigned a weight using the ICSD-based biasing, then one of those sets of Wyckoff sites is chosen at random to be used as the child that is carried to the next generation.

With some probability, each child that is generated during the genetic algorithm will be selected to have either or both of two mutation types applied to it. In the first mutation type, the positions of atoms are randomly perturbed while ensuring that those perturbations do not break the symmetry of the structure. This mutation allows for changes to the positions of atoms on each Wyckoff site. In the second mutation type, new combinations of Wyckoff sites can be randomly selected, and atoms moved from the old sites to the new ones. When these new Wyckoff sites are selected, atoms are placed at random positions while ensuring that symmetry is preserved.

Once a set of structures has been generated, Rietveld refinement is carried out for each. The objective function during refinement is defined by Pecharsky and Zavalij (2005):

$$ R = \frac{\sum_{2\theta} \left| I_{\text{calc}} - \left(I_{\text{obs}} - I_{\text{bg}} \right) \right|}{\sum_{2\theta} \left| I_{\text{obs}} - I_{\text{bg}} \right|} \tag{2}$$

where I_{calc}, I_{obs}, and I_{bg} are the calculated, observed, and background intensities, respectively, and the sums run over angles measured in the observed pattern. The background intensity, I_{bg}, is modeled as a Chebyshev polynomial of the angle, the features of which are treated as fitting parameters during refinement. The calculated pattern for diffraction off of the plane hkl (which can be directly converted to an angle 2θ) has the following definition:

$$ I_{hkl} = K \times m_{hkl} \times LP\left(\theta\right) \times T_{hkl} \times A_{hkl} \times |F_{hkl}|^2 . \tag{3}$$

K is a scaling factor that can be optimized during refinement. m_{hkl} is the number of planes that are equivalent by symmetry to hkl (since the intensity only needs to be computed for one). $LP\left(\theta\right)$ are the Lorentz and polarization factors. T_{hkl} is the preferred grain orientation or texturing. A_{hkl} is the volume absorption of the plane.

$|F_{hkl}|^2$ is the structure factor. Each peak is broadened by multiplying the computed intensity (I_hkl) by a psuedo-Voight function centered at the peak center (2Θ). The width parameters of the pseudo-Voight function are optimized during refinement, the parameters of which are optimized during refinement. Using this function for the calculated intensity, all free parameters are adjusted during refinement in order to minimize Eq. 2.

As described, the FPASS method sits at the intersection between energy-based structure prediction and experimental solution to powder XRD patterns. There are a number of difficulties when each of these methods is used individually that FPASS can help alleviate. First, the size of the search space can be very large ($3N + 3$ degrees of freedom for N atoms when the lattice shape is not known) in pure structure prediction. However, the use of a reference powder XRD pattern can bias the search in the space of structures that are most physical. In addition, the symmetry of the space group determined from the XRD pattern will, unless it is P1, impose constraints on atomic positions and reduce the number of free variables in the optimization. Second, as already discussed, the global search methods in FPASS allow for structure solution even in the absence of an expert to drive them. Third, some elements, such as hydrogen, do not scatter X-rays strongly, and the positions of these atoms can be impossible to determine through powder XRD alone. Since the energy of structures are calculated in FPASS, the positions of these atoms can be taken from the lowest-energy configurations. In the next section, example applications of FPASS are described that take advantage of each of these features.

3   Selected Results

There are two classes of problems to which FPASS has been applied: first, to validate FPASS by confirming existing structures in the literature and, second, to solve diffraction patterns that had no known structure associated with them. In this section, a number of examples of each of these types are highlighted in order to show the breadth of problems that have been solved using this method.

3.1 Li_2O_2

Li_2O_2 is a material that has been widely studied for use in lithium-air batteries. Given its technological importance, the structure of the material is of interest since it is needed, for example, to calculate diffusion energetics using DFT (which requires a complete crystal structure as input). Two crystal structures, both in the $P\bar{6}$ space group, had been proposed for this chemistry: one structure by Föppl (1957) with an 8-atom basis and the other by Fehér et al. (1953) with a 32-atom basis. DFT can be used to determine which of these structures has the lowest energy at $T = 0$ K. Indeed, the structure proposed by Föppl has a lower energy than that given by Fehér (Cota and de la Mora 2005; Chan et al. 2011).
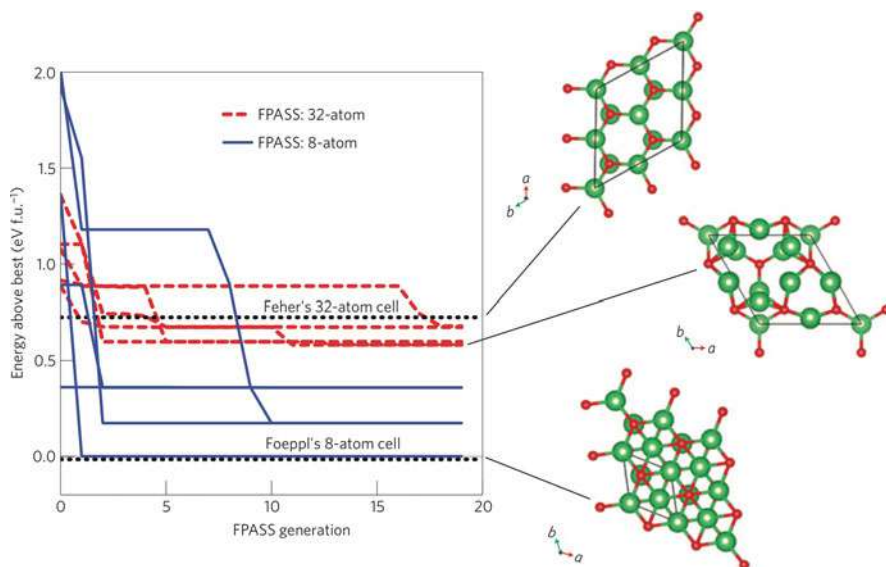
Fig. 1 Energy of the best structure at each step of an FPASS optimization of Li_2O_2. Red, dashed lines use a 32-atom cell while blue, solid lines use an 8-atom cell. The structure proposed by Föppl (1957) is the lowest-energy structure found during all simulations (Meredig and Wolverton 2013)

Given the questions around the structure in this system, FPASS was applied to determine whether there might exist some other structure in the same space group with a better match to the experimental pattern (Meredig and Wolverton 2013). A 32-atom cell was generated for Li_2O_2 in the $P\bar{6}$ space group. The results for this optimization are shown in Fig. 1. These findings support those of other studies (Cota and de la Mora 2005; Chan et al. 2011) in that the Föppl structure is the lowest-energy one that exists.

3.2 MgNH

Due to the presence of hydrogen, the structure of MgNH is difficult to resolve from an XRD pattern and is also difficult to predict from energetics alone due to the complexity of the structure. In this way, MgNH is an ideal candidate for solving with FPASS since a hybrid approach can mitigate both issues. Some historical information was used in applying FPASS to this system. Previous studies of MgNH have lead researchers to conclude that the structure contains 36 atoms in its unit cell and is one of four space groups; from XRD, the $P6_3$, $P6_3/m$, and $P6_322$ were proposed (Jacobs and Juza 1969) while neutron scattering was used to suggest the $P6/m$ space group (Dolci et al. 2011). Ten different optimizations were performed in each of these four space groups, all with 36 atoms. The results of these simulations are shown in Fig. 2 (Meredig and Wolverton 2013).
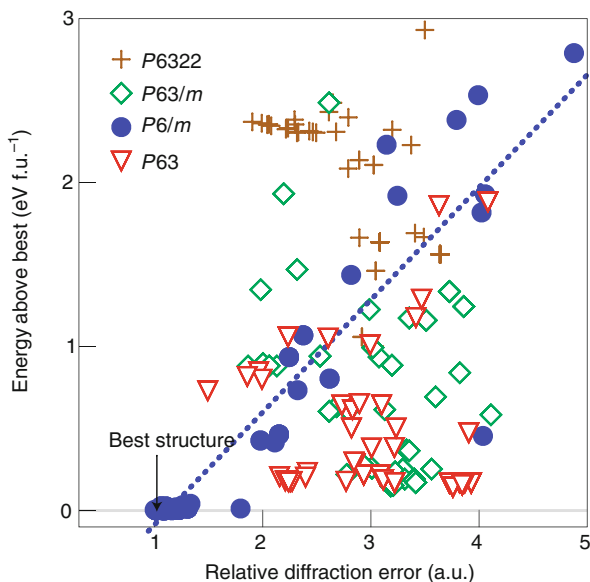
Fig. 2 The energy (y axis) and relative diffraction error (x axis) for all structures generated during FPASS runs of MgNH in the P6$_3$, P6$_3$/m, P6$_3$22, and P6/m space groups. Different symbols denote the different space groups that were studied. Structures in the P6/m space group had both the lowest energy and best match to diffraction patterns during these runs (Meredig and Wolverton 2013)

From the FPASS runs that were performed, the lowest-energy structures and best matches to experimental patterns were simultaneously found in the P6/m space group. Surprisingly, the top P6/m structure also had the best agreement with the diffraction pattern for which the space group had been reported as being P6$_3$, P6$_3$/m, or P6$_3$22. This shows that non-optimal solutions can be found even when experts drive the refinement process.

3.3   Na-Pb System

The Na-Pb binary system contains three phases (γ-Na$_{13}$Pb$_5$, δ-Na$_5$Pb$_2$, and δ'-Na$_9$Pb$_4$) for which the structure has been partially but not completely solved. In all three cases, the lattice parameters and Pb positions were known, but the Na positions were not. In Fig. 3, the convex hull of DFT energies for the completely-solved structures in the Na-Pb system is shown in orange. A goal in solving the structures of the γ, δ, and δ' phases was to determine whether they would be energetically favorable compared to structures that were already solved.

The space groups of these structures had been determined to be R$\bar{3}$m for δ-Na$_5$Pb$_2$ and P6$_3$/mmc for both γ-Na$_{13}$Pb$_5$ and δ'-Na$_9$Pb$_4$ (Weston and Shoemaker 1957). In all FPASS runs, the lattice parameters and Pb positions were fixed to the experimentally determined values, while the positions of Na atoms were optimized.

Fig. 3 Convex hull for the Na-Pb system. Orange points show the phases present on the hull prior to the FPASS calculations (energies calculated with DFT for known structures). The blue points are phases that were studied using FPASS. All three phases were found to lie on the hull or were very close to it, suggesting that they are energetically stable or near to it
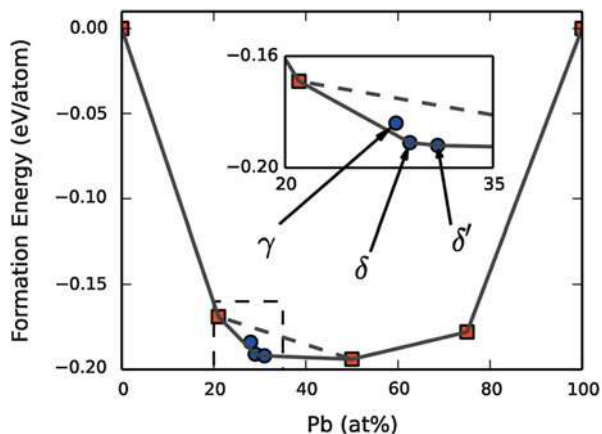
The structures of all three phases were determined using FPASS (Ward et al. 2015), and qualitative agreement was found with hypotheses from experiment about the Wyckoff site occupancies and positions of Na atoms (Weston and Shoemaker 1957). The energies of these newly solved structures are added to the convex hull in Fig. 3 as blue points. Clearly, all three structures are stable relative to the previously known structures. From the calculations that were performed, the δ-Na_5Pb_2 and δ'-Na_9Pb_4 phases were found to lie on the convex hull, suggesting that they are energetically stable at $T = 0$ K with respect to decomposition to other phases in the system. The γ-$Na_{13}Pb_5$ phase was found to be only slightly above the hull and is likely stabilized at finite temperature.

Importantly, it was hypothesized that the Na positions would be partially occupied in the γ-$Na_{13}Pb_5$ and δ-Na_5Pb_2 structures (Weston and Shoemaker 1957). However, the implementation of DFT used by FPASS currently supports only fully occupied sites, which means that partially occupied structures could not be explored. Therefore, these calculations put an upper bound on the energies of the structures. That is, the partially occupied structures may have even lower energies than the ones that are reported.

Of these solutions, γ-$Na_{13}Pb_5$ is particularly interesting. A common method for structure solution is to search the ICSD for structures with the same space group and relative stoichiometries of the phase in question (Curtarolo et al. 2003). However, in the case of $Na_{13}Pb_5$, no structure exists in the ICSD that has a 5:13 ratio of atoms, 36 atoms in the unit cell, and which is in the $P6_3/mmc$. In other words, the structure would not have been solvable had a global optimization method not been employed.

3.4 LiSbO₃

A monoclinic phase of $LiSbO_3$ was discovered in 2006. The structure was determined to be in the C2/m space group, while the positions of the atoms could not be
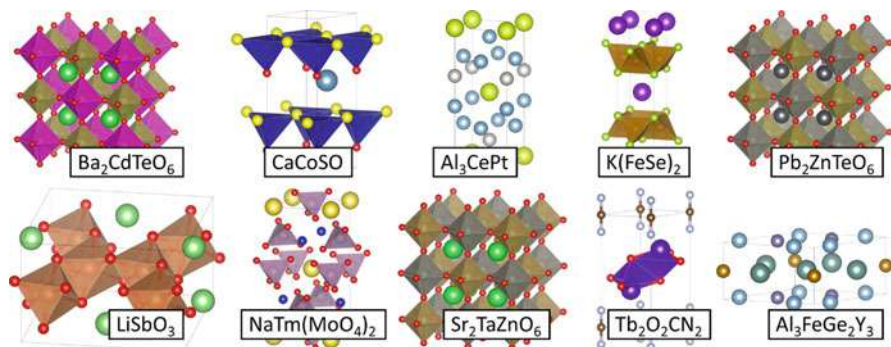
Fig. 4 Ten different structures that were solved using FPASS. Each solution started with an assumed space group and lattice parameters, and the positions within the unit were optimized for fixed composition. Each structure solution was checked for convergence and agreement with experimental patterns

determined. In addition to the C2/m structure, an orthorhombic phase is known to exist at the same composition.

Confined to the C2/m space group, FPASS found a layered structure of Li on alternate planes from Sb and O, where the O anions form an FCC lattice and the Sb anions are octahedrally coordinated with the O anions. This structure is shown in Fig. 4, and the positions of atoms are in agreement with qualitative suggestions made by (Nalbandyan et al. 2006) In DFT calculations, the energy of this structure was found to be within 1 meV/atom (higher) of the orthorhombic phase. That is, a purely energy-based structure prediction calculation would not have been able to differentiate this structure and the orthorhombic one. In fact, this is true of $T = 0$ K structure prediction for any material that has metastable or high-temperature polymorphs. The comparison to experimental XRD pattern is vital to the solution of the structure in these cases.

3.5   Other Solutions

FPASS has been used to solve many other systems, some of which are shown in Fig. 4 (Ward et al. 2017). In all cases, confidence in a solution is obtained in one or more of three ways: First, when multiple FPASS runs produce the same structure suggesting that the runs have converged. Second, when the difference between experimental and calculated patterns is sufficiently small. Third, when qualitative comparisons can be made to experimental hypotheses. Indeed, even though a complete structure might not exist, for many patterns, researchers have published theories of certain characteristics that should appear in the structure. Comparison between the FPASS structure and these hypotheses can give additional confirmation that a solution is likely correct. Of course each of these comparisons could fail on their own, but in validating FPASS calculations, as many of them are taken into account as possible before a structure is considered solved.

4   Discussion

With the reduced need for human intervention in the structure solution process, an obvious next step is to completely automate the solution of unknown diffraction patterns. To this end, a driver was developed for FPASS to which any number of diffraction patterns, along with the known lattice parameters and compositions, could be submitted (Ward et al. 2017). Each of those patterns was then solved without intervention from a human.

As a test of this system, 95 different crystal structures were selected from the monoclinic, orthorhombic, hexagonal, tetragonal, and cubic crystal families; each structure had between 1 and 30 atoms in the primitive cell. For each structure, FPASS was supplied with the composition, powder diffraction pattern, lattice parameters, and space group. FPASS was run at least 10 times for each of these 95 structures. In these runs, the correct structure was returned at least once for every one of the structures. Even more, the success rate was high for many of the structures that were studied; for 67 of the 95 structure, the best structure was found in at least 90% of the runs. One of the more difficult structures was the 29-atom primitive cell of α-Mn. Even in this case, the best structure was found in 11 out of 15 runs.

With such a large set of examples, some characteristics were observed that can be used to predict the likelihood that a new structure would be solved. It was determined that runs were more likely to succeed when diffraction patterns were of high quality. As the number of symmetrically unique atoms increased, the likelihood dropped for the same number of trials (since the number of degrees of freedom increased). Similarly, as the number of possible combinations of Wyckoff sites increase, the likelihood dropped (again due to an increase in the size of the search space). None of these observations are surprising but do confirm that high-quality diffraction patterns should be used wherever possible.

There are of course limitations to the FPASS method. One important consideration is that DFT produces energies of structures at $T = 0$ K. For structures that undergo phase transformations with increased temperature, or those structures that are metastable, the energy-based biasing will still favor structures that are lowest in energy at low temperatures. This effect should be minimized by the fact that the *best* structure is always taken as the one with optimal match between calculated and observed XRD patterns, and the search is constrained to symmetries observed in the high-temperature phase. However, in complex systems, the biasing of structures based on energies could result in too much effort being spent in the neighborhood of the $T = 0$ K structure.

Another consideration is that FPASS does not currently include any deterministic methods to find the space group or lattice parameters directly from the peak locations of an XRD pattern. Instead, FPASS requires that these values either be input to the calculation or determined during energetic minimization steps. As a result, mistakes in correctly identifying the space group or lattice parameters may result in the structure being incorrectly solved since FPASS cannot currently correct these mistakes directly from the pattern alone. Of course, integrating FPASS with

existing peak indexing software is an obvious addition to the method that would improve its accuracy.

Finally, as with all structure solution methods, FPASS requires a good-quality powder XRD pattern to match against. During refinement, a noisy pattern may produce a structure that is different than what would be obtained using a cleaner one. However, the requirement for a high-quality pattern is somewhat mitigated in hybrid structure solution methods, like FPASS, by the inclusion of structural energy.

Even with those limitations, it has been shown that the FPASS method can be used to solve complex systems that would otherwise not be tenable for structure prediction or refinement methods alone. A primary target for future work should be around automation of the steps outside of the core algorithms. Those steps include the determination of space group and lattice parameters; possible comparison to existing patterns in, for example, the Powder Diffraction File (PDF) (Faber and Fawcett 2002); and handling of multiphase systems. With those additional tools in place, FPASS can provide the foundation for solving XRD patterns with limited human intervention.

References

Nalbandyan VB, Avdeev M, Pospelov AA (2006) Ion exchange reactions of $NaSbO_3$ and morphotropic series $MSbO_3$. Solid State Sci 8:1430–1437. https://doi.org/10.1016/j.solidstatesciences.2006.05.017

Belsky A, Hellenbrandt M, Karen VL, Luksch P (2002) New developments in the inorganic crystal structure database (ICSD): accessibility in support of materials research and design. Acta Crystallogr B 58(3 Part 1):364–369. https://doi.org/10.1107/S0108768102006948

Černý R (2017) Crystal structures from powder diffraction: principles, difficulties and progress. Crystals 7(5):142. https://doi.org/10.3390/cryst7050142, http://www.mdpi.com/2073-4352/7/5/142

Chan MKY, Shirley EL, Karan NK, Balasubramanian M, Ren Y, Greeley JP, Fister TT (2011) Structure of lithium peroxide. J Phys Chem Lett 2(19):2483–2486. https://doi.org/10.1021/jz201072b

Cota LG, de la Mora P (2005) On the structure of lithium peroxide, Li_2O_2. Acta Crystallogr B 61(2):133–136. https://doi.org/10.1107/S0108768105003629

Cullity BD, Stock SR (2001) Elements of X-ray diffraction. Prentice Hall. https://books.google.com/books?id=YA03PQAACAAJ

Curtarolo S, Morgan D, Persson K, Rodgers J, Ceder G (2003) Predicting crystal structures with data mining of quantum calculations. Phys Rev Lett 91:135503. https://link.aps.org/doi/10.1103/PhysRevLett.91.135503

Deem MW, Newsam JM (1989) Determination of 4-connected framework crystal structures by simulated annealing. Nature 342(6247):260–262. http://www.nature.com/doifinder/10.1038/342260a0

Dolci F, Napolitano E, Weidner E, Enzo S, Moretto P, Brunelli M, Hansen T, Fichtner M, Lohstroh W (2011) Magnesium imide: synthesis and structure determination of an unconventional alkaline earth imide from decomposition of magnesium amide. Inorg Chem 50(3):1116–1122. https://doi.org/10.1021/ic1023778

Faber J, Fawcett T (2002) The powder diffraction file: present and future. Acta Crystallogr B 58(3 Part 1):325–332. https://doi.org/10.1107/S0108768102003312

Fehér F, Von Wilucki I, Dost G (1953) Beiträge zur Kenntnis des Wasserstoffperoxyds und seiner Derivate, VII. Mitteil.: Über die Kristallstruktur des Lithiumperoxyds, Li_2O_2. Chem Ber 86(11):1429–1437. https://doi.org/10.1002/cber.19530861111

Föppl H (1957) Die Kristallstrukturen der Alkaliperoxyde. Zeitschrift für anorganische und allgemeine Chemie 291(1–4):12–50. https://doi.org/10.1002/zaac.19572910104

Gao P, Tong Q, Lv J, Wang Y, Ma Y (2017) X-ray diffraction data-assisted structure searches. Comput Phys Commun 213:40–45. https://doi.org/10.1016/j.cpc.2016.11.007

Harris KDM, Tremayne M (1996) Crystal structure determination from powder diffraction data. Chem Mater 8(11):2554–2570. http://pubs.acs.org/doi/abs/10.1021/cm960218d

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871. https://link.aps.org/doi/10.1103/PhysRev.136.B864

Jacobs H, Juza R (1969) Darstellung und Eigenschaften von Magnesiumamid und -imid. Zeitschrift für anorganische und allgemeine Chemie 370(5–6):254–261. https://doi.org/10.1002/zaac.19693700508

Kariuki BM, Serrano-González H, Johnston RL, Harris KDM (1997) The application of a genetic algorithm for solving crystal structures from powder diffraction data. Chem Phys Lett 280(3–4):189–195. https://doi.org/10.1016/S0009-2614(97)01156-1, http://linkinghub.elsevier.com/retrieve/pii/S0009261497011561

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138. https://link.aps.org/doi/10.1103/PhysRev.140.A1133

Lanning OJ, Habershon S, Harris KDM, Johnston RL, Kariuki BM, Tedesco E, Turner GW (2000) Definition of a "guiding function" in global optimization: a hybrid approach combining energy and R-factor in structure solution from powder diffraction data. Chem Phys Lett 317:296–303

Meredig B, Wolverton C (2013) A hybrid computational–experimental approach for automated crystal structure solution. Nat Mater 12(2):123–127

Michel KJ, Wolverton C (2014) Symmetry building Monte Carlo-based crystal structure prediction. Comput Phys Commun 185(5):1389–1393. https://doi.org/10.1016/j.cpc.2014.01.015

Pecharsky VK, Zavalij PY (2005) Fundamentals of powder diffraction and structural characterization of materials. Springer New York. https://books.google.com/books?id=XnkA-tgdtzMC

Putz H, Schön JC, Jansen M (1999) Combined method for ab initio structure solution from powder diffraction data. J Appl Crystallogr 32(5):864–870. https://doi.org/10.1107/S0021889899006615, http://scripts.iucr.org/cgi-bin/paper?S0021889899006615

Rietveld HM (1969) A profile refinement method for nuclear and magnetic structures. J Appl Crystallogr 2(2):65–71. https://doi.org/10.1107/S0021889869006558

Ward L, Michel K, Wolverton C (2015) Three new crystal structures in the Na–Pb system: solving structures without additional experimental input. Acta Crystallogr A 71(5):542–548. https://doi.org/10.1107/S2053273315012516

Ward L, Michel K, Wolverton C (2017) Automated crystal structure solution from powder diffraction data: validation of the first-principles-assisted structure solution method. Phys Rev Mater 1:063802. https://link.aps.org/doi/10.1103/PhysRevMaterials.1.063802

Weston ME, Shoemaker DP (1957) Acta Crystallogr 10(12):735–863. https://doi.org/10.1107/S0365110X57002649

Computational Modeling and the Design of Perovskite Solar Cells

120

Wan-Jian Yin, Hui-Jun Gu, and Xin-Gao Gong

Contents

Abstract

Materials modeling of solar cell absorbers has been developing rapidly in the last few decades, thanks to the capability of density functional theory to calculate total energies and electronic structures and the development of computational algorithms to determine optical properties, band offsets, and defects in semiconductors. We give a brief introduction to the materials systems and the computation of key parameters for photovoltaic application such as the bandgap, effective mass, and optical absorption. Then we take perovskite solar cells as an example, to show how one can understand and engineer real materials through high-throughput first-principles calculations. This is followed by a short perspective.

W.-J. Yin
School of Energy and Soochow Institute for Energy and Materials InnovationS (SIEMIS), Soochow University, Suzhou, Jiangsu, China
e-mail: wjyin@suda.edu.cn

H.-J. Gu · X.-G. Gong (✉)
Key Laboratory for Computational Physical Science (Ministry of Education), Fudan University, Shanghai, China
e-mail: xggong@fudan.edu.cn

1   Introduction

Tetrahedral coordination structures (TCS), with silicon as prototype, have been dominating solar cell research and market, since the first semiconductor solar cell was fabricated based on silicon absorber in the 1950s (Fig. 1) and followed by GaAs and CdTe. The success of TCS in the solar cell market can be ascribed not only to its superior performance but also to its extremely long-term stability. However, as derived from typical covalent systems, their electronic properties are generally defect-intolerant. An extremely low concentration of defects can significantly degrade cell performance. For crystalline Si and GaAs, high-quality crystals have been grown in expensive ways using, e.g., epitaxial growth methods. For CdTe, post-treatments are required for defect passivation so as to achieve high power conversion efficiency (PCE). Chemical "mutation" of the cation based on CdTe has also been applied and results in more ionic quaternary compounds such as $Cu_2(In,Ga)Se_4$ and $Cu_2ZnSn(S,Se)_4$. Because of the presence of multiple elements, defect control is crucial to enhance their efficiency, in particular for $Cu_2ZnSn(S,Se)_4$ (Park et al. 2018; Chen et al. 2013).

In 2010s, hybrid organic-inorganic perovskites, with $CH_3NH_3PbI_3$ as prototype, have introduced a new revolutionary class of solar cell materials, with remarkable performance and economic synthesis methods. In contrast to TCS, in this class of perovskites, corner-sharing octahedra are key structural characteristics (octahedral coordination structure (OCS)), and a unique feature is its defect tolerance. However, stability and toxicity problems are the main obstacles for commercialization. This
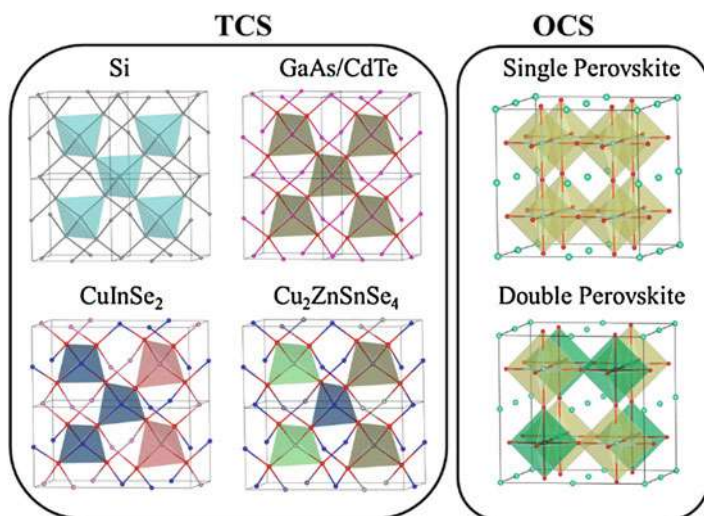
Fig. 1 Tetrahedral coordination structure (TCS), e.g., crystalline Si and the mutated compounds GaAs/CdTe, $CuInSe_2$, and $Cu_2ZnSnSe_4$, and octahedral coordination structure (OCS), e.g., perovskites and the mutated double perovskites, have represented two classes of structural family for solar cell absorbers. (Reorganized by permission of Wang et al. 2019)

situation has stimulated considerable interest in research, and many proposals have been made to provide suitable solutions. Unlike the success in TCS, chemical mutations on OCS are not able to result in functional semiconductors with performance comparable to $CH_3NH_3PbI_3$. Despite extensive theoretical and experimental research, unfortunately, both the stability and toxicity issues of those perovskites have not yet been solved.

As prototype compounds for TCS and OCS, respectively, silicon and α-phase $CH_3NH_3PbI_3$ have the highest symmetry as three-dimensional crystals. Alternatively, previous interest had also been directed to low-symmetry structures, which included Zn_3P_2 ($E_g \sim 1.5$ eV, PCE 6.08%) (Yin and Yan 2013; Fagen 1979; Bhushan and Catalano 1981), Cu_2S ($E_g \sim 1.21$ eV, PCE 11%) (Wang 2012; Hall et al. 1981), and FeS_2 ($E_g \sim 0.95$ eV, PCE 2.8%) (Hu et al. 2012; Bi et al. 2011). Inspired by the recent explosion in perovskite solar cell research, efforts are also being made to include perovskite derivatives in the search of improved stability, such as edge- or face-shared $A_3B_2X_9$ formula ($Cs_3Bi_2I_9$ (PCE ~1.0%), $Cs_3Sb_2I_9$ (PCE < 0.1%)), A_2BX_4 compounds, and Ruddlesden-Popper phases (Xiao et al. 2017). However, none of them has exhibited sufficiently promising properties to survive in the race for novel solar cell materials, and to date the search for novel stable high-efficiency solar cell absorbers with earth-abundant and environment-friendly elements remains an active area in materials modeling. In this chapter, by using perovskite solar cell as example, we will discuss a high-throughput screening approach to materials design for solar energy applications.

2   Perovskite Solar Cells

Organic-inorganic hybrid halide perovskite, with $CH_3NH_3PbI_3$ (MAPbI$_3$) as a prototype, represents an emerging solar cell absorber (Kojima et al. 2009; Kim et al. 2012; Lee et al. 2012; Burschka et al. 2013). Materials modeling has played an important role in understanding the properties and mechanisms responsible for their functionality (Frost et al. 2014; Yin et al. 2015). In order to optimize the strategy for designing high-performance and stable perovskite solar cells, it is crucial to understand the relationship between the performance and the intrinsic properties of the known perovskites.

2.1   Bulk Properties of MAPbI$_3$

2.1.1   Electronic Structure

MAPbI$_3$ can exhibit cubic (space group $Pm\overline{3}m$, >330 K), tetragonal (I4/mcm, 160–330 K), and orthorhombic (Pbnm, <160 K) phases (Kawamura et al. 2002; Baikie et al. 2013; Stoumpos et al. 2013). DFT calculation shows that three phases have similar electronic and optical properties (Yin et al. 2014a). The electronic structure of MAPbI$_3$ has very weak dependence on the orientation of the MA ion,
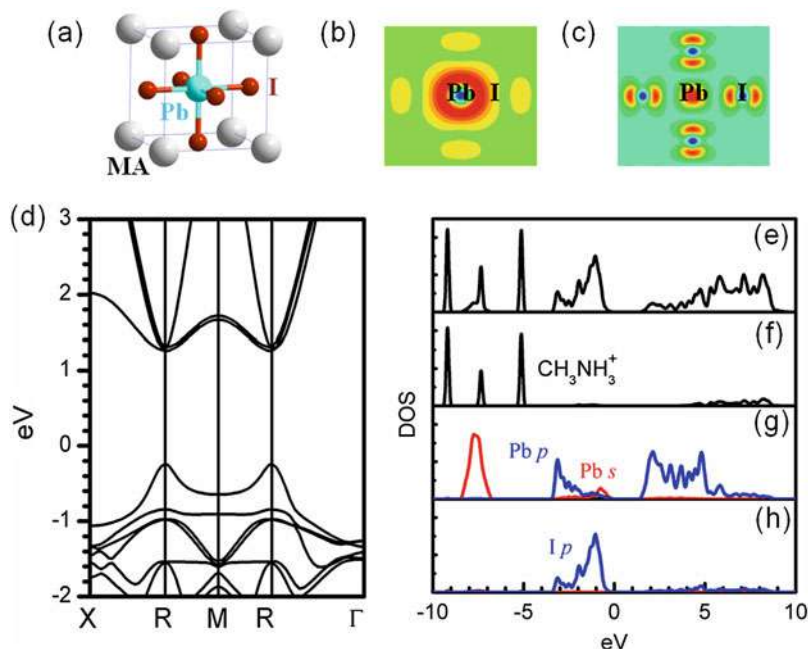
Fig. 2 (**a**) The schematic perovskite structure of $CH_3NH_3PbI_3$. (**b**) and (**c**) are the partial charge density at CBM and VBM, respectively. (**d**) is the band structure of MAPbI$_3$. (**e**–**f**) are total DOS and $CH_3NH_3^+$, Pb, I partial DOS, respectively. The zero in DOS is referred to VBM. The Pb partial DOS has been enlarged by a factor of 5 to make clear the *s* orbital contribution. (Reprint by permission of Yin et al. 2014b)

since MA does not have obvious contribution at the band edge (Fig. 2) but donates one electron to the Pb-I framework.

In perovskite MAPbI$_3$, MA and Pb donate one and two electrons, respectively, to three I ions, forming a bandgap about 1.50 eV (Yin et al. 2014b) between unoccupied Pb *p* orbital and occupied I *p* orbital (see Fig. 2e–h). In general, for a stable perovskite ABX$_3$, the tolerance factor, $t = (r_A + r_X) / \sqrt{2} (r_B + r_X)$, is required to be close to 1. Therefore the size of the A atom should be much larger than that of the B atom. However, Pb in the 2+ charged state is already a large ion so that it is difficult to find a single element A that is large enough to stabilize APbI$_3$. Large-size organic ions such as $CH_3NH_3^+$ can stabilize this perovskite structure but do not impact the intrinsic electronic structure of bulk APbI$_3$. CBM and VBM of MAPbI$_3$ are mainly composed of Pb *p* orbital with negligible coupling with I (Fig. 3b) and antibonding coupling between Pb^{2+} lone-pair *s* and I *p* (Fig. 3c and g) (Yin et al. 2014b). The strong *s-p* antibonding coupling makes upper valence bands dispersive (Fig. 3d) and results in small hole effective mass which is comparable with the electron effective mass, leading to ambipolar conductivity. Similarly to the case of silicon, the calculated hole and electron effective masses of $CH_3NH_3PbI_3$ (Table 1) are balanced, in contrast to conventional thin-film solar cell absorbers GaAs, CIS, CZTSe, and CdTe.
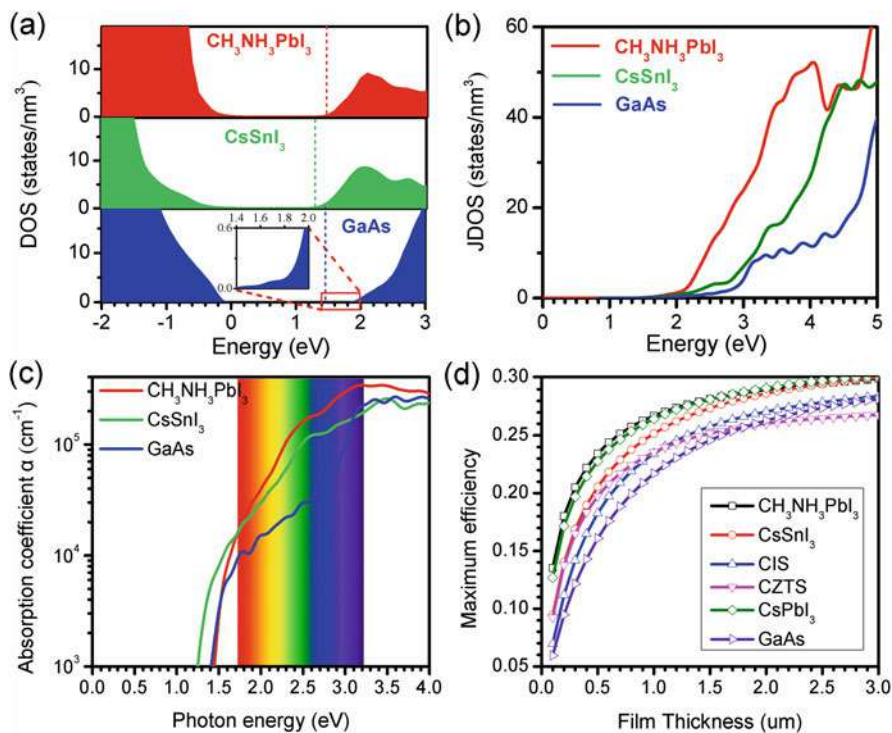
Fig. 3 (**a**) DOS of MAPbI₃, CsSnI₃, and GaAs. The VBMs are referred to as zero energy and CBMs are marked as dashed lines. (**b**) The JDOS of MAPbI₃, CsSnI₃, and GaAs. (**c**) The optical absorptions of MAPbI₃, CsSnI₃, and GaAs. (**d**) Calculated maximum efficiencies of halide perovskites, CIS, CZTS, and GaAs, as a function of film thickness. (Reprint from permission of Yin et al. 2014a)

Table 1 Calculated effective masses for electrons and holes. (Reformatted from permission of Yin et al. 2014a)

	Electron effective mass (m_e)	Hole effective mass (m_e)
CH₃NH₃PbI₃	0.35	0.31
Silicon	0.26	0.29
GaAs	0.07	0.34
CuInSe₂	0.09	0.75
Cu₂ZnSnSe₄	0.10	0.26
CdTe	0.09	0.28

2.1.2 Optical Properties

Strong optical absorption, particularly in the visible light region (1.7–3.0 eV), enables absorber to harvest enough sunlight with extremely thin layer, so that the photo-generated electrons and holes do not have to travel too far to be collected, thus leading to high quantum efficiency. According to the Fermi golden rule, the

optical absorption of a semiconductor at photonic energy hω can be calculated as $\frac{2\pi}{\hbar} \int |\langle v | \hat{p} | c \rangle|^2 \frac{2}{8\pi^3} \delta(E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega) d^3k$, where $\langle v | \hat{p} | c \rangle$ is the optical transition matrix from valence band states $|v\rangle$ to conduction band states $|c\rangle$. To make the model straightforward, the transition matrix $\langle v | \hat{p} | c \rangle$ is considered to be independent of k, and the above absorption formula can be approximated as $\frac{2\pi}{\hbar} |\langle v | \hat{p} | c \rangle|^2 \int \frac{2}{8\pi^3} \delta(E_c(\vec{k}) - E_v(\vec{k}) - \hbar\omega) d^3k$, where the second term is the joint density of states (JDOS) at energy hω. Therefore, optical absorption is directly correlated with dipole transition matrix and JDOS. Calculations (Yin et al. 2014a) show that the transition matrices of $CH_3NH_3PbI_3$ and GaAs near the optical absorption edges are similar; hence the JDOS determines the optical absorption difference between these two materials as shown in Fig. 3. Electronic structure calculations show a large difference in JDOS and therefore in optical absorptions, for $CH_3NH_3PbI_3$ and GaAs. The lower part of the GaAs conduction band is derived from the dispersive s band, whereas the lower part of the $CH_3NH_3PbI_3$ conduction band is mainly composed of Pb p bands. The atomic p states have two more orbitals and more localized orbital shape than the s states. As a result, the density of states (DOS) in the lower CB of the halide perovskites is significantly higher than that of GaAs, leading to higher JDOS and, therefore, stronger optical absorption for halide perovskites. The calculated optical absorption coefficients as in Fig. 3c show that the optical absorption coefficient of $CH_3NH_3PbI_3$ is up to 1 order of magnitude higher than that of GaAs within the visible light range. In comparison to conventional thin-film solar cell absorbers such as CdTe, CIGS, and $Cu_2ZnSnSe_4$ (CZTSe), which have similar chemical characters in their CBs as GaAs, $CH_3NH_3PbI_3$ exhibits sharp optical absorption edge and strong absorption strength in the visible light range.

2.1.3 Spectroscopic Limited Maximum Efficiency (SLME)

In the Shockley-Queisser limit, one considers radiative recombination based on black-body radiation to obtain the maximum theoretical efficiency of a semiconductor solar cell using a single p-n junction. Yu and Zunger (2012) have generalized Shockley-Queisser limit by taking into account the dependence of absorption coefficient of the absorber. A new theoretical limit was thus obtained, i.e., the so-called spectroscopic limited maximum efficiency (SLME), which has been used as a screening parameter for the design of solar cell absorbers. To calculate SLME, the optical absorption spectrum should be first calculated and then combined with the sunlight spectrum to evaluate the maximum conversion efficiency dependence on film thickness. Figure 3d shows the SLME of some common light absorbers as a function of the thickness of the absorber layers, based on DFT-calculated optical absorption spectrum at the GGA-PBE (Perdew et al. 1996) level. The bandgaps are corrected and adjusted to experimental values via the scissor operator. It shows that OCS halide perovskites with direct p-p transition exhibits much higher conversion efficiencies than the conventional TCS semiconductors in thin layers. For example, $CH_3NH_3PbI_3$-based cells can have a maximum efficiency up to 21% with a 0.3 μm absorber, while at the same thickness, GaAs can only achieve 13% (Yin et al. 2014a).

2.2   Influence of Defects

Semiconductor solar cells rely on the formation of *p-n* junction. Therefore, the properties of defects and dopants are crucial. Formation energy and transition energy levels are two key parameters to evaluate the properties of defects and dopants. Apart from point defects, extended defects such as grain boundaries and dislocations also play an important role for carrier separation and recombination. In this subsection, we show how DFT calculations at the PBE level can help understand the defect behavior in perovskites (Yin et al. 2014b; Walsh et al. 2015; Kim et al. 2014; Du 2014).

2.2.1   Formation Energies of Intrinsic Defects

Since the defect formation energies depend on the environmental chemical potentials, the chemical potential for equilibrium growth of $CH_3NH_3PbI_3$ is calculated for a range of compositions, from I-rich/Pb-poor to I-poor/Pb-rich conditions (Yin et al. 2014b). All point defects in $CH_3NH_3PbI_3$ are considered, including three vacancies (V_{MA}, V_{Pb}, V_I), three interstitials (MA_i, Pb_i, I_i), two cation substitutions (MA_{Pb}, Pb_{MA}), and four antisite substitutions (MA_I, Pb_I, I_{MA}, I_{Pb}). The formation energies as a function of Fermi level position at different chemical potentials are shown in Fig. 4a, b, and c, respectively. It is interesting to find out that the conductivity of $CH_3NH_3PbI_3$ can be tuned from intrinsic good *p*-type to good *n*-type by tuning chemical potential from I-rich/Pb-poor to I-poor/Pb-rich conditions. This behavior is different from that of conventional thin-film solar cell absorber CIS(Se) and CZTS, exhibiting only *p*-type conductivity. The formation energies of typical donors (Cu_i and In_{Cu}) in CIS is much higher than that of the V_{Cu} acceptor. Therefore, the V_{Cu} acceptor dominates the conductivity of CISe (Zhang et al. 1998). In $CH_3NH_3PbI_3$, although the formation energies of the dominant donor MA_i and the dominant acceptor V_{Pb} are dependent on chemical potentials, their values ranging from -1 to 1 eV are comparable in the whole range of Fermi level. The origins of the formation of dominating V_{Pb} in $CH_3NH_3PbI_3$ resemble that of V_{Cu} in CISe, ascribed to *s-p* and *p-d* coupling, respectively (Zhang et al. 1998). The fully occupied antibonding state between Pb *s* (Cu *d*) and I *p* (Se *p*) coupling tends to break the chemical bond and form a vacancy.

2.2.2   Transition Energies of Intrinsic Defects

The transition energy of a defect is the Fermi level in the conduction of equal probability for the defect to be either negatively or positively charged. A deep defect, namely, giving rise to a level in the middle of bandgap, will attract carriers and act as Shockley-Read-Hall nonradiative recombination center. The high performance of polycrystalline CIGS solar cell is partially attributed to its shallow dominant defects (V_{Cu}) with transition levels only 0.03 eV above VBM (Zhang et al. 1998). On the contrary, in CZTS the dominant defect (Cu_{Zn}) has a relatively deep transition level (0.10 eV) (Chen et al. 2009), responsible for the low V_{oc} of the CZTS solar cell (Polizzotti et al. 2013). The defect transition levels of $CH_3NH_3PbI_3$ are shown in
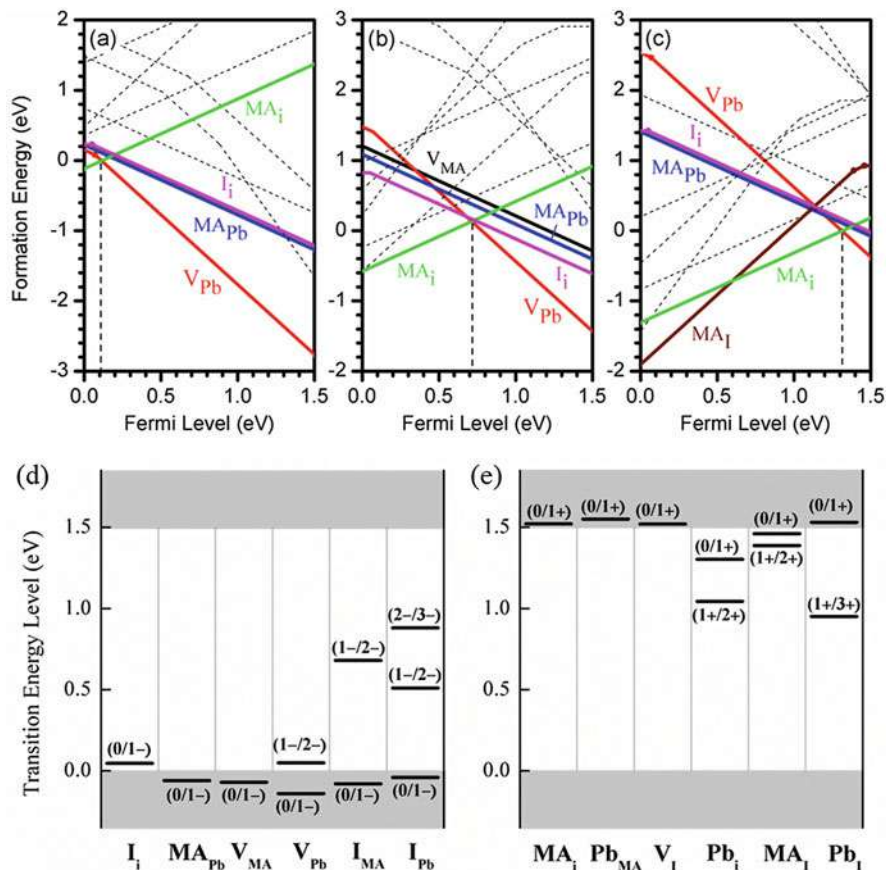
Fig. 4 The formation energies of intrinsic point defects in $CH_3NH_3PbI_3$ at chemical potential (**a**) Pb-poor, (**b**) intermediate, and (**c**) Pb-rich conditions. Defects with much higher formation energies are displayed as dashed lines. The transition energy levels of intrinsic acceptors and donors are shown in (**d**) and (**e**), respectively. (Reprint by permission of Yin et al. 2014b)

Fig. 4. When calculating transition levels, all lattices are fully relaxed to account for the lattice contribution. We can see that the defects with low formation energies (MA_i, V_{Pb}, MA_{Pb}, I_i, V_I, and V_{MA}) have transition level less than 0.05 eV around the band edge. Meanwhile, since defects that create deep levels (I_{Pb}, I_{MA}, Pb_i, Pb_I) have high formation energies, $CH_3NH_3PbI_3$ should have rather low nonradiative recombination rate, explaining the reported long electron-hole diffusion length up to 175 μm (Dong et al. 2015).

2.2.3   Effects of Grain Boundaries

Since $CH_3NH_3PbI_3$ are grown via solution process methods (Jung and Park 2015; Baikie et al. 2013), the materials exhibit polycrystalline phases with grain

Fig. 5 (**a**) The structural model of $\Sigma 5(310)$ GB for $CH_3NH_3PbI_3$. Two identical GBs are included in the supercell to enable the periodicity. (**b**) DOS of $\Sigma 5(310)$ GB model shown in (**a**) with comparison to $CH_3NH_3PbI_3$ bulk calculated from the unit cell. (**c**–**f**) show the partial DOS of selective atoms in (**a**). The (partial) DOS are average per atom, and (**b**) and (**c**) are enlarged twice for clarity. (Reprint by permission of Yin et al. 2014a)
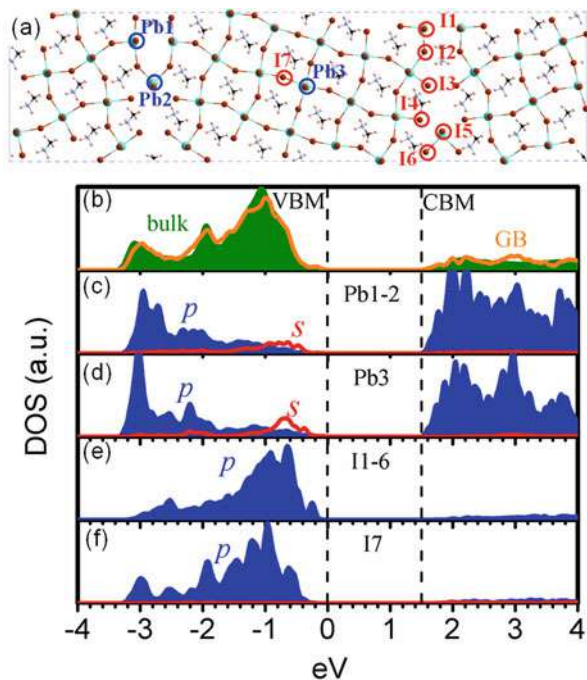
boundaries, which affect the photoelectric properties. In conventional thin-film solar cell absorbers such as GaAs, CIGS, CZTS, and CdTe, GBs usually create deep levels in the bandgaps that are considered detrimental for solar cell performance (Yin et al. 2013, 2014c). In contrast, DFT calculations at the PBE level show (Stoumpos et al. 2013) that GBs in $CH_3NH_3PbI_3$ are intrinsically benign which is consistent with some experimental reports (Edri et al. 2014), although other experiments show that the GBs are not as benign as proposed (Adhyaksa et al. 2018). Figure 5a shows the atomic structure of a supercell containing two $\Sigma 5(310)$ grain boundaries , a typical GB model containing wrong bonds, dangling bonds, and extra bonds. The calculated DOS (Fig. 5b) shows that the $CH_3NH_3PbI_3$ GBs do not generate any states in the gap. The shallow GB energy levels are correlated with the shallow point defects, since the atoms at GBs have similar chemical environments as individual interstitials and vacancies, which create dangling bonds and uncharacteristic bonds such as I-I bonds and Pb-Pb bonds.

3   Design of Novel Perovskite Solar Cell Absorbers

Early exploration of novel solar cell materials was based on trial and error in experiments. Inspired by the change in the nucleotide sequence of a gene or a chromosome in natural organisms, elemental "transmutation" can also be adopted in materials design (Walsh et al. 2012) and is an effective approach to tune the

bandgap, a crucial parameter for a solar cell. For example, elemental transmutation of ZnS (bandgap 2.7 eV (Madelung 2004)) via $Zn^{2+} -> Cu^{1+} -> In^{3+}$ leads to the formation of $CuInS_2$. Although the crystal structure has no major change, the bandgap is significantly reduced to 1.53 eV since the strong coupling between Cu $3d^{10}$ and O $2p^6$ pushes up the VBM, enabling $CuInS_2$ to be a good solar cell absorber. Further mutation on $CuInS_2$ leads to another environment-friendly earth-abundant solar cell absorber Cu_2ZnSnS_4. Elemental mutations have also been made to explore alternative perovskites by replacing divalent Pb^{2+} with monovalent cations such as Na^{1+}, K^{1+}, Rb^{1+}, Cu^{1+}, Ag^{1+}, Au^{1+}, In^{1+}, and Tl^{1+} and trivalent cations such as Ga^{3+}, In^{3+}, Sb^{3+}, and Bi^{3+}, all aiming to find high-efficiency Pb-free stable perovskite, even though the optoelectrical properties of double halide perovskites are found to be inferior to $CH_3NH_3PbI_3$ (Li and Yin 2018).

Apart from elemental transmutation, high-throughput calculations have been used for materials screening for novel solar cell materials, owing to increasing computational power and vast materials data accumulated in the last decades. Here, we take chalcogenide single perovskites as an example to show how high-throughput calculations can be used for screening of solar cell absorbers (Huo et al. 2018).

For high-throughput calculations, first one should choose the material systems of interest:

(i) Chemical composition – we choose 168 kinds of ABX_3 chalcogenide (including oxide) perovskites (A = Mg, Ca, Sr, Ba, Zn, Cd, Sn, Pb; B = Ti, Zr, Hf, Si, Ge, Sn, Pb; X = O, S, Se).
(ii) Second, Crystal structures – four phases including $Pm\overline{3}m$, Pnma, P63/mmc, and Pnma.

Second, key parameters including (a) stability, (b) bandgap, (c) effective mass, and (d) optical absorption should be chosen as the searching criteria. The schematic procedure for high-throughput calculations is shown in Fig. 6.

Thermodynamic stability of ABX_3 perovskites is calculated based on DFT-PBE by

$$H_f = \left(E_{ABX_3}/N_{ABX_3} - E_{AX}/N_{AX} - E_{BX_2}/N_{BX_2} \right)/5$$

where H_f, E_{ABX3}, E_{AX}, and E_{BX2} are formation energy of ABX_3 per atom, total energies of ABX_3, AX, and BX_2, respectively. Negative/positive formation energies indicate that compounds are stable/unstable. Due to the ease of octahedral tilt, perovskite phases could be dynamically unstable despite being thermodynamically stable ($H_f < 0$). Dynamic stabilities are evaluated based on the calculated phonon spectrum. Harmonic phonon spectrum at 0 K is calculated in linear response (also known as density functional perturbation theory, DFPT (Giannozzi et al. 1991; Gonze and Lee 1997; Baroni et al. 2001)) and finite-difference supercell approach (Togo and Tanaka 2015). The room temperature phonon spectrum is obtained by Dynaphopy code (Carreras et al. 2017) which allows us to extract the microscopic
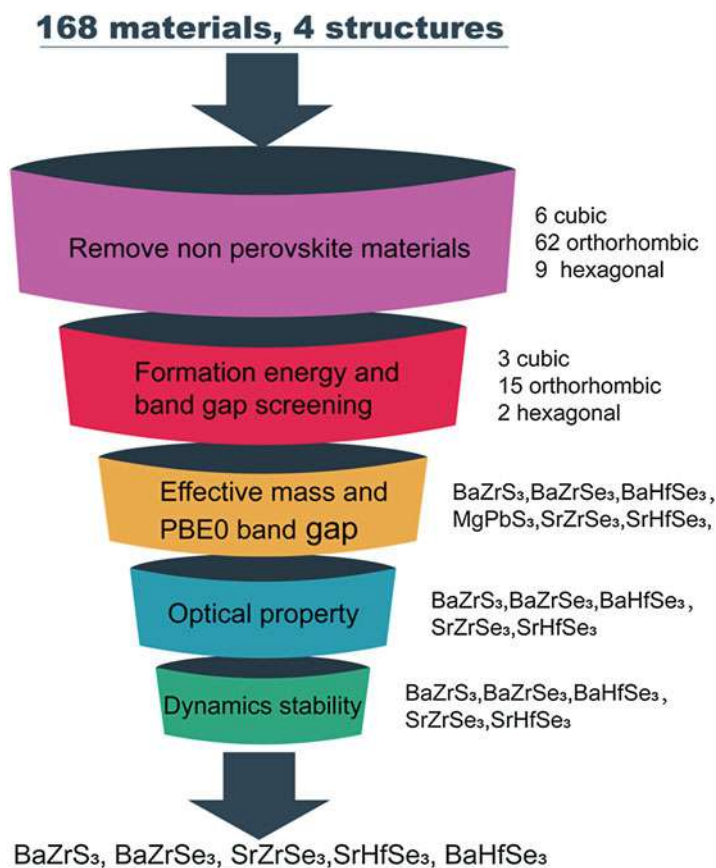
Fig. 6 Screening procedure of novel chalcogenide single perovskites for solar cell applications. (Reprint by permission of Huo et al. 2018)

anharmonic phonon properties from molecular dynamic trajectories using the mode decomposition technique (McGaughey and Larkin 2014) as reported by Sun et al. (Sun et al. 2014).

According to Shockley-Queisser limit, the optimal range of bandgap for efficient single-junction solar cell absorber is about 1.0–1.6 eV. The bandgaps of initial 168 materials with 4 different phases (totally 672 structures) are calculated based on DFT at the PBE level. Due to bandgap underestimation of PBE-DFT, materials with bandgaps in the range 0–1.1 eV were selected for further screening, and the hybrid PBE0 functional (Adamo and Barone 1999) was used to accurately determine the bandgap of selected perovskites.

The effective mass of electrons (holes) around the bottom of the conduction band (top of the valence band) is approximately fitted based on PBE-calculated band structure by $m^* = \hbar^2 \left[\frac{\partial^2 \varepsilon(k)}{\partial k^2} \right]^{-1}$, where $\varepsilon(k)$ is energy dispersion relation functions.

The absorption coefficients as a function of photon energy can be calculated via the expression $\alpha\,(E) = \frac{4\pi e}{hc} \left\{ \frac{\left[\varepsilon_1^2 + \varepsilon_2^2\right]^{1/2} - \varepsilon_1}{2} \right\}^{1/2}$, where $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ is frequency-dependent dielectric function. Usually, dense k-points are required to get a fine optical spectrum, in particular to dispersive bands, leading to high computational cost. Therefore, optical absorption spectra were calculated based on the efficient PBE functional and then simply shifted to reproduce the PBE0 value of the bandgap but keeping the spectral shape of the PBE calculations. The maximum PCE of absorber candidates are estimated via SLME as discussed in Sect. 2.1.3. Despite the absence of experimental results on chalcogenide perovskites, the predicted PCE for $CH_3NH_3PbI_3$ (26%) has been quoted as the ideal target to approach in experiment (Singh et al. 2016).

The procedures mentioned above have been recently applied to search solar cell absorbers among chalcogenide single perovskites. More computational details can be found in Huo et al. (2018). Based on the screening criteria of phase stability, thermodynamic stability, bandgap, carrier effective mass, and optical absorption, five chalcogenide perovskites $BaZrS_3$, $BaHfSe_3$, $SrZrSe_3$, $SrHfSe_3$, and $BaZrSe_3$ have been selected. Among them, $BaZrS_3$ has been successfully synthesized in experiments (Perera et al. 2016a) with a measured value of 1.73 eV for the bandgap, in good agreement with the DFT-PBE0 value of 1.78 eV (Perera et al. 2016b). Device fabrication based on chalcogenide perovskites has not yet been reported thus far.

4   Future Perspectives

The fundamental approach to materials design is to relate the structure of a material to its functional properties. Conventionally, structure-property relations can be understood in terms of existing human knowledge of materials physics and chemistry, such as elemental transmutation and high-throughput calculations as mentioned above. With the accumulation of experimental data, such as the continually expanding Inorganic Crystal Structure Database (ICSD) (https://icsd.fizkarlsruhe.de), and theoretical data, such as those from Materials Project (Jain et al. 2013), OQMD (Kirklin et al. 2015), and NoMAD (Ghiringhelli et al. 2017), the conventional style of materials design is now changing. One finds new routes of data-driven materials design to be merging.

Materials design of perovskite via machine learning (ML) has drawn much current interest. Pilania et al. trained a classification model based on a dataset of 185 experimentally known ABX_3 compounds and proposed 40 new stable perovskites (Pilania et al. 2016). Balachandran et al. applied two models based on a dataset of 390 experimentally reported ABO_3 compounds and predicted 235 new oxide perovskites (Balachandran et al. 2018). Bartel et al. proposed a new tolerance factor by ML based on 576 experimentally known ABX_3 compounds and predicted

the existence of more than a thousand inorganic and hybrid organic-inorganic double perovskites (Bartel et al. 2019). Schmidt et al. calculate 249, 984 potential cubic ABX_3 perovskites via high-throughput calculations and found that ML can reproduce the energetics data with errors around 100 meV/atom (Schmidt et al. 2017). Xie et al. established a smart crystal graph of convolutional neural networks based on DFT-calculated data in the Materials Project (Jain et al. 2013) to predict materials properties directly from crystal structures; and based on their model, 228 perovskites were deemed potentially synthesizable (Xie and Grossman 2018). Recently, through high-throughput DFT calculation, we established a database containing the decomposition energies, considered to be closely related to the thermodynamic stability of 354 halide perovskite candidates. A well-functioning ML model is trained over this theory-based database and further validated by experimental observations of perovskite formability (F_1 score, 95.9%) of 246 $A_2B(I)B(III)X_6$ compounds that are not present in the training database (Li et al. 2019).

   Nowadays machine learning is considered to represent the third generation of computational materials design, statistically driven design, preceded by first generation (computing properties from structures) and second generation (predicting structure from properties). Indeed, the basic methodology is not new and can be traced back to more than 50 years ago (Samuel 1959). Given the long-standing development of algorithms, the procedure of application is by now standard. Therefore, in our opinion, the reliability and validity of data is the crucial prerequisite for a reliable machine learning work. Natural laws are supposed to be buried in the existing training data, and the rules are only learned from these data. Much attention should be given to this aspect of the development in the field. For example, a data-cleaning procedure was recently proposed (Xu et al. 2018) to rationalize the data of perovskite formability for all the ABX_3 and $(AA´)(BB´)X_6$ perovskite compounds in Materials Projects database (Jain et al. 2013). Once combined with this type of "rationalization," we believe that machine learning has the potential to revolutionize current materials discovery (Butler et al. 2018).

References

Adamo C, Barone V (1999) Toward reliable density functional methods without adjustable parameters: The PBE0 model. J Chem Phys 110:6158

Adhyaksa GWP, Brittman S, Āboliņš H, Lof A, Li X, Keelor JD, Luo Y, Duevski T, Heeren RMA, Ellis SR, Fenning DP, Garnett EC (2018) Understanding detrimental and beneficial grain boundary effects in halide perovskites. Adv Mater 30:1804792

Baikie T, Fang YN, Kadro JM, Schreyer M, Wei FX, Mhaisalkar SG, Graetzel M, White TJ (2013) Synthesis and crystal chemistry of the hybrid perovskite ($CH_3 NH_3$) PbI_3 for solid-state sensitised solar cell applications. J Mater Chem A 1:5628

Balachandran PV, Emery AA, Gubernatis JE, Lookman T, Wolverton C, Zunger A (2018) Predictions of new ABO_3 perovskite compounds by combining machine learning and density functional theory. Phys Rev Mater 2:043802

Baroni S, De Gironcoli S, Dal Corso A, Giannozzi P (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73:515

Bartel CJ, Sutton C, Goldsmith BR, Ouyang RH, Musgrave CB, Ghiringhelli LM, Scheffler M (2019) New tolerance factor to predict the stability of perovskite oxides and halides. Science Advances 5:eaav0693, arXiv:1801.07700

Bhushan M, Catalano A (1981) Polycrystalline Zn_3P_2 schottky barrier solar cells. Appl Phys Lett 38:39

Bi Y, Yuan Y, Exstrom CL, Darveau SA, Huang J (2011) Air stable, photosensitive, phase pure iron pyrite nanocrystal thin films for photovoltaic application. Nano Lett 11:4953

Burschka J, Pellet N, Moon SJ, Humphry-Baker R, Gao P, Nazeeruddin MK, Grätzel M (2013) Sequential deposition as a route to high-performance perovskite-sensitized solar cells. Nature 499:316

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A (2018) Machine learning for molecular and materials science. Nature 559:547

Carreras A, Togo A, Tanaka I (2017) DynaPhoPy: A code for extracting phonon quasiparticles from molecular dynamics simulations. Comput Phys Commun 221:221

Chen S, Gong XG, Walsh A, Wei SH (2009) Crystal and electronic band structure of Cu_2ZnSnX_4 (X=S and Se) photovoltaic absorbers: First-principles insights. Appl Phys Lett 94:041903

Chen S, Walsh A, Gong XG, Wei SH (2013) Classification of lattice defects in the kesterite Cu_2ZnSnS_4 and $Cu_2ZnSnSe_4$ earth-abundant solar cell absorbers. Adv Mater 25:1522

Dong Q, Fang Y, Shao Y, Mulligan P, Qiu J, Cao L, Huang J (2015) Electron-hole diffusion lengths > 175 μm in solution-grown $CH_3NH_3PbI_3$ single crystals. Science 347:967

Du MH (2014) Efficient carrier transport in halide perovskites: Theoretical perspectives. J Mater Chem A 2:9091

Edri E, Kirmayer S, Mukhopadhyay S, Gartsman K, Hodes G, Cahen D (2014) Elucidating the charge carrier separation and working mechanism of $CH_3NH_3PbI_{3-x}Cl_x$ perovskite solar cells. Nat Commun 5:3461

Fagen EA (1979) Optical properties of Zn_3P_2. J Appl Phys 50:6505

Frost KB, Brivio F, Hendon C, Schilfgaarde M, Walsh A (2014) Atomistic origins of high-performance in hybrid halide perovskite solar cells. Nano Lett 14:2584

Ghiringhelli LM, Carbogno C, Levchenko S, Mohamed F, Huhs G, Lüders M, Oliveira M, Scheffler M (2017) Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. npj Comput Mater 3:46

Giannozzi P, De Gironcoli S, Pavone P, Baroni S (1991) Ab initio calculation of phonon dispersions in semiconductors. Phys Rev B 43:7231

Gonze X, Lee C (1997) Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. Phys Rev B 55:10355

Hall RB, Birkmire RW, Phillips JE, Meakin JD (1981) Thin-film polycrystalline $Cu_2S/Cd_{1-x}Zn_x$ S solar cells of 10% efficiency. Appl Phys Lett 38:925

Hu J, Zhang Y, Law M, Wu R (2012) Increasing the band gap of iron pyrite by alloying with oxygen. J Am Chem Soc 134:13216

Huo Z, Wei S-H, Yin W-J (2018) High-throughput screening of chalcogenide single perovskites by first-principles calculations for photovoltaics. J Phys D Appl Phys 51:474003

Inorganic Crystal Structure Database (ICSD). https://www.fiz-karlsruhe.de/en/produkte-und-dienstleistungen/inorganic-crystal-structure-database-icsd. Accessed in January 2019

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G et al (2013) Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Mater 1:011002

Jung HS, Park NG (2015) Perovskite solar cells: from materials to devices. Small 11:10

Kawamura Y, Mashiyama H, Hasebe K (2002) Structural study on cubic-tetragonal transition of $CH_3NH_3PbI_3$. J Phys Soc Jpn 71:1694

Kim HS, Lee CR, Im JH, Lee KB, Moehl T, Marchioro A, Moon SJ, Humphry-Baker R, Yum JH, Moser JE, Grätzel M, Park NG (2012) Lead iodide perovskite sensitized all-

solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9%. Sci Rep 2:591

Kim J, Lee S-H, Lee JH, Hong K-H (2014) The role of intrinsic defects in methylammonium lead iodide perovskite. J Phys Chem Lett 5:1312

Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, Aykol M, Rühl S, Wolverton C (2015) The open quantum materials database (OQMD): Assessing the accuracy of DFT formation energies. npj Comput Mater 1:15010

Kojima A, Teshima K, Shirai Y, Miyasaka T (2009) Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. J Am Chem Soc 131:6050

Lee MM, Teuscher J, Miyasaka T, Murakami TN, Snaith HJ (2012) Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites. Science 338:643

Li Z, Yin W-J (2018) Recent progress in Pb-free stable inorganic double halide perovskites. J Semicond 39:071003

Li Z, Xu Q, Sun Q, Hou Z, Yin W-J (2019) Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning. Adv Funct Mater 29:1807280

Madelung O (2004) Semiconductors: data handbook. Springer, Berlin/Heidelberg. ISBN 978-3-642-62332-5

McGaughey A, Larkin JM (2014) Predicting phonon properties from equilibrium molecular dynamics simulations. Ann Rev Heat Transf 17:49

Park JS, Kim S, Xie Z, Walsh A (2018) Point defect engineering in thin-film solar cells. Nat Rev Mater 3:194

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865

Perera S, Hui H, Zhao C, Xue H, Sun F, Deng C, Gross N, Milleville C, Xu X, Watson DF, Weinstein B, Sun Y-Y, Zhang S, Zeng H (2016a) Chalcogenide perovskites: an emerging class of ionic semiconductors. Nano Energy 22:129

Perera S, Hui H, Zhao C (2016b) Chalcogenide perovskites: an emerging class of ionic semiconductors. Nano Energy 22:129

Pilania G, Balachandran P, Kim C, Lookman T (2016) Finding new perovskite halides via machine learning. Front Mater 3:19

Polizzotti A, Repins IL, Noufi R, Wei SH, Mitzi DB (2013) The state and future prospects of kesterite photovoltaics. Energy Environ Sci 6:3171

Samuel A (1959) Some studies in machine learning using the game of checkers. II—Recent progress. IBM J Res Dev 3:210

Schmidt J, Shi J, Borlido P, Chen L, Botti S, Marques M (2017) Predicting the thermodynamic stability of solids combining density functional theory and machine learning. Chem Mater 29:5090

Singh T, Singh J, Miyasaka T (2016) Role of metal oxide electron-transport layer modification on the stability of high performing perovskite solar cells. ChemSusChem 9:2559

Stoumpos CC, Malliakas CD, Kanatzidis MG (2013) Semiconducting tin and lead iodide perovskites with organic cations: Phase transitions, high mobilities, and near-infrared photoluminescent properties. Inorg Chem 52:9019

Sun T, Zhang D-B, Wentzcovitch RM (2014) Dynamic stabilization of cubic perovskite at high temperatures and pressures from ab initio molecular dynamics. Phys Rev B 89:094109

Togo A, Tanaka I (2015) First principles phonon calculations in materials science. Scripta Mater 108:1

Walsh A, Chen S, Wei S-H, Gong X-G (2012) Kesterite thin-film solar cells: Advances in materials modelling of Cu_2ZnSnS_4. Adv Energy Mater 2:400

Walsh A, Scanlon DO, Chen S, Gong X-G, Wei S-H (2015) Self-regulation mechanism for charged point defects in hybrid halide perovskites. Angew Chem Int Ed 54:1791

Wang J, Chen H, Wei S-H, Yin W-J (2019) Materials design of solar cell absorbers beyond perovskites and conventional semiconductors via combining tetrahedral and octahedral coordination. Adv Mater 31:1806593

Wang LW (2012) High Chalcocite: A Solid-liquid hybrid phase. Phys Rev Lett 108:085703

Xiao Z, Meng W, Wang J, Mitzi DB, Yan Y (2017) Searching for promising new perovskite-based photovoltaic absorbers: The importance of electronic dimensionality. 4:206

Xie T, Grossman JC (2018) Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett 120:145301

Xu Q, Li Z, Yin W-J (2018) Rationalizing perovskite data for machine learning and materials design. J Phys Chem Lett 9:6948

Yin W-J, Yan Y (2013) The electronic properties of point defects in earth-abundant photovoltaic material Zn_3P_2: A hybrid functional method study. J Appl Phys 113:013708

Yin W-J, Wu Y, Noufi R, Al-Jassim M, Yan Y (2013) Defect segregation at grain boundary and its impact on photovoltaic performance of $CuInSe_2$. Appl Phys Lett 102:193905

Yin W-J, Shi TT, Yan YF (2014a) Unique properties of halide perovskites as possible origins of the superior solar cell performance. Adv Mater 26:4653

Yin W-J, Shi TT, Yan YF (2014b) Unusual defect physics in $CH_3NH_3PbI_3$ perovskite solar cell absorber. Appl Phys Lett 104:063903

Yin W-J, Wu Y, Wei S-H, Noufi R, Al-Jassim M, Yan Y (2014c) Engineering grain boundaries in $Cu_2ZnSnSe_4$ for better cell performance: A first-principle study. Adv Energy Mater 4:1300712

Yin W-J, Yang J-H, Kang J, Yan Y, Wei S-H (2015) Halide perovskite materials for solar cells: A theoretical review. J Mater Chem A 3:8926

Yu L, Zunger A (2012) Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. Phys Rev Lett 108:068701

Zhang SB, Wei SH, Zunger A, Katayama-Yoshida H (1998) Defect physics of the $CuInSe_2$ chalcopyrite semiconductor. Phys Rev B 57:9642

Index